

Evaluation de la saisonnalité mobile due au calendrier lunaire sur les séries chronologiques par application du modèle espace d'état

Ibrahim Amrani, Abedelhalim Skalli

Ecole Mohammadia d'Ingénieurs
Département Génie industriel
BP 765, Agdal, Rabat, Maroc
ibrahim_amrani@yahoo.fr, skalli@emi.ac.ma

Résumé

La décomposition des séries chronologiques en tendance, cycle, composante saisonnière et résidu, s'avère dans plusieurs cas, insuffisante dans la mesure où la présence d'autocorrélation est remarquée au niveau du résidu. L'extraction d'autres composantes est donc nécessaire. Parmi elles, il est possible de citer les composantes d'événements liés au calendrier Hégirien : Ramadan et les fêtes religieuses ainsi que la composante du calendrier. La particularité des premiers événements est leur mobilité par rapport au calendrier Grégorien. Dans cet article, le modèle espace d'état est utilisé pour modéliser des séries chronologiques mensuelles et les composantes précitées sont introduites. L'intérêt de ce modèle est qu'il permet de modéliser les effets fixes et variables. Les paramètres du modèle sont estimés par maximisation de la fonction vraisemblance en utilisant une méthode quasi newtonienne. Chaque composante n'est maintenue au modèle que si elle permet le respect des tests de validation. Le critère d'information de Akaike (AIC) est utilisé pour choisir entre les modèles qui vérifient ces tests.

Mots clés : Calendrier Hégirien; Composante du calendrier; Critère AIC; Modèle espace d'état; Tests de validation.

Summary

In many cases, the decomposition of time series into trend, cycle, seasonal component and residual, is insufficient insofar as there is an autocorrelation at the level of the residual. The extraction of other components is therefore necessary. Among them, we can mention the components of events related to the Hegirian calendar: Ramadan, religious feasts and also the trading day components. The specificity of the first events is their mobility in the Gregorian calendar. In this paper, the state space method is used to model monthly time series and the above-mentioned components are introduced. The advantage of this form is that it can model both the fixed and variable effects. The parameters of this model are estimated through the maximization of the likelihood function using a quasi-Newton method. Each component is maintained in the model only if the validation tests are satisfied. The Akaike information criterion (AIC) is used to choose amongst the models which respect these tests.

Key words: AIC criterion; Hegirian calendar; State space model; Trading day components; Validation tests.

1 Introduction

Toute activité économique est influencée par des événements qui sont soit exceptionnels (accidents, grève etc...) ou qui présentent des caractères qui leurs sont propres (fêtes, vacances etc...). En prenant le calendrier Grégorien comme référence pour les dates, un événement est dit mobile si sa date change d'une année à une autre. L'exemple le plus connu en occident est la fête de Pâques qui a lieu chaque année entre le 22 mars et le 25 avril. D'un autre côté, Ramadan est le neuvième mois du calendrier Hégirien qui fixe les dates des fêtes musulmanes. Ce calendrier lunaire recule chaque année d'une dizaine de jours par rapport au calendrier Grégorien et ce recule est l'origine de la mobilité des événements qui lui sont associés. Ce type de fête engendre dans la société, des comportements spécifiques et similaires à un effet de saison. D'où l'appellation de saisonnalité mobile.

L'objet de cette étude est d'évaluer l'effet des événements mobiles sur les séries chronologiques unidimensionnelles. Ignorer ces événements peut compliquer l'identification des modèles et empêcher leur validation, voir Bell, R.W et Hillmer, S.C. (1983). Ce problème des événements mobiles a une dimension universelle dans la mesure où il touche diverses régions du monde. Plusieurs auteurs l'ont abordé, Liu, M.L (1980) a développé un moyen de détection de la présence d'une saisonnalité mobile sur les séries chronologiques. Bell, R.W et Hillmer, S.C (1983) ont comparé la qualité des modèles (avec et sans la composante de saisonnalité mobile). Une telle comparaison peut être considérée comme un moyen de détection. Morris, N.D et Pfeffermann, D (1984) ont utilisé le modèle dynamique linéaire pour évaluer l'effet de la saisonnalité mobile. L'intérêt de cet outil, est qu'il permet de modéliser les phénomènes stochastiques, alors que Liu, M.L (1980) a imposé une évolution linéaire et Bell, R.W et Hillmer, S.C (1983) se sont restreints aux phénomènes déterministes. Alper, E.C et Aruoba, S.B (2001) se sont limités également au cas déterministe. Leur étude dans le domaine des fréquences peut être considérée comme un autre moyen pour justifier l'intégration d'une composante de saisonnalité mobile dans un modèle.

Le modèle espace d'état est équivalent au modèle dynamique linéaire. Il est appliqué dans cet article comme il a été étudié dans Harvey, A.C (1989) et Durbin, J et Koopman, S.J (2001). Les modèles construits comportent une tendance, un cycle, une composante saisonnière, une composante de la saisonnalité mobile et la composante du calendrier. L'objectif est d'obtenir des modèles conformes aux tests de validation et le critère AIC est utilisé pour choisir entre les modèles valides. Chaque composante dont la présence dans le modèle permet la validation des tests, n'est retenue que si elle améliore le critère AIC.

Le modèle espace d'état est constitué de deux équations : la première appelée équation de mesure, exprime la série sous forme d'un vecteur d'état, la seconde est appelée équation d'état, elle traduit le fait que le vecteur d'état suit un processus autorégressif d'ordre 1. L'application du modèle espace d'état revient à appliquer un algorithme connu sous le nom de filtre de Kalman. Le filtrage à un instant t , est la projection de l'état du système sur les valeurs de la série, disponibles à cet instant. Les composantes du vecteur d'état, considérées dans cet article sont supposées non stationnaires. Dans ce cas, l'algorithme de Rosenberg étudié par Harvey A.C (1989), est utilisé pour initialiser le filtre de Kalman. Les paramètres inconnus du modèle espace d'état sont estimés par maximisation de la fonction vraisemblance. Une méthode quasi-newtonienne est utilisée pour traiter ce problème d'optimisation.

Dans la section 2, le modèle espace d'état est défini avec les différentes composantes qui peuvent modéliser une série. La section 3 comporte la présentation du filtre de Kalman et le traitement de son initialisation. La méthode BFGS (Broyden, Fletcher, Goldfarb, Shanno) utilisée pour

maximiser la fonction vraisemblance, est décrite, ainsi que les différents tests de validation et le critère AIC. La section 4 traite l'application du modèle espace d'état sur des séries mensuelles marocaines.

2 Le modèle espace d'état : un outil pour modéliser les séries chronologiques.

Soit y_t une série temporelle, l'équation de mesure est : $y_t = z_t \alpha_t + \varepsilon_t$ et l'équation de transition est $\alpha_t = T_t \alpha_{t-1} + R_t \eta_t, t = 1, \dots, n$, z_t est un vecteur $(m, 1)$ déterministe, ε_t est un bruit blanc normal, d'espérance nulle et de variance σ_ε^2 , T_t est une matrice (m, m) déterministe, appelée matrice de transition et η_t est un vecteur bruit blanc normal, centré, de matrice de covariance $\sigma_\varepsilon^2 Q$. Les composantes de η_t , sont souvent supposées indépendantes pour simplifier la forme de la matrice Q . La matrice R_t est une matrice déterministe : elle peut par exemple sélectionner les composantes du vecteur d'état α_t dont les perturbations sont non nulles.

Pour compléter cette définition, il va falloir supposer que :

- (i) le vecteur initial α_0 suit une loi normale de moyenne a_0 et de matrice de covariance P_0
- (ii) les perturbations ε_t et η_t sont indépendantes entre elles et avec α_0

$$E(\varepsilon_s \eta_t') = 0, s, t = 1, \dots, n, E(\varepsilon_t \alpha_0') = 0 \text{ et } E(\eta_t \alpha_0') = 0, t = 1, \dots, n.$$

Dans cette définition, la matrice de covariance de la perturbation η_t est exprimée en fonction de la variance de ε_t . Dans la suite de cet article, les variances vont être exprimées de la même façon. Cette manière de procéder, permet d'exprimer σ_ε^2 en fonction des autres variances, ce qui diminue le nombre des paramètres inconnus.

La première équation indique que le vecteur α_t explique la série dans la mesure où ce dernier est constitué des composantes qui l'influencent :

$$y_t = z_t \alpha_t + \varepsilon_t = \mu_t + c_t + \gamma_t + x_t' \delta_t + \sum_{i=1}^{nb} d_{it} \varphi_{it} + \varepsilon_t$$

La forme de cette équation indique que la série étudiée suit un schéma additif. Si la série suit un schéma multiplicatif, il suffit de lui appliquer la fonction logarithmique. Le test de Buys-Ballot proposé par Bourbonnais, R et Terraza, M (2004) permet de choisir entre les deux schémas.

- La tendance indique le mouvement à long terme de la série, à chaque instant cette composante μ_t est composée de deux variables : la première indique le niveau et l'autre indique la pente.

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \zeta_t$$

$$\beta_t = \beta_{t-1} + \zeta_t$$

Si ζ_t est nul alors la pente est constante, sinon elle varie dans le temps. Les variances de ζ_t et ζ_t sont respectivement $\sigma_\varepsilon^2 \sigma_\beta^2$ et $\sigma_\varepsilon^2 \sigma_\zeta^2$

- Le cycle est un mouvement oscillatoire qui se superpose à la tendance et qui s'étend sur plusieurs années. Dans le modèle non stationnaire proposé par Harvey, A.C (1989), intervient un processus auxiliaire \tilde{c}_t tel que :

$$\begin{pmatrix} c_t \\ \tilde{c}_t \end{pmatrix} = \begin{pmatrix} \cos(\lambda_c) & \sin(\lambda_c) \\ -\sin(\lambda_c) & \cos(\lambda_c) \end{pmatrix} \begin{pmatrix} c_{t-1} \\ \tilde{c}_{t-1} \end{pmatrix} + \begin{pmatrix} \tilde{w}_t \\ \tilde{w}_t^* \end{pmatrix}$$

La variance de \tilde{w}_t et de \tilde{w}_t^* est $\sigma_\varepsilon^2 \sigma_w^2$. λ_c est la fréquence du cycle, elle peut être estimée en maximisant la vraisemblance du modèle. $2\pi/\lambda_c$ est la période du cycle.

- La composante saisonnière est une fluctuation périodique qui s'inscrit dans le cadre d'une année et réapparaît chaque année. Cette variation est détectée par des données mensuelles ou trimestrielles. L'effet des fêtes dont la date est fixée par le calendrier Grégorien (calendrier de référence) est inclus dans la composante saisonnière. Pour modéliser cette composante, la forme trigonométrique proposée par Harvey, A.C (1989) a été choisie :

$$\gamma_t = \sum_{j=1}^{[S/2]} \gamma_{jt}$$

$$\text{Où } \gamma_{jt} = \gamma_{j,t-1} \cos \lambda_j + \gamma_{j,t-1}^* \sin \lambda_j + w_{jt}$$

$$\gamma_{jt}^* = -\gamma_{j,t-1} \sin \lambda_j + \gamma_{j,t-1}^* \cos \lambda_j + w_{jt}^*$$

S étant le nombre de saisons ($S = 12$ pour des données mensuelles) et $\lambda_j = \frac{2\pi j}{S}$ $j = 1, \dots, [S/2]$.

La variance de w_{jt} et w_{jt}^* est $\sigma_\varepsilon^2 \sigma_w^2$ et $[S/2]$ désigne la partie entière de $S/2$

- La composition des jours peut différer d'un mois à un autre et cette différence peut influencer l'allure d'une série chronologique. Parmi les exemples les plus remarquables, on peut citer le chiffre d'affaires des centres commerciaux, qui augmente sensiblement pendant les week-ends. Pour un mois qui comprend cinq week-ends, il est très probable que son chiffre d'affaires soit supérieur à celui d'un mois qui ne comprend que quatre week-ends. Le calcul de l'effet calendrier a été effectué par Dagum, E.B et al (1992), Harvey, A.C (1989) et Bell, W.R (2004), qui ont considéré le modèle stochastique.

$$\text{L'effet de la composante du calendrier est } x_t' \delta_t = \sum_{i=1}^7 n_{it} (\bar{y}_{it} - \bar{y}_{0t}) = \sum_{i=1}^6 (n_{it} - n_{7t}) (\bar{y}_{it} - \bar{y}_{0t})$$

n_{it} est le nombre de jour i appartenant au mois t , \bar{y}_{it} est la moyenne du jour i , relative au mois t

$$\text{et } \bar{y}_{0t} = \frac{1}{7} \sum_{i=1}^7 \bar{y}_{it}$$

Dans les articles de Dagum, E.B et al (1992) et Bell, W.R et Martin, Donald E.K (2004),

$$\delta_{it} = \bar{y}_{it} - \bar{y}_{0t} \text{ et } x_{it} = n_{it} - n_{7t}, i = 1, \dots, 6.$$

δ_{it} est le paramètre du jour i relatif au mois t , $i = 1, \dots, 6$. Pour dimanche considéré comme jour de

$$\text{référence : } \delta_{7t} = \bar{y}_{7t} - \bar{y}_{0t} = -\sum_{i=1}^6 \delta_{it}$$

δ_{it} peut être soit fixe ou variable. Dans ce dernier cas $\delta_{it} = \delta_{it-1} + \chi_{it}$ avec $\chi_t' = (\chi_{1t}, \dots, \chi_{6t})'$ est un vecteur bruit blanc centré, de matrice de covariance $\sigma_\varepsilon^2 Q_\chi$. Dans Dagum, E.B et al (1992), Q_χ est choisie égale à $\sigma_\varepsilon^2 I_6$, (I_6 étant la matrice identité d'ordre (6,6)), alors que dans Bell W.R (2004),

$$Q_\chi = \begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_6^2 \end{pmatrix}. \text{ Notons que dans ces deux articles, les perturbations sont non corrélées entre}$$

elles mais elles sont corrélées avec la perturbation du septième jour qui est considéré comme jour de référence, en effet, si $\chi_{7t} = \delta_{7t} - \delta_{7t-1}$, alors $\chi_{7t} = -\sum_{i=1}^6 \delta_{it} - \delta_{it-1} = -\chi_{1t} - \dots - \chi_{6t}$. Le choix du jour de référence est arbitraire, et n'importe quel jour peut être choisi comme jour de référence, mais les propriétés de ce jour diffèrent des autres jours. Bell, W.R et Martin, Donald E.K (2004) considèrent cette caractéristique comme un inconvénient de ce modèle. Cette remarque est acceptée dans le cadre d'un jugement global de ce modèle. Mais en ne considérant que le côté calcul, aucun reproche n'apparaît, dans la mesure qu'une fois le jour de référence est choisi, six perturbations non corrélées sont utilisées dans le modèle et cette propriété de corrélation déduite de l'équation $\chi_{7t} = -\chi_{1t} - \dots - \chi_{6t}$ n'apparaît en aucun moment au niveau des calculs relatifs à l'estimation des paramètres inconnus.

D'un autre côté, Harvey, A.C (1989) a implicitement supposé que $\bar{y}_{it} = \bar{y}_{it-1} + \chi_{it}^*$, $i = 1, \dots, 7$, et χ_{it}^* sont des bruit blancs centrés, mutuellement indépendants, de variance commune $\sigma_\varepsilon^2 \sigma_{\chi^*}^2$. Dans ce cas les bruits χ_{it} , $i = 1, \dots, 7$, sont corrélés et il est facile de calculer $\text{cov}(\chi_{it}, \chi_{jt})$ qui est

$$\text{égale à } \begin{cases} \frac{6}{7} \sigma_\varepsilon^2 \sigma_{\chi^*}^2 & \text{si } i = j \\ -\sigma_\varepsilon^2 \frac{\sigma_{\chi^*}^2}{7} & \text{si } i \neq j \end{cases}$$

Bell, W.R & Martin, Donald E.K (2004) considèrent cette forme plus restrictive que celle de Bell, W.R (2004). En effet, nous pensons qu'il est peut-être plus pratique de construire des bruits non autocorrélés plutôt que des bruits autocorrélés d'une manière précise. Bell, W.R et Martin, Donald E.K (2004) ont également proposé une transformation trigonométrique pour donner à la matrice de covariance du modèle de Harvey, A.C. (1989), une forme diagonale et se conformer au logiciel qu'ils utilisent. Dans cet article, le modèle de Harvey, A.C (1989) peut être utilisé sans avoir à transformer la matrice de covariance.

- Pour chaque événement mobile i , il est attribué un intervalle de longueur τ_i jours, dans lequel l'influence de cet événement, exercée sur la série est la même. Bell, R.W et Hillmer, S.C (1983) ont pris une même longueur de cet intervalle pour les différentes années de la série, alors que dans cet article, τ_i peut changer d'une année à une autre pour un même événement. En effet, un nombre est fixé à τ_i mais chaque année où le premier ou le dernier jour de cet intervalle coïncide avec un week-end alors ce dernier y est inclus.

d_{it} est la proportion des jours du mois t , qui appartiennent à τ_i . Pour les mois qui ne coïncident pas avec cet intervalle, d_{it} est nul.

φ_{it} est le paramètre de l'événement mobile i , il est généré par un processus de premier ordre :

$$\varphi_{it} = \varphi_{it-1} + \theta_i, \theta_i \text{ étant un bruit blanc centré de variance } \sigma_\varepsilon^2 \sigma_\theta^2.$$

Le calendrier de l'Hégire est composé de 12 mois, la durée des mois est 29 ou 30 jours. C'est ainsi que le nombre de jours d'une année Hégirienne est inférieur à celui d'une année Grégorienne. Cette différence entraîne le recul du premier calendrier par rapport au second. Au Maroc, chaque 29^{ème} jour d'un mois Hégirien, la lune est surveillée, si elle est observée, le mois suivant débute le lendemain, sinon le mois courant dure 30 jours. Les événements mobiles considérés dans cet article sont les suivants :

Ramadan est le neuvième mois du calendrier Hégirien, les musulmans jeûnent pendant ce mois qui est caractérisé par des habitudes et des comportements bien spécifiques.

Fitr est le premier jour du mois de Chaoual, successeur du mois de ramadan. C'est une fête religieuse dans laquelle deux jours de congé sont accordés

Adha est la fête du sacrifice, chaque famille égorge un mouton et cette pratique est fortement suivie par la société. Cette fête a lieu le dixième jour du mois de Hija et deux jours de congé sont accordés. C'est aussi la période du congé annuel de certains métiers comme les maçons.

Maoulid est l'anniversaire du prophète Mohammed, il a lieu le douzième jour du mois de Rabi 2. Deux jours de congé sont accordés.

Suite au recul du calendrier Hégirien par rapport au calendrier Grégorien, des périodes d'un mois Hégirien peuvent être rencontrées à deux reprises dans une année Grégorienne. Par exemple, dans l'année 2000, le mois de janvier a coïncidé avec 7 jours du mois de Ramadan de l'année de l'Hégire 1420, et le mois de décembre a coïncidé avec 27 jours du mois de Ramadan de l'année 1421. Ce scénario ne pose aucun problème pour l'évaluation de l'effet d'un événement mobile dans la mesure où ce dernier est représenté par une variable indicatrice dans chaque mois du calendrier Grégorien.

Les composantes et événements précités sont stochastiques si la variance du bruit correspondant est strictement positive. Lorsqu'elle est nulle, ils sont fixes ou déterministes et dans ce cas $\varphi_t = \varphi_{t-1}$, $t = 1, \dots, n$.

Lorsqu'on modélise une série donnée, il est courant de remarquer la présence de valeurs aberrantes. Celles-ci sont dues à des situations exceptionnelles : accidents, grèves, etc. Pour palier à ce problème, on ajoute au niveau de l'équation de mesure une variable indicatrice δ_t relative à la valeur aberrante ($\delta_t = 1$ si $t = t^*$ et $\delta_t = 0$ si $t \neq t^*$, t^* étant l'instant de la valeur aberrante), et multiplié par un paramètre b_t inconnu et fixe ($b_t = b_{t-1}$) qui est une composante du vecteur d'état α_t .

Le résidu regroupe les fluctuations irrégulières et imprévisibles autour de la tendance et qui ne peuvent être attribuées à l'influence des composantes précitées.

3 Le modèle espace d'état

3-1) Le filtre de Kalman

Soit $a_{t/t}$ la meilleure approximation de α_t basée sur l'information disponible à la date t :

$$a_{t/t} = E(\alpha_t / y_1, \dots, y_t).$$

L'erreur quadratique moyenne de filtrage sur α_t à la date t est $\sigma_e^2 P_{t/t} = E[(\alpha_t - a_{t/t})(\alpha_t - a_{t/t})']$

A la date $t-1$, $a_{t/t-1} = E(\alpha_t / y_1, \dots, y_{t-1})$ est la prévision de α_t

L'erreur quadratique moyenne de prévision est $\sigma_e^2 P_{t/t-1} = E[(\alpha_t - a_{t/t-1})(\alpha_t - a_{t/t-1})']$

La prévision de y_t faite à l'instant $t-1$ est $\tilde{y}_{t/t-1} = E(y_t / y_1, \dots, y_{t-1})$.

L'erreur de prévision à l'horizon 1 est $v_t = y_t - \tilde{y}_{t/t-1}$ et f_t est tel que $\sigma_\varepsilon^2 f_t = \text{var}(v_t) = E(y_t - \tilde{y}_{t/t-1})^2$.

Le filtre de Kalman est un algorithme qui permet un calcul récursif de ces grandeurs.

v_t peut être considérée comme la partie de y_t non corrélée avec y_1, \dots, y_{t-1} , ce qui entraîne que $E(\alpha_t / y_1, \dots, y_{t-1}, y_t) = E(\alpha_t / y_1, \dots, y_{t-1}, v_t)$

$$\begin{aligned} a_{t/t} &= E(\alpha_t / y_1, \dots, y_{t-1}, v_t) = E(\alpha_t / y_1, \dots, y_{t-1}) + \text{cov}(\alpha_t, v_t)(\sigma_\varepsilon^2 f_t)^{-1} v_t \\ &\Rightarrow a_{t/t} = a_{t/t-1} + P_{t/t-1} z_t' f_t^{-1} v_t \text{ (voir Durbin J et Koopman S.J (2001))} \\ \sigma_\varepsilon^2 P_{t/t} &= E[(\alpha_t - a_{t/t})(\alpha_t - a_{t/t})'] = \text{var}(\alpha_t / y_1, \dots, y_{t-1}, v_t) = \sigma_\varepsilon^2 P_{t/t-1} - \text{cov}(\alpha_t, v_t)(\sigma_\varepsilon^2 f_t)^{-1} \text{cov}(\alpha_t, v_t)' \\ &\text{(voir Durbin J et Koopman S.J (2001))} \\ &\Rightarrow P_{t/t} = P_{t/t-1} - P_{t/t-1} z_t' f_t^{-1} z_t P_{t/t-1} \end{aligned}$$

L'équation de transition implique que $a_{t/t-1} = T_t a_{t-1/t-1}$ et $P_{t/t-1} = T_t P_{t-1/t-1} T_t' + R_t Q R_t'$

L'équation de mesure implique que $\tilde{y}_{t/t-1} = z_t a_{t/t-1}$

$$\sigma_\varepsilon^2 f_t = \text{var}(z_t((\alpha_t - a_{t/t-1}) + \varepsilon_t)) = z_t \sigma_\varepsilon^2 P_{t/t-1} z_t' + \sigma_\varepsilon^2 \Rightarrow f_t = z_t P_{t/t-1} z_t' + 1$$

3-2) Estimation des conditions initiales

L'équation de transition est supposée non stationnaire. Pour s'en assurer, il suffit de vérifier que les modules des valeurs propres de la matrice de transition sont supérieurs à 1.

Dans l'algorithme de Rosenberg, α_0 est supposé fixe, ce qui entraîne $a_0 = \alpha_0$ et $P_0 = 0$. Le filtre de Kalman est appliqué une première fois lorsque $\alpha_0 = 0$ et $P_0 = 0$ pour obtenir les erreurs de prévision $v_1^*, v_2^*, \dots, v_n^*$. Il est appliqué une seconde fois lorsque α_0 est fixé à un vecteur non nul donné et $P_0 = 0$, pour obtenir les erreurs de prévision v_1, v_2, \dots, v_n . Sachant que la matrice $P_{t/t-1}$ est la même pour les deux filtres, l'idée est d'exprimer $v_t, t = 1, \dots, n$, en fonction de v_t^* dans la fonction vraisemblance.

Il est facile de prouver en manipulant les équations du filtre de Kalman, qu'à chaque instant t , $a_{t/t-1} = a_{t/t-1}^* + G_t \alpha_0$, $a_{t/t-1}^*$ est la prévision de α_t calculée pour le premier filtre de Kalman et G_t se calcule d'une manière récursive par $G_{t+1} = (T_{t+1} - T_{t+1} P_{t/t-1} z_t' f_t^{-1} z_t) G_t$ avec $G_1 = T_1$

L'erreur de prévision $v_t = y_t - \tilde{y}_{t/t-1} = y_t - z_t a_{t/t-1} = v_t^* - z_t G_t \alpha_0$

3-3) Calcul de la fonction vraisemblance

Par un raisonnement simple, effectué dans Harvey, A.C (1989), le logarithme de la fonction vraisemblance est donné par :

$$-\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma_\varepsilon^2) - \frac{1}{2} \sum_{i=1}^n \log(f_i) - \frac{1}{2\sigma_\varepsilon^2} \sum_{i=1}^n \frac{(v_i^* - z_i G_i \alpha_0)^2}{f_i} \quad (3-3-1)$$

L'annulation de la dérivée partielle de cette équation par rapport à α_0 et σ_ε^2 , permet de les estimer :

$$\hat{\alpha}_0 = \left[\sum_{i=1}^n G_i' z_i' f_i^{-1} z_i G_i \right]^{-1} \sum_{i=1}^n G_i' z_i' f_i^{-1} v_i^* \text{ et } \hat{\sigma}_\varepsilon^2 = \frac{1}{n} \sum_{i=1}^n \frac{(v_i^* - z_i G_i \hat{\alpha}_0)^2}{f_i}$$

$\hat{\alpha}_0$ fourni une estimation des composantes du vecteur d'état qui sont choisies fixes. Cependant, il n'existe que si la matrice entre crochets est inversible.

En remplaçant les expressions de $\hat{\alpha}_0$ et $\hat{\sigma}_\varepsilon^2$ dans l'équation (3-3-1), la fonction vraisemblance concentrée est obtenue : $\log(Lc) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\hat{\sigma}_\varepsilon^2) - \frac{1}{2} \sum_{i=1}^n \log(f_i) - \frac{n}{2}$, c'est cette fonction qui est maximisée pour estimer les paramètres inconnus du modèle.

3-4) Maximisation de la fonction vraisemblance du modèle espace d'état

La fonction vraisemblance concentrée du modèle espace d'état, est une fonction à plusieurs variables, pour laquelle nous cherchons à calculer le maximum. Nous proposons pour cela une méthode quasi-newtonienne. La description de cette méthode va porter sur un problème de minimisation, il est tout à fait simple de ramener un problème de maximisation d'une fonction g à un problème de minimisation : $\max(g) = -\min(-g)$.

La méthode BFGS est une méthode quasi newtonienne qui consiste à construire une suite d'approximations x_1, x_2, \dots, x_k du minimum x^* recherché, à l'étape $k+1$, $x_{k+1} = x_k + \mu_k d_k$ avec $d_k = -D_k \nabla g(x_k)$ et D_k une matrice définie positive qui se calcule de manière itérative :

$$D_{k+1} = D_k + \frac{p_k p_k'}{p_k' q_k} - \frac{D_k q_k q_k' D_k}{q_k' D_k q_k} + v_k v_k v_k'$$

$$\text{avec } p_k = x_{k+1} - x_k$$

$$q_k = \nabla g(x_{k+1}) - \nabla g(x_k)$$

$$v_k = \frac{p_k}{p_k' q_k} - \frac{D_k q_k}{v_k}$$

$$v_k = q_k' D_k q_k$$

D_0 est une matrice définie positive quelconque, en général D_0 est prise égale à la matrice identité.

μ_k est un réel tel que : $\begin{cases} \text{wp1} : g(x_k + \mu_k d_k) \leq g(x_k) + m_1 \mu_k \nabla g(x_k)' d_k & m_1 \in]0,1[\\ \text{wp2} : g(x_k + \mu_k d_k)' d_k \geq m_3 \nabla g(x_k)' d_k & m_3 \in]m_1,1[\end{cases}$ Ce sont les règles

de Wolfe Powell, proposées par Minoux M (1983) qui a indiqué que 0.1 et 0.7 sont des valeurs courantes respectivement pour m_1 et m_3 . La première inégalité assure que l'algorithme est décroissant : $g(x_{k+1}) < g(x_k)$. Bertsekas, D.P (1982) a démontré en utilisant l'inégalité de Cauchy Schwartz que la seconde inégalité assure que D_{k+1} est définie positive si D_k l'est aussi. La mise en œuvre des règles de Wolfe Powell est expliquée dans la figure 1.

Le logiciel STAMP (Koopman, S.J et al (2000)) utilise la méthode BFGS pour maximiser la fonction vraisemblance d'un modèle espace d'état. Pour le choix du pas, il utilise une méthode fondée sur des aspects empiriques (Koopman, S.J et al (2000), p. 164-165). D'un point de vue théorique, les règles de Wolfe Powell sont plus intéressantes.

Lors de l'application de la méthode BFGS, il faut réinitialiser l'algorithme d'une manière périodique pour assurer sa convergence. Minoux M (1983) recommande cette procédure pour chaque p itérations, p est le nombre d'éléments du vecteur x .

Le calcul du gradient de la fonction vraisemblance est compliqué. Chaque composante va être approchée par $\nabla g(x)_i \approx \frac{g(x + hv_i) - g(x - hv_i)}{2h}$, avec v_i le vecteur où la composante i est égale à 1 et les autres sont nulles, et h est un scalaire très petit.

Les paramètres inconnus de la fonction vraisemblance concentrée peuvent être les variances et la fréquence du cycle λ_c . Les variances sont positives et $0 < \lambda_c < \pi$. Un changement de variables est effectué pour se ramener à un problème d'optimisation sans contraintes, voir par exemple (Durbin, J et Koopman, S.J. (2001), p. 143-144)

L'algorithme de calcul des x_k est arrêté après avoir atteint les deux conditions suivantes :

i) $|g(x_{k+1}) - g(x_k)| < c_1$

ii) $\|\nabla g(x_k)\| < c_2$

c_1 et c_2 sont deux seuils à choisir et il faut s'assurer que chaque inégalité est vérifiée pour un certain nombre d'itérations successives.

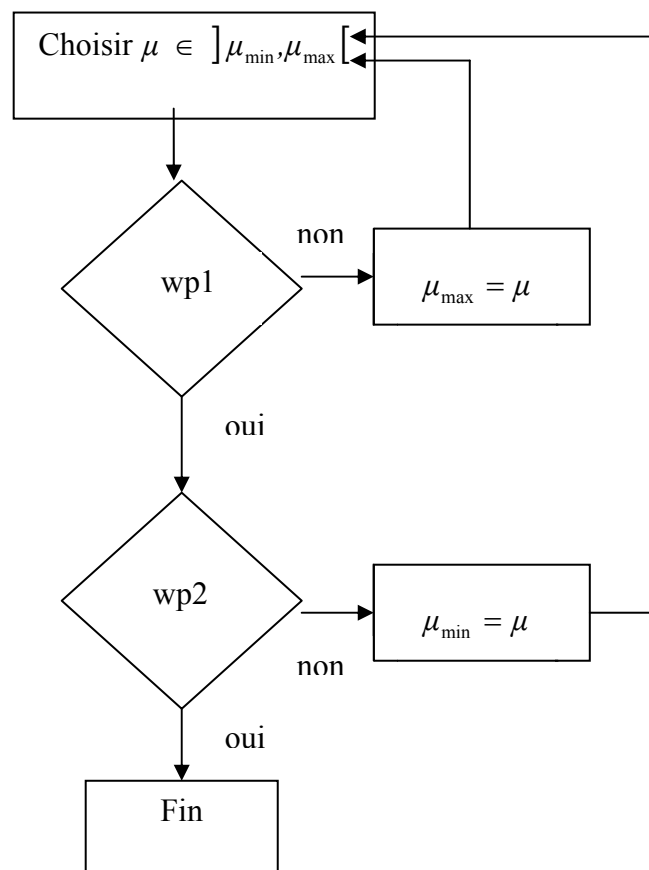


Figure 1 : Organigramme des règles de Wolfe Powell ($\mu_{\min} = 0 ; \mu_{\max} = \infty$)

3-5) Les tests de validation

ε_t et η_t sont supposés être des bruits blancs normaux. Dans ce cas l'erreur de prévision standard définie par $e_t = \frac{v_t}{\sigma_\varepsilon f_t^{1/2}}$ est aussi un bruit blanc normal. Les tests de normalité, d'homoscédasticité et de recherche d'autocorrélation sont appliqués.

Harvey, A.C. (1989) a indiqué que l'hypothèse de normalité de e_t est approximative lorsque les paramètres inconnus du modèle sont estimés. Malgré cette approximation, le test de normalité est appliqué dans cet article. D'ailleurs il est retrouvé dans les application du modèle espace d'état, étudiée dans Harvey, A.C et Durbin J (1986) et Durbin J et Koopman S.J (2001).

3-5-1) Test de normalité.

Les moments de l'erreur de prévision standard sont :

$$m_1 = \frac{1}{n} \sum_{i=1}^n e_i, \quad m_q = \frac{1}{n} \sum_{i=1}^n (e_i - m_1)^q \quad q = 2,3,4$$

Le test de normalité, utilise la statistique $N = n \left(\frac{\beta_1}{6} + \frac{(\beta_2 - 3)^2}{24} \right)$ qui suit une loi χ^2 à deux degrés

de liberté sachant que n est assez grand, $\beta_1^{1/2} = \frac{m_3}{m_2^{3/2}}$ est le coefficient de Skewness et $\beta_2 = \frac{m_4}{m_2^2}$

est le coefficient de Kurtosis. Si $N \geq \chi_{1-\alpha}^2(2)$, l'hypothèse de normalité de l'erreur de prévision standard est rejetée au seuil α .

3-5-2) L'homoscédasticité

Le test d'homoscédasticité de l'erreur de prévision standard, utilise la statistique $H(h) = \frac{\sum_{i=n-h+1}^n e_i^2}{\sum_{i=1}^h e_i^2}$ qui

doit être inférieure au Fisher $F_{h,h}^\alpha$ pour h un entier positif proche de $\frac{n}{3}$ et α un degré de confiance.

3-5-3) La recherche d'autocorrélation

La statistique de Ljung Box est utilisée pour tester l'absence d'autocorrélation entre les $e_i, i = 1, \dots, n$, $Q(k) = n(n+2) \sum_{j=1}^k \frac{c_j^2}{n-j}$ avec $c_j = \frac{1}{nm_2} \sum_{t=j+1}^n (e_t - m_1)(e_{t-j} - m_1)$

$Q(k)$ suit d'une manière asymptotique une loi χ^2 à $(k-p)$ degrés de liberté. Il y a présence d'autocorrélation au seuil α , si $Q(k)$ est supérieure au $\chi^2_{1-\alpha}(k-p)$, p est le nombre de paramètres à estimer.

3-6) Le critère AIC

Le critère AIC est utilisé pour comparer entre les modèles qui vérifient les tests de validation. Par exemple, comparer un modèle qui comporte une composante choisie fixe ou déterministe à un modèle qui comporte la même composante lorsqu'elle est choisie stochastique. Comparer également un modèle qui comporte une composante au modèle qui ne la comporte pas. Le modèle choisi est celui qui dispose du critère le plus faible.

Koopman S.J et al (2000) ont défini le critère AIC pour un modèle espace d'état :

$$AIC = \text{Log}(pev) + 2r/n,$$

$pev = \sigma_\varepsilon^2 \bar{f}$ et $\bar{f} = \lim_{t \rightarrow +\infty} f_t$. Si \bar{f} ne peut pas être obtenue, il est possible d'utiliser f_i avec i assez grand, par exemple $i = n$, r est le nombre de paramètres à estimer plus le nombre de composantes non stationnaires du vecteur d'état.

3-7) Les principales étapes des calculs

Deux algorithmes sont présentés, le premier consiste à calculer la fonction vraisemblance concentrée en appliquant le filtre de Kalman et en partant d'un vecteur x donné, des paramètres inconnus. Dans cet algorithme, chaque fois qu'une variable s_t est écrite sous forme $s_t(t)$, cela signifie tout simplement qu'il est demandé au calculateur de sauvegarder s_t et qu'il est possible de faire appel à cette variable tout au long de ce programme. Le second algorithme optimise la fonction vraisemblance concentrée en utilisant la méthode BFGS et les règles de Wolfe Powell pour le choix du pas. La matrice D est réinitialisée à la matrice unité après chaque p itérations.

Algorithme 1

Choisir x

$$a_0^* = 0, P_0 = 0, G_1(1) = T_1$$

$$M\alpha_0 = 0, V\alpha_0 = 0, l_v = 0, l_f = 0$$

$$t = 1$$

Tant que $t \leq n$

$$a_{t/t-1}^* = T_t a_{t-1/t-1}^*$$

$$P_{t/t-1} = T_t P_{t-1/t-1} T_t' + RQR'$$

$$v_t^*(t) = y_t - z_t a_{t/t-1}^*$$

$$f_t(t) = z_t P_{t/t-1} z_t' + 1$$

$$G_{t+1}(t+1) = (T_{t+1} - T_{t+1} P_{t/t-1} z_t' f_t^{-1} z_t) G_t(t)$$

$$a_{t/t}^* = a_{t/t-1}^* + P_{t/t-1} z_t' f_t^{-1} v_t^*$$

$$P_{t/t} = P_{t/t-1} - P_{t/t-1} z_t' f_t^{-1} z_t P_{t/t-1}$$

$$M\alpha_0 = M\alpha_0 + G_t' z_t' f_t^{-1} z_t G_t$$

$$V\alpha_0 = V\alpha_0 + G_t' z_t' f_t^{-1} v_t^*$$

$$l_f = l_f + \log(f_t)$$

$$t = t + 1$$

Fin Tant que

$$\hat{\alpha}_0 = (M\alpha_0)^{-1}V\alpha_0$$

$$t = 1$$

Tant que $t \leq n$

$$v_t = v_t^*(t) - z_t G_t(t) \hat{\alpha}_0$$

$$lv = lv + \frac{v_t^2}{f_t}$$

$$a_{t/t-1} = a_{t/t-1}^*(t) + G_t(t) \hat{\alpha}_0$$

$$t = t + 1$$

Fin tant que

$$\hat{\sigma}_\varepsilon^2 = \frac{lv}{n}$$

$$\text{Log}(Lc) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\hat{\sigma}_\varepsilon^2) - \frac{lf}{2} - \frac{n}{2}$$

$$g(x) = \text{Log}(Lc)$$

Algorithme 2

$$c_1 = 1, c_2 = 1$$

Tant que $c_1 > 1e-18$ et $c_2 > 1e-6$

$$iter = 0$$

$$d = -\nabla g(x)$$

$$D = I_p$$

Tant que $iter \leq p-1$ et $c_1 > 1e-18$ et $c_2 > 1e-6$

$$\mu = 1$$

$$\mu_{\max} = 100, \mu_{\min} = 0$$

$$x_{suc} = x + \mu d$$

$$wp1 = g(x_{suc}) - g(x) - \nabla g(x)' d (\mu m_1)$$

$$iter = iter + 1$$

Si $wp1 \leq 0$

$$wp2 = (\nabla g(x_{suc}) - \nabla g(x) m_3)' d$$

Si $wp2 \geq 0$

Calculer D_{suc}

$$D = D_{suc}$$

$$d = -D \nabla g(x_{suc})$$

$$c_1 = |g(x_{suc}) - g(x)|$$

$$c_2 = \|\nabla g(x_{suc})\|^{1/2}$$

$$x = x_{suc}$$

Sinon $\mu_{\min} = \mu$

Fin Si

```

Sinon  $\mu_{\max} = \mu$ 
Fin Si
Tant que  $wp1 > 0$  ou  $wp2 < 0$ 
     $\mu = (0.9 \mu_{\max} + 0.1 \mu_{\min})$ 
     $x_{suc} = x + \mu d$ 
     $wp1 = g(x_{suc}) - g(x) - \nabla g(x)' d (\mu m_1)$ 
    Si  $wp1 > 0$ 
         $\mu_{\max} = \mu$ 
    Sinon
         $wp2 = (\nabla g(x_{suc}) - \nabla g(x) m_3)' d$ 
        Si  $wp2 < 0$ 
             $\mu_{\min} = \mu$ 
        Fin Si
    Fin Si
Fin Tant que
Calculer  $D_{suc}$ 
 $D = D_{suc}$ 
 $d = -D \nabla g(x_{suc})$ 
 $c_1 = |g(x_{suc}) - g(x)|$ 
 $c_2 = \|\nabla g(x_{suc})\|^{1/2}$ 
 $x = x_{suc}$ 

```

Fin Tant que

Fin Tant que

4 Applications

Dans cette section, le modèle espace d'état est appliqué sur des séries chronologiques marocaines. L'objectif de cette application est l'obtention d'un modèle qui respecte d'abord les tests de validation et fourni un critère AIC minimal. Le calcul de la log vraisemblance, sa maximisation et les calculs nécessaires aux tests de validation, ont été programmés sur Matlab 6.5.

Chaque valeur aberrante corrigée dans les deux applications qui suivent, est caractérisée par une erreur absolue standard e_i assez grande en valeur absolue. Sa présence dans la série octroie à la statistique N du test de normalité une très grande valeur. La correction de cette valeur aberrante, améliore la vraisemblance optimale du modèle et diminue la statistique N .

4-1) Le trafic ferroviaire des voyageurs exprimé en million de voyageurs kilomètres
janvier 1980 – décembre 2004

L'application du test de Buys-Ballot indique la préférence pour le schéma multiplicatif : La fonction logarithmique est alors appliquée à la série.

L'intervalle de l'événement Ramadan τ_{ram} est choisi égale au nombre de jours de ce mois qui est en général 30. Les intervalles des événements Adha et Maoulid, τ_{adh} et τ_{mao} sont fixés selon le jour de l'événement :

Lundi	Mardi	Mercredi	Jeudi	Vendredi	Samedi	Dimanche
5	3	3	5	4	3	4

Tableau 1 : Les valeurs prises par τ_{adh} et τ_{mao} selon le type de jour de l'événement.

Lors de chaque événement, deux jours de congé sont offerts. Si l'événement est un mercredi alors τ_{adh} ou τ_{mao} sont égales à 3 jours, en effet le trafic des voyageurs s'active en principe dès la veille de l'événement. Si par exemple l'événement est un jeudi alors τ_{adh} (ou τ_{mao}), comprend mercredi qui est la veille, jeudi et vendredi qui sont les deux jours de congé, et samedi et dimanche qui constituent le week-end.

Sachant que Fitr est le premier jour du mois qui succède au mois de Ramadan, la veille de cet événement est incluse dans ce mois. Les valeurs prises par l'intervalle τ_{fit} de l'événement Fitr, sont données dans le tableau 2.

Lundi	Mardi	Mercredi	Jeudi	Vendredi	Samedi	Dimanche
2	2	2	4	3	2	2

Tableau 2 : Les valeurs prises par τ_{fit} selon le type de jour de l'événement Fitr.

Modèles	Log vraisemblance	N	Autocorrélation	Homoscédastisité	AIC
tend + sais + cyc f + aberr(3)	-297	30,42		test accepté	
tend + sais + cyc f + aberr (4)	-314	3,05	test rejeté	test accepté	
tend + sais + cyc f + Adh + aberr (4)	-365	4,09	test rejeté	test accepté	
tend + sais + cyc f + Adh + Ram f + aberr (4)	-401	4,35	test rejeté	test accepté	
tend + sais + cyc f + Adh + Ram f + Fit f + aberr (4)	-425	4,39	test accepté	test accepté	-5,52
tend + sais + cyc f + Adh + Ram f + Fit f + Mao f + aberr (4)	-426	4,19	test accepté	test accepté	-5,51
tend + sais + cyc f + Adh + Ram f + Fit f + calend (Dag)f + aberr (4)	-428	3,55	test accepté	test accepté	-5,48
tend + sais + cyc f + Adh + Ram f + Fit f + calend (Har)f + aberr (4)	-429	3,38	test accepté	test accepté	-5,49

(La lettre f qui suit une composante signifie que celle-ci est choisie fixe)

Tableau 3 : Les modèles construits pour la série trafic des voyageurs

Les composantes fixes introduites dans les différents modèles du tableau 3 ont d'abord été choisies variables mais les variances de leurs perturbations étaient très faibles. La composante du calendrier a été introduite au modèle en tant que variable fixe, dans ce cas, le modèle de Bell, W.R (2004), est le même que celui de Dagum, E.B et al (1992).

Le critère AIC est calculé pour les modèles validés par les tests. Le modèle ayant le critère le plus faible (AIC = -5.52 dans le tableau 3) est celui constitué d'une tendance, d'une composante

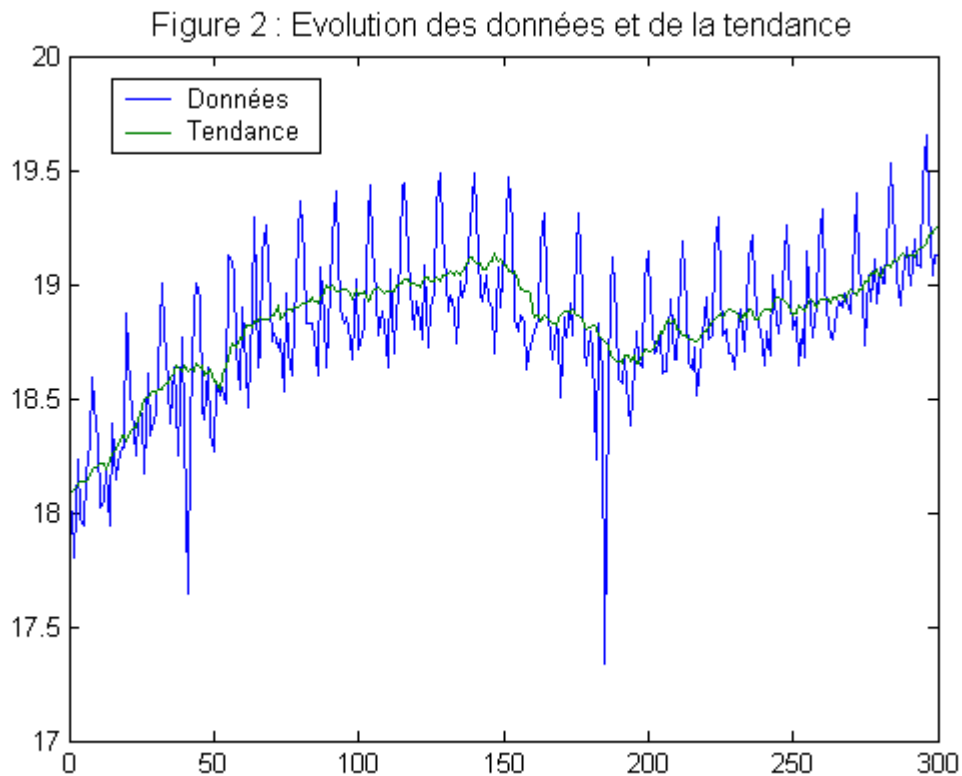
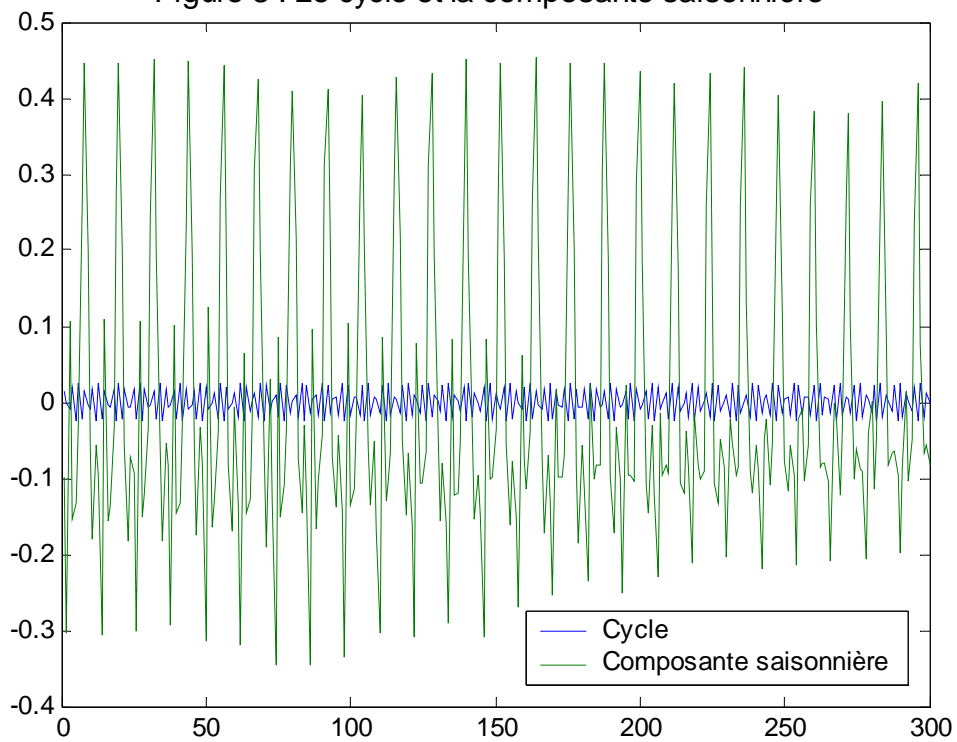


Figure 3 : Le cycle et la composante saisonnière



Adha		Ramadan				Fitr			
oct.-80	5/31	juil.-80	18/31	mars-93	24/31	août-80	2/31	nov.-04	1/15
oct.-81	4/31	août-80	12/31	févr.-94	17/28	août-81	2/31		
sept.-82	1/10	juil.-81	29/31	mars-94	13/31	juil.-82	3/31		
sept.-83	2/15	août-81	1/31	févr.-95	1	juil.-83	2/31		
sept.-84	1/6	juin-82	4/15	mars-95	2/31	juin-84	1/30		
août-85	3/31	juil.-82	22/31	janv.-96	10/31	juil.-84	1/31		
août-86	3/31	juin-83	19/30	févr.-96	10/29	juin-85	2/15		
août-87	5/31	juil.-83	11/31	janv.-97	22/31	juin-86	1/15		
juil.-88	5/31	mai-84	1/31	févr.-97	2/7	mai-87	2/31		
juil.-89	4/31	juin-84	29/30	déc.-97	1/31	mai-88	2/31		
juil.-90	3/31	mai-85	11/31	janv.-98	29/31	mai-89	2/31		
juin-91	2/15	juin-85	19/30	déc.-98	12/31	avr.-90	1/10		
juin-92	1/6	mai-86	22/31	janv.-99	18/31	avr.-91	1/15		
mai-93	1/31	juin-86	4/15	déc.-99	23/31	avr.-92	1/15		
juin-93	1/15	avr.-87	1/30	janv.-00	7/31	mars-93	4/31		
mai-94	3/31	mai-87	29/31	nov.-00	1/10	mars-94	2/31		
mai-95	3/31	avr.-88	13/30	déc.-00	27/31	mars-95	3/31		
avr.-96	1/6	mai-88	17/31	nov.-01	7/15	févr.-96	2/29		
avr.-97	2/15	avr.-89	4/5	déc.-01	16/31	févr.-97	1/14		
avr.-98	1/10	mai-89	6/31	nov.-02	5/6	janv.-98	2/31		
mars-99	4/31	mars-90	4/31	déc.-02	5/31	févr.-98	1/28		
mars-00	3/31	avr.-90	13/15	oct.-03	5/31	janv.-99	2/31		
mars-01	3/31	mars-91	15/31	nov.-03	5/6	janv.-00	2/31		
févr.-02	3/28	avr.-91	1/2	oct.-04	17/31	déc.-00	4/31		
févr.-03	3/28	mars-92	27/31	nov.-04	13/30	déc.-01	2/31		
janv.-04	2/31	avr.-92	1/10			déc.-02	3/31		
févr.-04	2/29	févr.-93	3/14			nov.-03	1/15		

Tableau 4 : Les proportions des mois qui coïncident avec les composantes Adha, Ramadan et Fitr

σ_ε^2	$\sigma_\varepsilon^2 \sigma_\zeta^2$	$\sigma_\varepsilon^2 \sigma_{adh}^2$	$\sigma_\varepsilon^2 \sigma_w^2$	$\sigma_\varepsilon^2 \sigma_\zeta^2$	λ_c
0,0012	0,0003	0,0078	0	0	2,69

Tableau 5 : Estimation des variances et de la fréquence du cycle

Figure 4 : La composante Adha

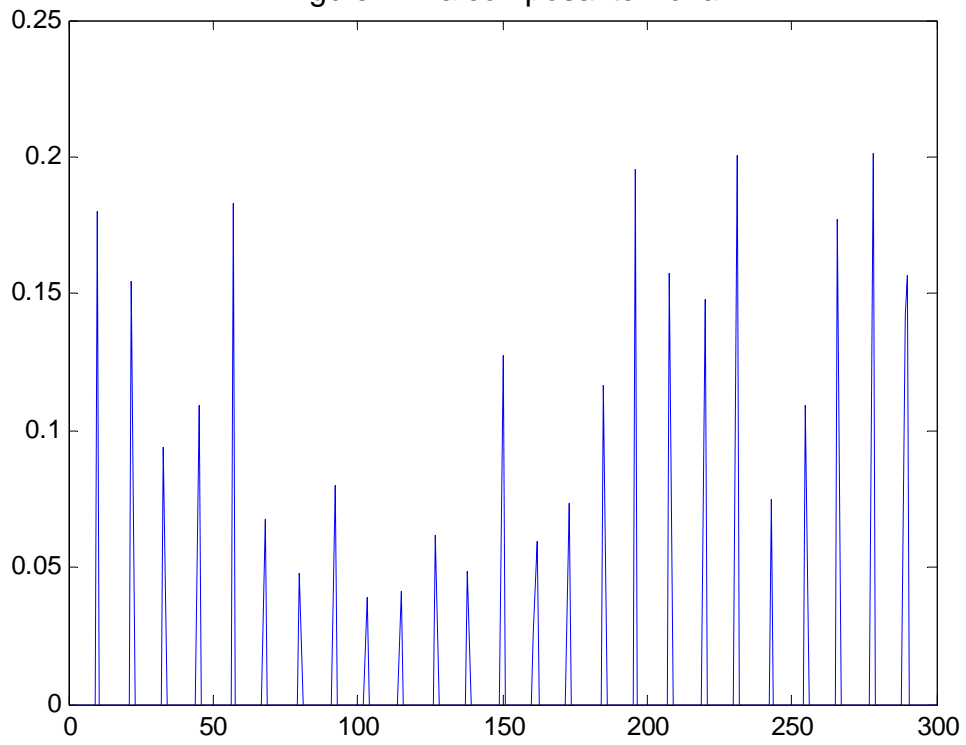


Figure 5 : La composante Ramadan

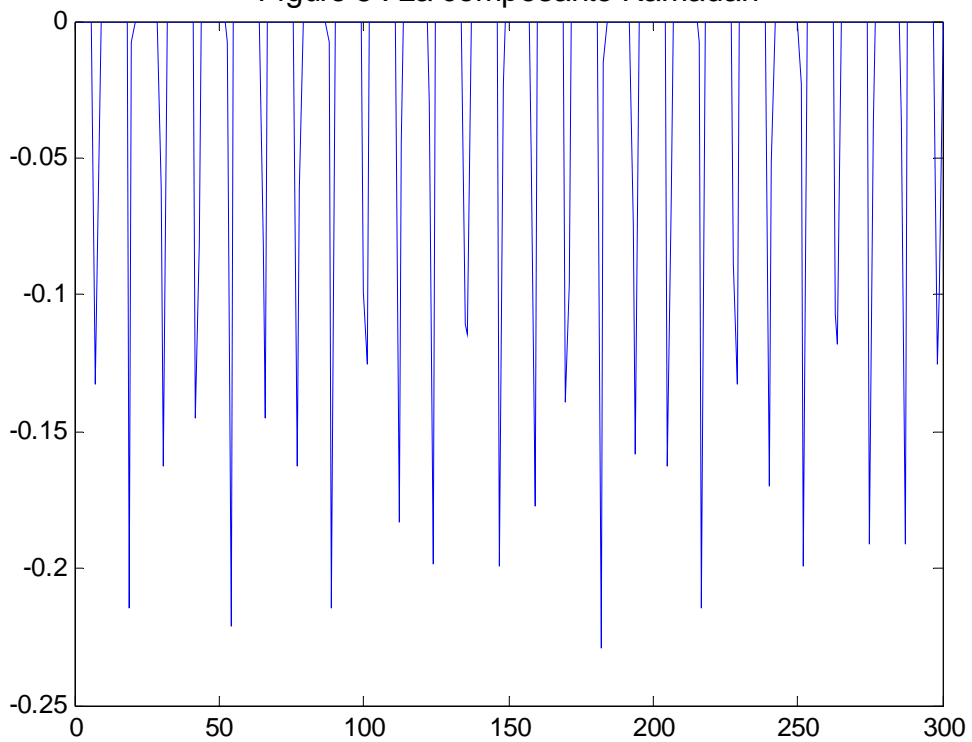
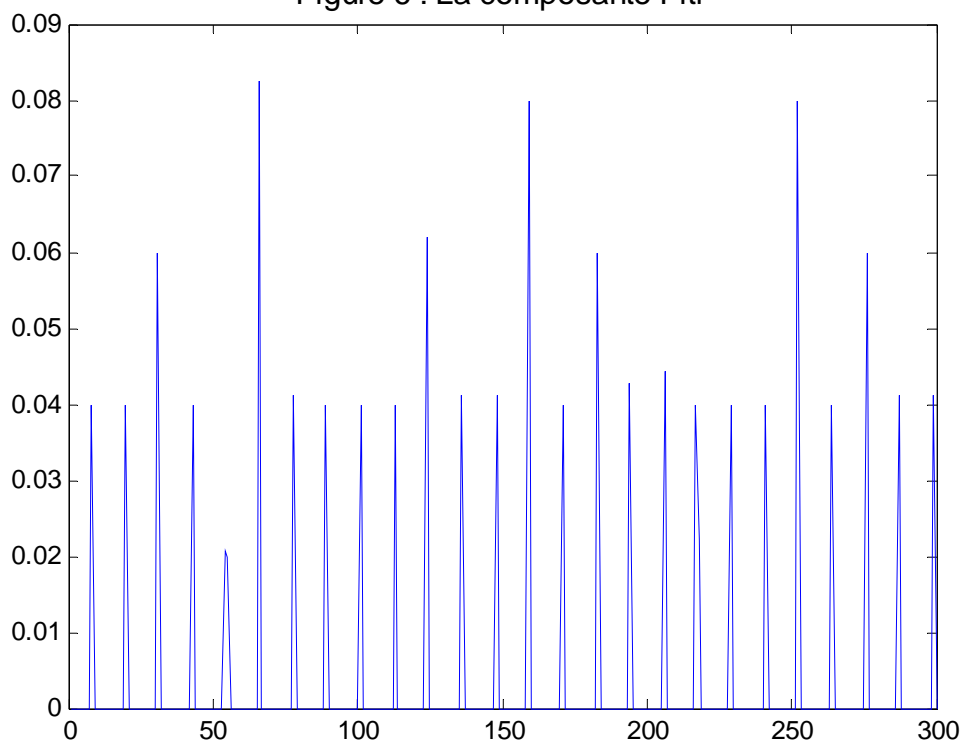


Figure 6 : La composante Fitr



saisonnière, d'un cycle fixe, des paramètres Adha, Ramadan fixe, Fitr fixe et quatre valeurs aberrantes (41, 64, 147, 185), la première et la dernière valeur sont manifestement des valeurs aberrantes (figure 2) et d'après les responsables, correspondent à des grèves.

Les paramètres fixes de Ramadan et Fitr sont respectivement -0.22 et 0.61. La présence de Ramadan réduit le trafic des voyageurs, alors que la présence de Fitr le fait augmenter. Ces deux événements se succèdent et leurs effets sont opposés. Ce modèle va être recalculé en modifiant les valeurs de τ_{fit} et τ_{ram} : l'intervalle τ_{fit} va prendre les valeurs du tableau 1, et τ_{ram} va être réduit du nombre de jours qui ont été inclus à τ_{fit} . Ce choix signifie tout simplement que l'effet réducteur du mois de Ramadan s'arrête la veille de Fitr dont l'effet favorise le trafic. Cette nouvelle modélisation est validée par les tests mais n'améliore pas le critère AIC. Le meilleur modèle du tableau 3 est finalement retenu.

Les figures 4 et 6 indiquent qu' Adha et Fitr favorisent le trafic. L'effet du premier est en général plus important que le second.

4-2) Le trafic ferroviaire des marchandises exprimé en millions de tonnes kilomètres janvier 96 – avril 2004

Le test de Buys-Ballot indique la préférence au schéma additif. Les intervalles de temps sont ceux du tableau 1 pour Adha, ceux du tableau 2 pour Fitr et τ_{ram} est en général égale à 30.

Le tableau 6 indique que la présence d'une valeur aberrante avec la tendance et la composante saisonnière, a suffi pour vérifier les tests de validation. L'introduction des événements Ramadan et Adha diminue le critère AIC. Ces deux événements sont donc retenus pour la modélisation de la série. Par contre ni la présence de la composante du calendrier ni de celle de l'événement Fitr ne

diminue le critère AIC. Enfin, l'introduction d'un cycle a crée des autocorrélations au niveau de l'erreur de prévision standard.

Selon le critère AIC, le meilleur modèle du tableau 6, est constitué d'une tendance, d'une composante saisonnière, d'une valeur aberrante et des paramètres Ramadan fixe et Adha fixe. La valeur aberrante (69) est sans doute une valeur erronée.

Reprenons cette modélisation avec $\tau_{adh} = 2$ et $\tau_{adh} = 3$ (1 jour avant la fête et 2 jours après). Les résultats de cette nouvelle modélisation, sont donnés dans le tableau 7 qui indique que le fait de choisir $\tau_{adh} = 2$ améliore la modélisation par rapport au meilleur modèle du tableau 6, Donc l'effet de l'événement Adha a une durée de deux jours seulement. Ce choix de τ_{adh} est semblable à celui de Bell R.W et Hillmer S.C. (1983) effectué pour la fête de Pâques et est retrouvé dans Bell, W.R. (2004) et Bell, W.R. et Martin, Donald E.K. (2004).

Modèles	Log vraisemblance	N	Autocorrélation	Homoscédastisité	AIC
tend + sais	1796	711,79			
tend + sais + aberr (1)	1764	2,93	test accepté	test accepté	32,86
tend + sais + aberr (1) + calend(Dag)	1760	1,08	test accepté	test accepté	32,96
tend + sais + aberr (1) + ram + calend(Dag)	1751	0,40	test accepté	test accepté	32,71
tend + sais + aberr (1) + adh + calend(Dag)	1756	0,34	test accepté	test accepté	32,81
tend + sais + aberr (1) + adh + ram + calend(Har)	1745	1,05	test accepté	test accepté	32,65
tend + sais + aberr (1) + adh + ram + calend(Bel)	1745	0,88	test rejeté	test accepté	
tend + sais + aberr (1) + adh + ram + calend(Dag)	1745	0,86	test accepté	test accepté	32,63
tend + sais + aberr (1) + adh f + ram f	1749	0,82	test accepté	test accepté	32,53
tend + sais + aberr (1) + adh + ram + cyc	1747	0,26	test rejeté	test accepté	
tend + sais + aberr (1) + adh + ram + fit	1748	0,09	test accepté	test accepté	32,59

(La lettre f qui suit une composante signifie que celle-ci est choisie fixe)

Tableau 6 : Les modèles construits pour la série trafic ferroviaire des marchandises

τ_{adh}	Log vraisemblance	N	Autocorrélation	Homoscédastisité	AIC
2	1745	0,38	test accepté	test accepté	32,45
3	1747	0,19	test accepté	test accepté	32,49

Tableau 7

Dans le meilleur modèle du tableau 7, les paramètres fixes d'Adha et de Ramadan sont négatifs et sont égales respectivement à $-2.05e8$ et $-2.39e7$. Ces événements réduisent alors les valeurs de la série. La figure 8 montre quatre niveaux pour la composante Adha qui correspondent aux proportions 2/30, 2/31, 2/28 et 2/29.

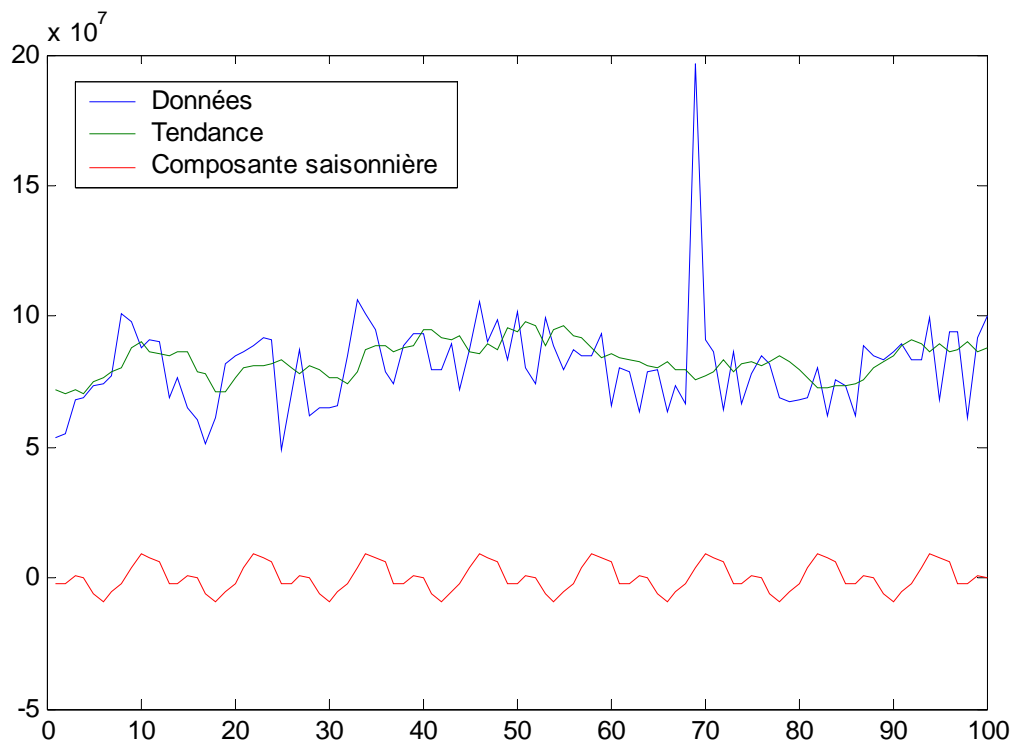


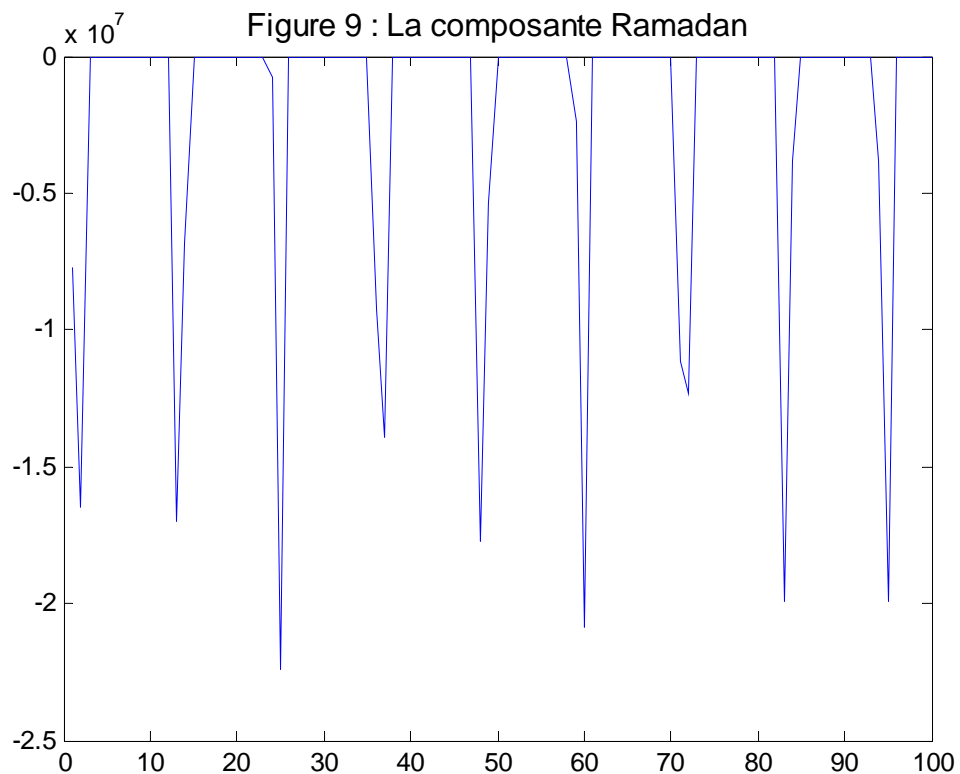
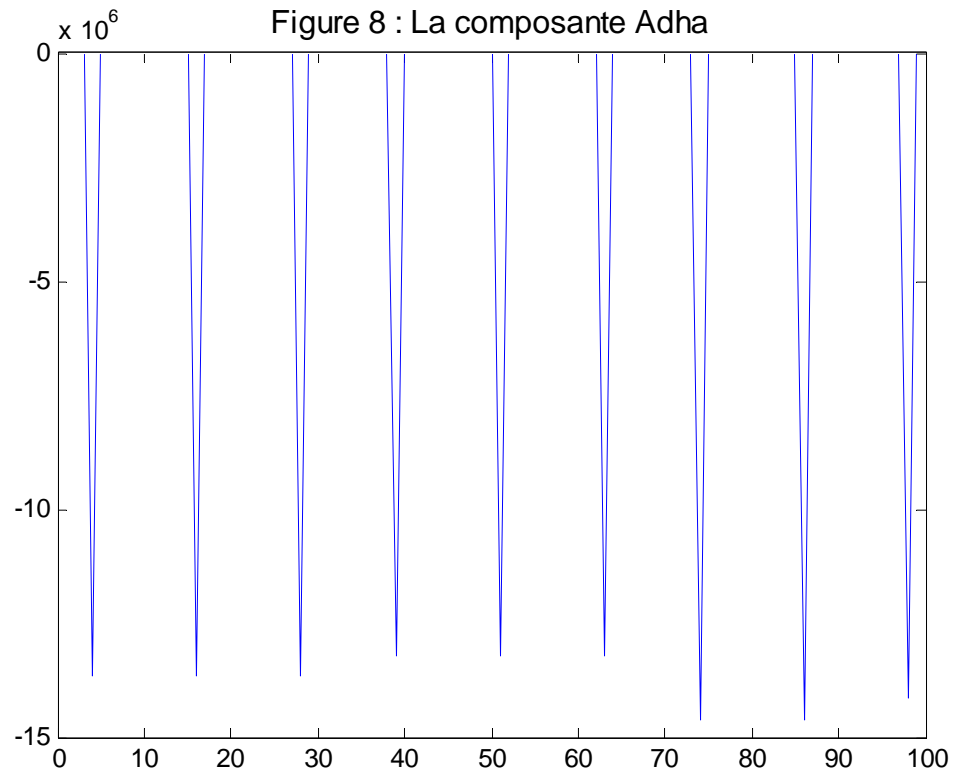
Figure 7 : Evolution des données, de la tendance et de la composante saisonnière

janv.96	fev-96	janv.97	fev.97	dec.97	janv.98	dec.98	janv.99	dec.99
10/31	20/29	22/31	2/7	1/31	29/31	12/31	18/31	23/31
janv.00	nov.00	dec.00	nov.01	dec.01	nov.02	dec.02	oct.03	nov.03
7/31	1/10	27/31	7/15	16/31	5/6	5/31	5/31	5/6

Tableau 8 : Les proportions des mois qui coïncident avec Ramadan

σ_{ε}^2	$\sigma_{\varepsilon}^2 \sigma_{\xi}^2$	$\sigma_{\varepsilon}^2 \sigma_{\zeta}^2$	$\sigma_{\varepsilon}^2 \sigma_w^2$
5,63E+13	9,90E+12	1,76E+07	2,33E+07

Tableau 9 : Estimation des variances



5 Conclusion

Dans un modèle espace d'état, une série chronologique est exprimée sous forme d'une combinaison linéaire de diverses composantes plus une erreur. Chaque composante peut être soit fixe, soit elle évolue dans le temps. Lorsqu'une nouvelle composante est introduite au modèle, la valeur de la

fonction vraisemblance est généralement augmentée. Malgré cela la décision du maintien de cette composante au modèle revient au critère AIC dont l'expression définie au paragraphe 3-6, pénalise les modèles trop paramétrés. Pour cette raison, l'effet calendrier n'a pas été retenu dans les deux applications.

Concernant la recherche du maximum de la fonction vraisemblance, il est nécessaire de vérifier que le hessien est défini positif. En fait on s'est contenté de maximiser au mieux la fonction vraisemblance en s'assurant que le modèle obtenu répond bien aux tests de validation. Cependant la valeur de la fonction vraisemblance calculée, reste tributaire du choix des valeurs initiales des paramètres inconnus.

Bell R.W et Hillmer S.C. (1983) ont utilisé le modèle reg-arima qui est un modèle de régression déterministe avec l'erreur sous forme d'un processus ARIMA. La recherche d'autocorrélation s'est faite au niveau du bruit du processus ARIMA. Dans le modèle espace d'état, la recherche d'autocorrélation se fait au niveau de l'erreur de prévision standard $e_t = \frac{v_t}{\sigma_\varepsilon f_t^{1/2}}$.

L'application du modèle espace d'état est simple et facilement programmable, il suffit d'utiliser un logiciel qui effectue facilement le calcul matriciel. Dans la première série de l'application, le modèle n'a pu être validé qu'après avoir inclus les composantes de saisonnalité mobile. Dans la seconde série, la tendance et la composante saisonnière ont suffi pour valider le modèle, la présence des composantes de saisonnalité mobile a cependant amélioré la qualité de la modélisation. Pour déterminer le nombre de jours des intervalles des événements mobiles, plusieurs scénarios sont mis en compétition. La connaissance des caractéristiques de la série, est d'une grande importance dans la conception des différents scénarios. En effet, la modélisation de celle-ci peut facilement être améliorée en modifiant simplement le choix de ces intervalles.

Cet article a montré la nécessité de tenir compte des phénomènes mobiles liés au calendrier hégirien dans la modélisation des séries chronologiques. Les événements cités concernent l'ensemble du monde musulman. Leur introduction au modèle permet le respect des tests de validation et améliore la qualité de la modélisation.

Références :

- Alper, E.C. et Aruoba, S.B. (2001). Moving holidays and seasonality: an application in the time and the frequency domains for Turkey, www.econ.edu/aruoba/papers/paper1.html
- Bell, W.R. (2004). "On RegComponent time series models and their applications," in *State Space and Unobserved Component Models: Theory and Applications*, eds. Andrew C. Harvey, Siem Jan Koopman, and Neil Shephard, Cambridge, UK: Cambridge University Press.
- Bell, W.R. et Hillmer, S.C. (1983). Modeling time series with calendar variation. *Journal of the American Statistical Association* 78, 526-534
- Bell, W.R. et Martin, Donald E.K. (2004). Modeling time-varying trading-day effects in monthly time series, ASA conference, Toronto.
- Bertsekas D.P. (1982). *Constrained optimization and Lagrange multiplier methods*, Academic Press
- Bourbonnais, R et Terraza, M. (2004). *Analyse des séries temporelles*, Dunod
- Dagum, E.B. Quenneville, B. et Sutradhar B. (1992). Trading day variations multiple regression models with random parameters. *International Statistical Review* 60, 57-73
- Durbin, J et Koopman, S.J. (2001). *Time series analysis by state space methods*, Oxford statistical science series.
- Kazmier, L.J. (1982). *Statistiques de la gestion, théorie et problèmes*, Serie Schaum
- Harvey, A.C. (1989). *Forecasting, structural time series and the kalman filter*, Cambridge

- Harvey, A.C et Durbin J (1986). The effects of seat belt legislation on British road casualties: A case study in structural time series modelling, (with discussion), *Journal of the Royal Statistical Society A* 149, 187-227
- Koopman, S.J. Harvey, A.C, Doornik J.A.D et Shephard N (2000) Stamp: Structural time series analyser, modeller and predictor, Timberlake Consultants Ltd
- Liu, M.L. (1980). Analysis of time series with calendar effects, *Management Science*, 26, 106-112
- Ministère de l'intérieur, (1994). Table de concordances des calendriers grégorien et hézirien.
- Minoux, M. (1983). Programmation mathématique : théorie et algorithmes tome 1, Dunod.
- Morris, N.D. et Pfeffermann, D. (1984). A Kalman filter approach to the forecasting of monthly time series affected by moving festivals, *Journal of Time Series Analysis* 5, 255-268