2.1 Détection de la saisonnalité et désaisonnalisation

Lors de l'analyse d'une chronique, il est nécessaire d'identifier la saisonnalité qui peut éventuellement être observée. Divers types de tests sont mis en place pour détecter la présence de cette composante. Par ailleurs, les méthodes de désaisonnalisation sont classées en deux catégories, à savoir : l'approche paramétrique et l'approche non paramétrique[1].

2.1.1 Approches non paramétriques de désaisonnalisation

2.1.1.1 Tests classiques de présence de saisonnalité et ajustement saisonnier par la technique Census X-11 et X-12-ARIMA

i) Décomposition de la série temporelle avec la méthode X-12-ARIMA et tests de saisonnalité :

Quand la saisonnalité est très apparente dans les séries chronologiques, une première approche consiste à enlever de telles fluctuations saisonnières en utilisant un programme d'ajustement saisonnier. Il s'agit de techniques permettant d'identifier les différentes composantes de la série brute (tendance-cycle, saisonnalité, irrégulier) en lui appliquant des filtres linéaires, ce qui annule ou conserve une composante bien déterminée (tendance-cycle ou saisonnalité). L'irrégulier est par la suite représenté par le résidu de la décomposition.
Ces filtres linéaires sont des moyennes mobiles qui constituent l'outil principal des méthodes Census X-11 construites à partir d'itérations successives de moyennes mobiles d'ordres différents pour mieux estimer les composantes tendancielle, cyclique, saisonnière et irrégulière de la chronique.

Toutefois, cette technique entraîne une perte d'information au niveau de l'extrémité terminale de la série. Cette lacune est comblée par la prévision de valeurs futures de la chronique avant sa désaisonnalisation, et ce à l'aide du modèle ARIMA. C'est ce qui a permis d'étendre la technique X-11 à X-11-ARIMA et ensuite à X-12-ARIMA. Cette dernière contient le module RegARIMA qui permet

[1] Bourbonnais et Terraza (2008) proposent une autre classification selon la nature de la saisonnalité qui peut être soit souple (stochastique : aléatoire en amplitude et/ou en période), soit rigide (déterministe : bien marquée et répétitive).

de détecter et d'enlever tout effet indésirable de la série (valeurs aberrantes, effets du calendrier…).

Une hypothèse centrale pour l'ajustement saisonnier est que la série x_t peut être décomposée en deux composantes non observables, à savoir : une composante non saisonnière x_t^{ns} et une composante saisonnière x_t^{s}. La décomposition peut se faire en suivant un schéma additif tel que $x_t = x_t^{ns} + x_t^{s}$. Or, dans la plupart des cas, la saisonnalité apparaît multiplicative, c'est-à-dire que la variation saisonnière augmente avec un trend, la décomposition se fait alors selon un schéma multiplicatif comme suit : $x_t = x_t^{ns}.x_t^{s}$.

La composante non saisonnière x_t^{ns} contient la tendance et les composantes cyclique et irrégulière (x_t^{TC}, x_t^{I}).

On note également que la tendance et le cycle se comportent de manière similaire, ils sont traités généralement comme étant une seule composante appelée l'extra-saisonnalité. Le trend ou tendance décrit le comportement de long terme de la série. Le cycle est aussi de long terme, mais plus court que la tendance (entre deux et cinq ans). Il provient d'habitude de certains types de cycles d'affaires. Les mouvements croissants sont dus à la croissance, les mouvements décroissants sont appelés « récession ».

La composante saisonnière traite avec les variations annuelles dans la série chronologique, ce qui peut être causé par des effets annuels répétitifs. Ceci est différent de la composante cycle, car la période de répétition est toujours une année. Par exemple, en période hivernale, le nombre de touristes arrivant en Tunisie baisse, en période estivale ce nombre augmente. Cette variation annuelle peut être décrite par la composante saisonnière. Les éléments de cette composante saisonnière sont les facteurs saisonniers estimés par la méthode X-12-ARIMA. L'application des facteurs saisonniers à la série brute permet d'obtenir une série ajustée ou encore corrigée des variations saisonnières (dans le cas d'un modèle multiplicatif, l'application des facteurs saisonniers signifie la division, et dans le cas de modèle additif, on utilise la soustraction). Quant à la composante irrégulière, elle capte les variations aléatoires causées par les événements inhabituels. En effet, les catastrophes naturelles, les grèves et tout autre effet non saisonnier ou ne représentant pas une tendance-cycle peuvent engendrer des fluctuations inattendues dans la chronique économique. Pour illustrer ceci, on considère une série relative aux données d'emploi d'une économie avec comportement cyclique, alors on constate ce qui suit :

- la composante tendance-cycle ou l'extra-saisonnalité (x_t^{TC}) montre d'abord une légère augmentation atteignant un certain maximum, ensuite commence à baisser, puis augmente de nouveau, et ainsi de suite ;
- la composante saisonnière (x_t^s) augmente pendant le deuxième et le troisième trimestre de chaque année, et diminue le reste de l'année. Ce comportement se répète chaque an ;
- la composante irrégulière (x_t^I) ne montre aucune forme régulière, on sait uniquement les deux faits suivants : premièrement, on ne peut pas prévoir sa valeur puisqu'elle varie selon une marche au hasard ; et deuxièmement, la moyenne de long terme de ses valeurs est presque nulle.

Si on disposait d'une information complète sur la série chronologique, il serait possible de séparer les composantes de manière exacte. En effet, on ne peut qu'estimer les valeurs de chaque composante. Ceci nous amène à insister sur le fait que le résultat de l'ajustement saisonnier est une estimation et n'a pas de valeur réelle.

Puisque pour plusieurs séries économiques, la tendance et l'aspect saisonnier ne sont pas parfaitement constants, on considère toujours certains filtres de moyenne mobile pour caractériser un changement de tendance et un changement saisonnier. Dans plusieurs occasions pratiques, tels filtres sont linéaires, symétriques et centrés autour de l'observation courante.

En notant F : l'opérateur avance, i.e., $F = L^{-1}$ et donc $F^k\ x_t = x_{t+k}$, ce filtre de moyenne mobile linéaire est donnée par :

$$C_m(L,F) = c_0 + \sum_{i=1}^{m} c_i (L^i + F^i) \qquad (1)$$

Où $c_0, c_1, \ldots, c_m$ sont les poids de moyenne mobile.

Maravall (1995), et Crether et Nerlove (1970) traitent ces filtres avec plus de détails. Dans la dernière étude, on montre que les filtres comme dans l'équation (1)

aboutissent à des propriétés d'optimalité des estimateurs des composantes saisonnière et non saisonnière.
Comme le révèle Hylleberg (1986) dans son étude, la fameuse méthode d'ajustement saisonnier « Census X-11 » utilise largement les filtres comme dans (1). Cette méthode est un algorithme lisse développé par Shiskin, Young et Musgrave au Bureau Census. Techniquement, elle utilise une série de calculs de moyennes mobiles pour séparer les différentes composantes de la série. Le modèle ARIMA est employé pour prévoir la variable initiale avant de lancer l'algorithme X-11/12. Cette prévision améliore l'exactitude des calculs de moyennes mobiles symétriques. Le programme X-11/12-ARIMA calcule la composante saisonnière et la série ajustée en six étapes qui sont les suivantes :

Etape 1 : Faire entrer la série des données brutes saisonnières ;

Etape 2 : Faire des ajustements préliminaires si nécessaire.
Pour ce faire, on recourt aux facteurs d'ajustement préliminaire, ou encore « facteurs de grèves » du moment qu'ils sont habituellement causés par les grèves ;

Etape 3 : Sélectionner le modèle ARIMA approprié, et prévoir la série temporelle ;

Etape 4 : Décomposer la série de données avec l'algorithme X-11.
Comme c'est mentionné précédemment, cet algorithme est basé sur les calculs de moyennes mobiles. Une moyenne mobile est un type de lissage linéaire, qui remplace un élément d'une chronique avec les moyennes pondérées des éléments voisins.
Une moyenne mobile symétrique prend en considération le même nombre d'éléments de chaque côté de la valeur présente. En revanche, une moyenne mobile asymétrique prend des nombres différents des deux côtés. A la fin d'une chronique, seulement les moyennes asymétriques peuvent être utilisées.
Par ailleurs, le programme X-11 utilise la moyenne mobile de Henderson pour l'estimation de la partie tendance-cycle. La moyenne mobile de Henderson est définie telle qu'elle minimise la somme des carrés des différences tertiaires de la série lissée.

En effet, on considère la série : $X_t = \begin{cases} 1 & \text{si } t = 0 \\ 0 & \text{sinon} \end{cases}$

Sa transformation par une moyenne mobile centrée M d'ordre 2p+1 avec des coefficients $\{\theta_i\}$ est donnée par :

$$MX_t = \begin{cases} 0 & \text{si } t < -p \\ \theta_t & \text{si } -p \le t \le p \\ 0 & \text{si } t > p \end{cases}$$

Cette transformation sera donc lisse si la courbe des coefficients de la moyenne mobile est lisse. Henderson (1916, 1924) a proposé de mesurer la flexibilité de la courbe des coefficients en utilisant la quantité : $H = \sum_i (\Delta^3 \theta_i)^2$. Cette quantité est nulle quand les coefficients $\{\theta_i\}$ sont situés le long d'une parabole, et dans le cas général, elle mesure la différence entre la forme parabolique et la forme de la fonction qui donne les coefficients $\{\theta_i\}$. Ensuite, Henderson a cherché les moyennes centrées d'ordre 2p+1 qui retiennent les polynômes d'ordre 2 et minimisent la quantité *H*. La moyenne mobile de Henderson d'ordre 2p+1 est la solution du problème de minimisation suivant :

$$\begin{cases} Min \sum_i (\Delta^3 \theta_i)^2 \\ \sum_{i=-p}^{i=+p} \theta_i = 1, \sum_{i=-p}^{i=+p} i\theta_i = 0 \text{ et } \sum_{i=-p}^{i=+p} i^2 \theta_i = 0 \end{cases}$$

Etape 5 : La X-11 affiche des tables statistiques pour l'évaluation de la procédure. Ces tables révèlent les résultats de certains tests statistiques. Le test de Fisher est capable de distinguer deux types de saisonnalité.

Le test « F » de saisonnalité évolutive peut déterminer si une saisonnalité évolutive existe dans la série observée. Une saisonnalité évolutive est représentée par un effet d'évolution d'un mois à un autre. Le manque de stabilité rend sa prévision difficile.

Un effet typique de saisonnalité évolutive peut être observé dans l'industrie de construction, où le début et la fin d'une période pic peuvent se produire dans différents moments de l'année, en dépendant du temps présent. Formellement, le test « F » de saisonnalité évolutive vérifie si après la suppression de la composante tendance-cycle, la série chronologique restante peut être considérée comme une série saisonnière évolutive.

En revanche, le test « F » de saisonnalité stable permet de vérifier si la série contient ou non une saisonnalité stable. Il s'agit d'un type de saisonnalité qui se répète au même moment chaque année. Cet aspect stable facilite les prévisions.

Ainsi, pour tester la présence de saisonnalité, une multiplicité de tests paramétriques et non paramétriques ont été suggérés.

a) Tests de saisonnalité stable :

Dans ce cas, on distingue deux tests : l'un paramétrique et l'autre non paramétrique, et dont les hypothèses nulle et alternative sont définies comme suit :

$$H_0 : m_1 = m_2 = \ldots = m_k$$
$$H_1 : m_i \neq m_j \quad \text{pour au moins un couple } (i, j)$$

Avec m_1, …, m_k étant les coefficients saisonniers stables pour les k périodes de saisonnalité.

- le premier test dit « de saisonnalité stable » est un test paramétrique basé sur un modèle d'analyse de la variance à un facteur. On possède k échantillons (les estimations des composantes saisonnière et irrégulière, k = 12 mois ou k = 4 trimestres) de tailles respectives n_1, n_2, …, n_k et on considère que la saisonnalité affecte seulement les moyennes des distributions et non leur variance. Il s'agit donc de tester l'égalité des k moyennes $\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_k$. Ainsi, en considérant chaque échantillon comme provenant d'une variable aléatoire Y_j suivant une loi de moyenne m_j et d'écart-type σ, on obtient les hypothèses formulées supra. L'équation d'analyse de la variance s'écrit :

$$\sum_{q=1}^{k}\sum_{p=1}^{n_q}(y_{pq} - \bar{y})^2 = \sum_{q=1}^{k} n_q(\bar{y}_{\bullet q} - \bar{y}_{\bullet\bullet})^2 + \sum_{q=1}^{k}\sum_{p=1}^{n_q}(y_{pq} - \bar{y}_{\bullet q})^2$$

La variance totale se décompose donc en variance des moyennes, variance due au facteur saisonnalité, et en une variance résiduelle, tel que : $S^2 = S_s^2 + S_R^2$.

Sous l'hypothèse H_0, la statistique de ce test : $F_S = \dfrac{S_s^2/(k-1)}{S_R^2/(n-k)}$ suit une loi de Fisher à $(k-1, n-k)$ degrés de liberté. Si la statistique de Fisher calculée est supérieure la Fisher tabulée, on peut conclure de l'effet significatif du facteur saisonnalité, i.e., les moyennes mensuelles ou trimestrielles ne sont pas toutes égales.

- le second test qui permet de détecter une saisonnalité stable est un test non-paramétrique de Kruskal-Wallis.

Dans ce cas, on suppose que les données sont issues de k échantillons indépendants $A_1, A_2, \ldots, A_k$ de tailles respectives $n_1, n_2, \ldots, n_k$. La statistique du test s'écrit de la manière suivante :

$$W = \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{S_j^2}{n_j} - 3(n+1)$$

Avec S_j étant la somme des rangs des observations de l'échantillon A_j dans la série des n observations où $n = \sum_{j=1}^{k} n_j$. Si H_0 est vraie, la quantité W suit une Chi-deux à $(k-1)$ degrés de libertés.

b) Test de saisonnalité évolutive :

Higginson (1975) a proposé un test de saisonnalité évolutive basé sur un modèle d'analyse de la variance à deux facteurs : le mois ou le trimestre, et l'année. Ce test est fondé sur la modélisation des valeurs de la composante saisonnière-irrégulière pour les années complètes uniquement:

$$Y_{ij} = b_i + m_j + \varepsilon_{ij}$$

Avec :

b_i : définit l'effet de l'année i ($i = 1, \ldots, N$) où N est le nombre d'années complètes.

m_j : définit l'effet du mois, ou trimestre j ($j = 1, \ldots, k$) ;

ε_{ij} : désigne l'effet résiduel, réalisation de lois indépendantes et identiquement distribuées de moyenne nulle.

Le test repose sur la décomposition suivante : $S^2 = S_{IM}^2 + S_{IA}^2 + S_R^2$, avec :

- S^2 : la somme des carrés totale : $S^2 = \sum_{j=1}^{k} \sum_{i=1}^{N} (Y_{ij} - \bar{Y}_{\bullet\bullet})^2$ avec $\bar{Y}_{\bullet\bullet} = \sum_{j=1}^{k} \sum_{i=1}^{N} Y_{ij} /(kN)$

- S_{IM}^2 : la somme des carrés « Inter Mois » :

$$S_{IM}^2 = N(\sum_{j=1}^{k}(\overline{Y}_{\bullet j} - \overline{Y}_{\bullet\bullet})^2 \quad \text{avec} \quad \overline{Y}_{\bullet j} = \sum_{i=1}^{N} Y_{ij} / N$$

- S_{IA}^2 : la somme des carrés « Inter Année » :

$$S_{IA}^2 = k\sum_{i=1}^{N}(\overline{Y}_{i\bullet} - \overline{Y}_{\bullet\bullet})^2 \quad \text{avec} \quad \overline{Y}_{i\bullet} = \sum_{j=1}^{k} Y_{ij} / k$$

- S_R^2 : la somme des carrés « résiduelle » : $S_R^2 = \sum_{j=1}^{k}\sum_{i=1}^{N}(Y_{ij} - \overline{Y}_{i\bullet} - \overline{Y}_{\bullet j} + \overline{Y}_{\bullet\bullet})^2$.

On cherche à tester l'hypothèse nulle selon laquelle la saisonnalité n'évolue pas au cours des années : $H_0 : b_1 = b_2 = \ldots = b_N$. La statistique utilisée dans ce test est la suivante :

$F_M = \dfrac{S_{IA}^2 / (N-1)}{S_R^2 / (N-1)(k-1)}$. Si H_0 est vraie, elle suit une loi de Fisher $(N - 1)$ et $(N - 1)(k - 1)$ degrés de libertés.

c) Test de saisonnalité identifiable :

Ce test vient compléter les tests évoqués ci-dessus. Il est construit à partir des valeurs des statistiques de Fisher du test paramétrique de saisonnalité stable et du test de saisonnalité évolutive (Lothian et Morry, 1978).
La statistique de test, notée T, s'exprime comme suit :

$$T = \left(\frac{T_1 + T_2}{2}\right)^{1/2} \quad \text{avec} \quad T_1 = \frac{7}{F_S} \quad \text{et} \quad T_2 = \frac{3F_M}{F_S}$$

Le diagramme suivant résume les différents tests de présence de saisonnalité présentés ci-dessus.

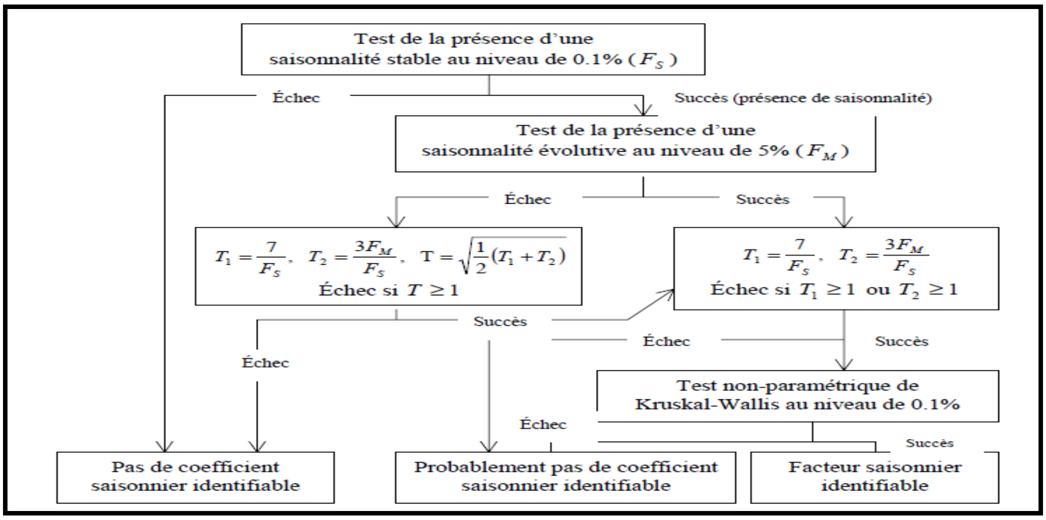

Fig. 3.1 – Test de présence de saisonnalité identifiable[1]

Etape 6 : Si le test révèle la présence d'une saisonnalité dans la série, alors on doit la désaisonnaliser. Pour ce faire, on utilise la méthode X-12-ARIMA qui est une amélioration de X-11-ARIMA.

Enfin, l'analyste décide si la chronique résultant de l'ajustement est appropriée en se basant sur les résultats du test statistique ou une nouvelle analyse est requise.

Le programme aide à prendre cette décision avec une liste de contrôles, indiquant les valeurs critiques pour chaque test statistique.

ii) Les procédures de l'ajustement saisonnier :

L'ajustement saisonnier d'une chronique peut se faire de quatre manières différentes.

- la méthode la plus simple consiste à ajuster la série avec X-11/12-ARIMA à la fin de chaque mois, quand les nouvelles données arrivent : il s'agit de l'ajustement synchrone direct ;

[1] Ladiray et Quenneville (1999).

- les facteurs d'ajustement saisonnier sont calculés uniquement tous les six mois, et les facteurs prévus sont employés pour ajuster les données actuelles : c'est l'ajustement direct prévu. Il est à noter que la prévision est d'abord effectuée sur la base d'une série finissant au mois de décembre de l'année précédente, et ensuite la seconde prévision de l'année est faite avec une chronique finissant en Juin ;

- une autre méthode consiste à séparer les parties bien définies d'une série principale (exemple : séparer les nouveaux entrants dans le marché de travail dans une série relative au chômage). Ces variables élémentaires sont ajustées séparément et la série principale sera simplement la somme de deux séries ajustées. Si cette procédure est effectuée chaque mois, la méthode est l'ajustement synchrone indirect ;

- la combinaison des approches 2 et 3 permet d'obtenir la méthode de l'ajustement indirect prévu.

Plusieurs études ont évalué l'impact de la méthode Census X-11 d'ajustement saisonnier.

Maravall (1995) montre que la variable résultante peut être décrite par un processus moyenne mobile non inversible rendant difficile la construction des modèles des séries temporelles uni et multivariées. Ghysels et Perron (1993) montrent que la puissance des tests de racine unité est plus faible pour les séries corrigées des variations saisonnières. En outre, Ghysels, Granger et Siklos (1995) révèlent que l'ajustement saisonnier peut conduire à des propriétés non linéaires indésirables dans la série chronologique. Ces études semblent toutes suggérer que l'ajustement saisonnier d'une chronique peut fausser certaines de ses propriétés importantes et peut aussi compliquer davantage l'analyse. D'autre part, un aspect commun de plusieurs phénomènes économiques est que la variation saisonnière peut changer à cause des changements climatiques ou du calendrier (e.g., le Pâques tombe parfois en Mars, parfois en Avril), mais aussi le changement de saisonnalité peut être causé par le comportement des agents économiques. Alternativement, ce comportement, peut être causé par des tendances ou des cycles perçus dans l'économie. En d'autres termes, il peut se produire que la tendance, la composante saisonnière et la composante cyclique soient difficiles à séparer, et par suite que l'identification de x_t^{ns} et x_t^{s} soit difficile. Un modèle économique théorique qui souligne cette complication est présenté dans Ghysels (1988). Si on ne peut pas éliminer la tendance et les saisons, il peut arriver que l'ajustement saisonnier enlève une information d'une série économique qui peut être utile pour décrire le comportement des agents économiques.

Ceci étant, on peut tirer les conclusions suivantes. La première est que la variation saisonnière peut constituer une grande partie de la variation totale dans les séries macroéconomiques. La seconde est que la saisonnalité semble varier au cours du temps. La troisième conclusion qu'on peut tirer est que la variation saisonnière et la variation non saisonnière peuvent ne pas être indépendantes pour certaines variables économiques. En effet, il y a une évidence d'une relation entre les périodes de saisonnalité et de cycle économique et/ou la tendance. Ce résultat met en cause l'utilité des méthodes d'ajustement saisonnier du moment que l'indépendance des fluctuations saisonnières et non saisonnières est une hypothèse clé.

Ultérieurement, nous allons étudier certaines approches qui incorporent explicitement la variation saisonnière dans le cadre d'un modèle économétrique. On trouve que, même si ces modèles économétriques peuvent ne pas être aussi parcimonieux que les modèles pour les données corrigées des variations saisonnières, il est important au moins de considérer de tels modèles dans des occasions pratiques. Enfin, un avantage important de la modélisation de la saisonnalité au lieu de son ajustement, est qu'en analysant les mêmes données, on peut évaluer les mérites de certains modèles. En effet, la plupart des approches comptent sur des tests statistiques formels. Ainsi, les tests emboîtés et non-emboîtés peuvent être utilisés pour comparer les modèles variés. Ceci est dans un contraste net à la pratique courante de l'ajustement saisonnier où cette pratique est habituellement considérée comme étant un processus boite noire.

Il est difficile d'obtenir une connaissance précise sur la manière avec laquelle les agences statistiques génèrent leurs séries temporelles ajustées, e.g., comment elles traitent les changements en moyenne et tendances.

2.1.1.2 Désaisonnalisation par la méthode du ratio aux moyennes mobiles[1]

Soit une série (X_t) saisonnière. On cherche à la transformer de manière à annuler sa composante saisonnière. Pour cela, on doit trouver un filtre, i.e. une transformation mathématique de la série chronologique. Le filtre le plus utilisé pour corriger une série des variations saisonnières est le filtre de la moyenne mobile.

Une moyenne mobile peut être définie comme étant un opérateur linéaire, combinaison linéaire d'opérateurs retard qui peut s'écrire de la manière suivante :

$$M = L^{m_1} \sum_{i=0}^{m_1+m_2} \theta_{i-m_1} L^{-i} = L^{m_1} \sum_{i=0}^{m_1+m_2} \theta_{i-m_1} F^{i} = L^{m_1} \Theta(F)$$

Où $\Theta(.)$ est un polynôme appelé polynôme caractéristique de *M*, de degré $m_1 + m_2$, et $m_1 + m_2 + 1$ représentera l'ordre de *M*, i.e. le nombre théorique de termes de *M*.

Si $m_1 = m_2 = m$, on dit que la moyenne mobile est centrée. De surcroît, si *M* est centrée et que $\theta_i = \theta_{-i}$ pour tout *i*, alors la moyenne mobile est dite symétrique.

D'autre part, il est à noter que les moyennes mobiles centrées symétriques sont nécessairement d'ordre impair (pour qu'elles soient centrées). Ainsi, pour *m* impair, on considère les moyennes mobiles d'ordre $m = 2p + 1$ définie par :

$$M_m(X_t) = \frac{1}{m}\left[X_{t-p} + X_{t-p+1} + \ldots + X_{t-1} + X_t + X_{t+1} + \ldots + X_{t+p-1} + X_{t+p} \right]$$

Par exemple, la moyenne mobile d'ordre 3 (ayant pour coefficients : 1/3, 1/3, 1/3) est telle que : $M_3(X_t) = \frac{1}{3}\left[X_{t-1} + X_t + X_{t+1} \right]$.

Cependant, on peut construire des moyennes mobiles centrées et symétriques d'ordre pair d'une manière artificielle. Pour ce faire, si $m = 2p$, les moyennes mobiles seront définies par : $M_m(X_t) = \frac{1}{m}\left[X_{t-p+1/2} + \ldots + X_{t-1/2} + X_{t+1/2} + \ldots + X_{t+p-1/2} \right]$

Avec $(X_{t-1/2})$ étant une valeur intermédiaire entre X_{t-1} et X_t, ce qui revient à écrire :

[1] Les définitions et les formules de cette section sont inspirées du manuel de Arthur Charpentier (2003).

$$M_{2\times p}=\frac{1}{2m}\left[X_{t-p}+2X_{t-p+1}+\ldots+2X_{t-1}+2X_{t}+2X_{t+1}+\ldots+2X_{t+p-1}+X_{t+p}\right]$$

La moyenne mobile 2×12 permet d'estimer des tendances dans le cas de données mensuelles. Elle est d'ordre 13 et de coefficients 1/24, 1/12, 1/12, …, 1/12, 1/24 :

$$M_{2\times 12}(X_t)=\frac{1}{24}\left[X_{t-6}+2X_{t-5}+2X_{t-4}+\ldots+2X_{t+5}+X_{t+6}\right]$$

Dans A. Charpentier (2003), on révèle que cette moyenne mobile élimine les saisonnalités mensuelles des séries mensuelles, conserve les tendances linéaires et réduit de plus de 90% la variance d'un bruit blanc.

2.1.2 Approches paramétriques de désaisonnalisation

2.1.2.1 Désaisonnalisation par la méthode de régression

Cette approche se base sur le modèle de Buys-Ballot (1847) qui consiste à effectuer la régression ci-dessous, en utilisant des variables indicatrices saisonnières ($S_{t,i}$) tel que $S_{t,i}$ prend la valeur 1 si t correspond à la période saisonnière, et 0 sinon. Le modèle s'écrit comme suit : $X_t=\beta_0+\beta_1 t+\sum_{i=1}^{T-1}\gamma_i S_{t,i}+\varepsilon_t$.

Avec : T étant la période de la saisonnalité (T = 4 pour une chronique trimestrielle, T = 12 pour des données mensuelles). L'utilisation de (T – 1) variables indicatrices uniquement permet d'éviter le problème de colinéarité qui pourrait exister avec le vecteur unité relatif à la constante[1]. On estime ainsi (T – 1) coefficients de saisonnalité et on vérifie le $T^{\text{ième}}$ à l'aide du principe de conservation des aires[2] : $\sum_{i=1}^{T}\gamma_i=0$.

[1] On peut aussi considérer T variables dichotomiques dans le modèle et enlever la constante.
[2] R. Bourbonnais & M. Terraza (2008).

2.1.2.2 Modèles saisonniers et tests de racines unitaires saisonnières

Comme c'est expliqué précédemment, plusieurs séries chronologiques ont un « profil saisonnier » accentué, c'est-à-dire que les données relatives à un mois (ou trimestre) de différentes années ont tendance à se situer de façon analogue par rapport à la moyenne annuelle. C'est ce qui a permis de penser qu'il serait intéressant, dans un modèle ARIMA, de faire intervenir des décalages multiples de 12 (pour des données mensuelles par exemple).

En théorie, rien ne nous interdit de prendre des valeurs de p et q suffisamment grandes pour que ces décalages soient pris en considération. Toutefois, il est clair que cela engendrerait une augmentation importante du nombre de paramètres et qu'il deviendrait très difficile de les estimer, ce qui aboutirait donc à un modèle difficile à interpréter.

Pour contourner ce problème « d'inflation » de paramètres, Box et Jenkins (1970) ont proposé un type particulier de modèles saisonniers : il s'agit du modèle multiplicatif ARIMA saisonnier ou SARIMA(p,d,q)(P,D,Q)$_s$, qui s'écrit sous cette forme :

$$\phi_p(L)\Phi_P(L^s)\Delta^d\Delta_s^D y_t = \theta_q(L)\Theta_Q(L^s)\varepsilon_t \qquad (2)$$

Où : s est la période de saisonnalité (s = 12 pour des séries mensuelles, s = 4 pour des séries trimestrielles, ..., etc) ; $\Delta = 1 - L$, $\Delta_s = 1 - L^s$, $\phi_p, \Phi_P, \theta_q, \Theta_Q$ sont des polynômes de degrés : p, P, q, Q et dont les racines sont de module supérieur à 1 ; (ε_t) est un bruit blanc ; d et D sont respectivement les ordres de différenciation non saisonnière et saisonnière.

Ces modèles supposent l'existence de racines unitaires saisonnières dans la chronique et cela peut être vérifié à l'aide de tests appropriés.

i) Autocorrélations

La saisonnalité peut être détectée graphiquement en examinant les fonctions d'autocorrélation (ACF) et d'autocorrélation partielle (PACF) nécessaires pour l'identification des modèles ARIMA appropriés.

ii) Tests de racines unitaires saisonnières

La détection de la saisonnalité peut se faire à l'aide de tests de racines unitaires saisonnières. A cette fin, Un certain nombre de tests ont été mis en œuvre dans les années 80 et 90, en particulier pour tester la saisonnalité à l'ordre 4 et à l'ordre 12.

a) Tests de Hasza et Fuller (1982) et de Osborn, Chui, Smith et Birchenhall (OCSB, 1988) :

Hasza et Fuller ont considéré le modèle suivant :

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-s} + \alpha_3 y_{t-s-1} + \varepsilon_t \tag{3}$$

Où (ε_t) est un bruit blanc. L'hypothèse nulle est H_0 : $\alpha_1 = \alpha_2 = -\alpha_3 = 1$.

Dans le test (OCSB) qui est une extension à cette approche, on a pris le modèle suivant :

$$\Theta(L)(1-L)(1-L^s)y_t = \sum_{i=1}^{s-1} \delta_i D_{it} + \lambda(1-L^s)y_{t-1} + \mu(1-L)y_{t-s} + \varepsilon_t \tag{4}$$

Si l'on accepte l'hypothèse μ = 0, la différence à l'ordre s est appropriée ; en revanche, si λ = 0, alors on maintient le filtre de différenciation première. Enfin, si $\lambda = \mu = 0$, alors le filtre $(1-L)(1-L^s)$ est nécessaire.

Par ailleurs, la littérature sur la notion de « racines unités » (e.g., Dickey, Bell et Miller (1986)) montre que l'hypothèse d'existence de certains filtres de différenciation revient à émettre l'hypothèse de présence d'un certain nombre de racines unités saisonnières et non saisonnières dans une série chronologique. Ceci peut être facilement vu en écrivant : $\Delta_s = (1-L^s)$, et en résolvant l'équation : $(1-z^s) = 0$.

La solution générale à cette équation est : $\{1, \cos(2\pi k/S) + i\sin(2\pi k/S)\}$; avec le terme $(2\pi k/s)$ pour k = 1,2,..., représente la fréquence saisonnière correspondante, donnant S solutions différentes qui se trouvent toutes sur le cercle unité. Par exemple, dans le cas où S = 4, les solutions à $(1-z^4) = 0$ sont $\{1, i, -1, -i\}$.

Ceci signifie qu'on peut écrire :

$$
\begin{aligned}
\Delta_4 = (1-L^4) &= (1-L)(1+L)(1-iL)(1+iL) \\
&= (1-L)(1+L)(1+L^2) \\
&= (1-L)(1+L+L^2+L^3)
\end{aligned}
$$

La racine unitaire « 1 » est appelée la racine unitaire « non saisonnière », alors que les racines unités (-1) et ($\pm i$) sont les racines unitaires « saisonnières » (Hylleberg et al. (1990)).

Tout filtre de différenciation (Δ_s) peut s'écrire tel que :

$\Delta_s = (1-L)(1+L+\ldots+L^{S-1})$, et tout (Δ_s) peut ainsi être décomposé en une partie avec racine unité non-saisonnière et une partie avec (S – 1) racines unitaires saisonnières. D'où le test suivant.

b) Test de Hylleberg, Engle, Granger et Yoo ((HEGY), 1990) :

Hylleberg et al.(1990) [HEGY] propose une méthode pour tester la présence des racines unitaires saisonnières et non saisonnières dans des séries trimestrielles. Cette méthode est une extension à l'approche de Dickey, Hasza et Fuller (1984) qui considère le test de l'hypothèse $\rho = 1$ dans $x_t = \rho x_{t-4} + \varepsilon_t$.

La méthode HEGY, principalement conçue pour les séries trimestrielles, a été adaptée au cas mensuel grâce aux travaux de Franses (1990) et de Beaulieu et Miron (1993).

Dans ce cas où s = 12, les solutions de l'équation $(1-z^{12}) = 0$ sont : « 1 » pour la racine unitaire non saisonnière correspondante à la fréquence 0 ; et les 11 racines unitaires saisonnières $\left\{-1, \pm i, -\frac{1}{2}(1\pm\sqrt{3}i), \frac{1}{2}(1\pm\sqrt{3}i), -\frac{1}{2}(\sqrt{3}\pm i), \frac{1}{2}(\sqrt{3}\pm i)\right\}$ correspondant respectivement aux fréquences suivantes : $\left\{\pi, \pm\frac{\pi}{2}, \pm\frac{2\pi}{3}, \pm\frac{\pi}{3}, \pm\frac{5\pi}{6}, \pm\frac{\pi}{6}\right\}$ et aux opérateurs de différenciation : $(1+L)$, $(1+L^2)$, $(1+L+L^2)$, $(1-L+L^2)$, $(1+\sqrt{3}L+L^2)$ et $(1-\sqrt{3}L+L^2)$.

Dans ce test, on recourt à la décomposition des polynômes $(1\text{-}L^4)$ et $(1\text{-}L^{12})$, avec respectivement 4 et 12 racines unités. Dans le cas d'une saisonnalité d'ordre s = 12, on considère la forme suivante :

$$\Theta(L)z_{8t} = \mu_t + \pi_1 z_{1,t-1} + \pi_2 z_{2,t-1} + \pi_3 z_{3,t-1} + \pi_4 z_{3,t-2} + \pi_5 z_{4,t-1} + \pi_6 z_{4,t-2} + \pi_7 z_{5,t-1} + \pi_8 z_{5,t-2}$$
$$+\pi_9 z_{6,t-1} + \pi_{10} z_{6,t-2} + \pi_{11} z_{7,t-1} + \pi_{12} z_{7,t-2} + \varepsilon_t \qquad (5)$$

Les variables z_{it} sont telles que : $z_{it} = P_i(L)y_t$, où les polynômes P_i sont définis comme suit[1] :

$$\begin{cases} P_1(L) = (1+L)(1+L^2)(1+L^4+L^8) \\ P_2(L) = -(1-L)(1+L^2)(1+L^4+L^8) \\ P_3(L) = -(1-L^2)(1+L^4+L^8) \\ P_4(L) = -(1-L^4)(1-\sqrt{3}L+L^2)(1+L^2+L^4) \\ P_5(L) = -(1-L^4)(1+\sqrt{3}L+L^2)(1+L^2+L^4) \\ P_6(L) = -(1-L^4)(1-L^2+L^4)(1-L+L^2) \\ P_7(L) = -(1-L^4)(1-L^2+L^4)(1+L+L^2) \\ P_8(L) = (1-L^{12}) \end{cases}$$

Avec aussi : $\Theta(L)$ est un polynôme autorégressif en L, et μ_t peut contenir une constante, 11 variables dummy saisonnières et/ou une tendance.
Les variables z_{it} sont alors associées aux différentes racines du polynôme. L'équation (5) est estimée par la méthode des moindres carrés ordinaires.
On peut donc effectuer des tests de type « t » pour les paramètres π_1 et π_2, ainsi que des tests de type « F » associés aux couples $(\pi_3,\pi_4),(\pi_5,\pi_6),(\pi_7,\pi_8),(\pi_9,\pi_{10})$ et (π_{11},π_{12}) : il s'agit de tester la significativité jointe des coefficients. Cela revient à tester l'hypothèse d'existence de racines unitaires aux différentes fréquences. Pour ce faire, on doit comparer les statistiques de test rattachées aux paramètres estimés aux valeurs critiques fournies par Franses (1990) et Beaulieu et Miron (1993).

Pour vérifier l'existence des racines « 1 » et « -1 » correspondantes respectivement aux fréquences 0 et π, on effectue deux tests individuels sur chacun des paramètres π_1 et π_2. Quant aux autres racines unitaires saisonnières, on peut effectuer soit des tests joints dont l'hypothèse nulle prend la forme $\pi_k = \pi_{k-1} = 0$ et ce pour les valeurs paires de k, de 4 à 12 ; soit, tout simplement, des

[1] Arthur Charpentier (2003).

tests individuels, suggérés dans Franses (1990), permettant de vérifier la non stationnarité de la chronique à toutes les fréquences saisonnières et ce en testant l'hypothèse nulle selon laquelle il existe une racine unitaire saisonnière ($\pi_k = 0 \, , k \in [3,12]$). Cependant, il convient de noter que l'application des MCO sur la régression (5) se fait où l'ordre de $\Theta(L)$ est déterminé tel que les erreurs sont approximativement des bruits blancs ou au moins des résidus non autocorrélés. A cette fin, HEGY (1990) et EGHL (1993) proposent d'introduire des retards supplémentaires de la variable jusqu'à ce qu'on obtienne des résidus non autocorrélés.

Pour monter qu'il n'existe pas de racines unitaires à toutes les fréquences saisonnières, il faut que π_k soit différent de zéro pour k = 2 et pour au moins un membre de chacun de ces couples : $(\pi_3, \pi_4), (\pi_5, \pi_6), (\pi_7, \pi_8), (\pi_9, \pi_{10})$ et (π_{11}, π_{12}) (Franses, 1991 ; Beaulieu et Miron, 1993).

Compte tenu de cela, on peut conclure que l'utilisation du filtre de $(1 - L^{12})$, suggéré par Box et Jenkins (1970), pour stationnariser les séries mensuelles saisonnières, dépend du fait que la variable est non stationnaire à la fréquence 0 et à toutes les fréquences saisonnières (Pichery et Ouerfelli, 1998).

2.1.2.3 Désaisonnalisation par la méthode TRAMO-SEATS

Le programme TRAMO-SEATS (Gomez & Maravall, 1997) fait partie des méthodes paramétriques de désaisonnalisation fondées sur l'extraction du signal. Il est composé de deux sous-programmes indépendants mais qui sont complémentaires du moment qu'ils sont le plus souvent utilisés ensemble :

- le programme TRAMO (**T**ime series **R**egression with **A**RIMA noise, **M**issing observations and **O**utliers) s'inscrit dans la même optique que la modélisation ARIMA, ou plus exactement, il s'agit d'une extension à ces modèles. Son principe est en effet de modéliser la série brute à l'aide de l'approche univariée de Box & Jenkins via les modèles ARIMA ou ARIMA saisonnier (SARIMA), en détectant, estimant et corrigeant au préalable les points atypiques, les valeurs manquantes, les effets calendaires (vacances, jours fériés…) ainsi que les changements de régime, susceptible de perturber l'estimation des coefficients du modèle.

- le programme SEATS (**S**ignal **E**xtraction **A**RIMA **T**ime **S**eries) vient compléter la procédure TRAMO par la décomposition de la série initiale ainsi modélisée en ses composantes (la tendance, le cycle, l'irrégulier et la saisonnalité) par l'extraction du signal, et ce en utilisant l'analyse spectrale de la série initiale.