

REVUE DE STATISTIQUE APPLIQUÉE

J.-L. MARCHAL

B. SEYS

B. SPAGO

P. HOUBEN

L. VANSNICK

Agrégation de prévisions mensuelles de consommations médicamenteuses à l'aide d'un modèle ARIMA

Revue de statistique appliquée, tome 53, n° 2 (2005), p. 5-28

http://www.numdam.org/item?id=RSA_2005__53_2_5_0

© Société française de statistique, 2005, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

AGRÉGATION DE PRÉVISIONS MENSUELLES DE CONSOMMATIONS MÉDICAMENTEUSES À L'AIDE D'UN MODÈLE ARIMA

J.-L. MARCHAL, B. SEYS, B. SPAGO, P. HOUBEN, L. VANSNICK

*Institut Pharmaco-Epidémiologique Belge,
Rue Archimède 11, 1000 Bruxelles*

RÉSUMÉ

Comment obtenir des prévisions de consommation médicamenteuse suffisamment précises à partir d'une série chronologique limitée de données mensuelles présentant une tendance croissante et des fluctuations périodiques? Comment estimer des prévisions agrégées en trimestres et semestres, et leurs intervalles de confiance, qui sont plus commodes à utiliser? Une même méthode permet-elle d'analyser les variables de volume et de coût caractérisant la consommation? Un modèle ARIMA $(2, 1, 3)(1, 0, 0)_{12}$, dont les deux premiers paramètres de moyenne mobile sont nuls, est ajusté à la série mensuelle par la méthode du maximum de vraisemblance. Il permet d'obtenir des prévisions mensuelles. Sans nouvel ajustement, les prédictions trimestrielles et semestrielle sont obtenues en agrégeant les prévisions mensuelles. L'addition des erreurs de prévision mensuelles, obtenues par transformation du modèle ARIMA sous sa forme en moyenne mobile d'ordre infini ($MA(\infty)$), permet la détermination des erreurs de prévision trimestrielle et semestrielle.

Les séries (84 observations, années 1996 à 2002) concernent la consommation totale des spécialités remboursées en Belgique. L'agrégation augmente la stabilité des observations et des prévisions et améliore la précision : les écarts-types des prévisions mensuelles, trimestrielles et semestrielles sont respectivement, selon les variables analysées, de l'ordre de 3 à 4 %, de 2 % et de 1 à 2 % de la valeur prédite correspondante. Entre les horizons 2 et 6, l'écart-type de prévision mensuelle augmente de 3 à 5 % par rapport à l'écart-type résiduel. Les intervalles de prévision sont d'une étendue acceptable pour l'objectif fixé.

Mots-clés : *Agrégation, consommation médicamenteuse, dose journalière définie (DDD), modèle ARIMA, pharmaco-épidémiologie, prévision, série chronologique.*

ABSTRACT

How to achieve reliable drug consumption forecasts from a limited monthly time series characterized by a growing trend and periodic fluctuations? How to obtain quarterly and half-yearly aggregated forecasts, and their confidence limits, which are more convenient? Could one unique method be found to process volume and cost variables characterizing the consumption? An ARIMA $(2, 1, 3)(1, 0, 0)_{12}$ model, where the first two moving average parameters are null, is fitted on the monthly time series following the maximum likelihood method. It gives monthly predictions. Quarterly and half-yearly forecasts are obtained without new adjustment by aggregating monthly predictions. Adding the monthly forecast errors,

obtained by transformation of the ARIMA model into its moving average representation of infinite order ($MA(\infty)$), gives rise to the quarterly and half-yearly forecast errors.

The time series (84 observations, years 1996 to 2002) represent the total consumption of reimbursed drugs in Belgium. The aggregation enhances the steadiness of the data and forecasts, and improves the precision : the standard deviations of the monthly, quarterly and half-yearly predictions had respectively, according to the variable studied, a magnitude order of 3 to 4 % , 2 % and 1 to 2 % of the corresponding forecasts. From 2- till 6-step ahead, the monthly forecast standard deviation raised from 3 to 5 % with respect to the residual standard deviation. Forecast intervals are of acceptable length with respect to the objectives.

Keywords : *Aggregation, ARIMA model, defined daily dose (DDD), drug consumption, forecast, pharmaco-epidemiology, time series.*

1. Introduction

L'Institut Pharmaco-Épidémiologique Belge (IPhEB) exploite depuis 1992 la banque de données IFSTAT[©], qui enregistre la consommation de spécialités remboursées en Belgique. Elle collecte des données de délivrance en pharmacie, transmises à des offices de tarification qui collaborent avec l'Association Pharmaceutique Belge (APB), union professionnelle majoritaire des pharmaciens en Belgique. La tarification consiste à facturer aux organismes assureurs la partie du prix du médicament à charge de la collectivité. Ceux-ci se font rembourser par la suite auprès de l'organisme fédéral de sécurité sociale, l'Institut National d'Assurance Maladie-Invalidité (INAMI), qui dispose lui-même d'une banque de données nationales sur la consommation médicamenteuse, appelée Pharmanet (INAMI, 1999).

En 2002, la population-univers de la collecte des données comprend 9,2 millions de personnes, soit 88,9 % de la population belge. Il s'agit des bénéficiaires du remboursement des médicaments par l'INAMI, à l'exception des affiliés de la Société Nationale des Chemins de Fer Belges (142 218 personnes en 2002). IFSTAT[©] collecte les données provenant de 80,1 % des officines belges ouvertes au public (de 61,9 % à 92,0 % selon les provinces). L'exhaustivité de la collecte, mesurée annuellement par le rapport entre le coût pour l'INAMI des médicaments calculé dans IFSTAT[©] et les dépenses de l'INAMI, augmente régulièrement de 1996 (74,74 %) à 2001 (78,44 %). La valeur de cette année, la dernière connue, est provisoirement utilisée en 2002. Les données sont redressées par l'exhaustivité pour estimer la consommation dans la population-univers.

Elles ont été validées à partir de janvier 1996, en s'appuyant principalement sur une étude de concordance (SEYS *et al.*, 1999) avec les autres sources de consommation disponibles en Belgique, Pharmanet et International Medical Statistics (IMS). Cette étude mettait en évidence l'exhaustivité importante, la stabilité et la fiabilité du processus de collecte; elle démontrait qu'un biais de sélection est possible, quoique les discordances avec les données officielles soient faibles au niveau national. Les données sont fiables à partir d'une délivrance supérieure ou égale à 60 boîtes par mois. Elles ont permis de réaliser des études sur les antibiotiques (VANDER STICHELE *et al.*, 2001; BAURAIND *et al.*, 2001) et sur l'exposition à certains médicaments en Belgique (HOUBEN *et al.*, 2001).

Les données, qui ne contiennent aucune information individuelle, sont obtenues sur une base mensuelle. Elles sont disponibles quatre mois après la délivrance. Leur traitement suit la méthodologie ATC/DDD¹, préconisée par l'Organisation Mondiale de la Santé (World Health Organization, 2001; SEYS *et al.*, 2001). Les informations sont mises à jour chaque trimestre. Dans ce système, les volumes sont représentés par le nombre d'unités standardisées de principe actif, plutôt qu'en nombre de boîtes.

Les observations mensuelles sont très variables, ce qui se reflète aussi au niveau des prévisions obtenues par un modèle ARIMA². L'agrégation des données mensuelles, qui permet d'obtenir finalement des prévisions trimestrielles et semestrielles plus stables, rencontre l'intérêt de l'IPhEB.

L'objectif est de trouver un modèle ARIMA de type général qui s'adapte aux différentes variables et d'obtenir par agrégation des prévisions stables et précises à partir d'un nombre limité de données mensuelles.

Pour parvenir à ce but, deux approches théoriques sont possibles. La première consiste à d'abord agréger par sommation les observations mensuelles en une série trimestrielle (ou semestrielle) et à ensuite ajuster à cette série dérivée un modèle. Dans la seconde approche, le modèle choisi est d'abord ajusté aux observations mensuelles et les prévisions obtenues sont ensuite agrégées. La première étape fournit l'estimation des paramètres du modèle ARIMA et de la variance résiduelle. À l'aide de ces derniers, il est ensuite possible de calculer les valeurs prédites agrégées et leur intervalle de prévision.

Le modèle ARIMA choisi dans la première approche dérive de celui sélectionné dans la seconde, mais leur ordre et leur variance résiduelle seront différents (PINO *et al.*, 1987; ROSSANA et SEATER, 1995). Dans certaines circonstances, le modèle réellement estimé sur la série agrégée est différent de celui que la théorie prévoit (ABRAHAM, 1982).

Vu le nombre limité de données mensuelles disponibles, seule la seconde approche, qui est aussi la plus précise, est réalisable. L'agrégation préalable éliminerait les variations de basse fréquence (ROSSANA et SEATER, 1995). Elle entraînerait dans le cas présent l'apparition de nouveaux termes de moyenne mobile. Pour les séries courtes, la perte d'observations occasionnerait une augmentation de la variance résiduelle.

La présentation des données (Point 2) sera suivie de l'évocation des étapes (Point 3) qui conduisent au choix du modèle ARIMA (DROESBEKE, FICHET et TASSI, 1989; DIGGLE, 1990; MELARD, 1990; WEI, 1994). L'ajustement de ce dernier aux données mensuelles permet l'obtention des prévisions mensuelles, ainsi que de l'erreur et de l'intervalle qui y sont associés (Point 4). Les notions utiles à l'agrégation des séries chronologiques sont rappelées dans un point suivant, avant de présenter l'application (Point 6) et d'envisager finalement la discussion (Point 7).

¹ Anatomical, Therapeutic and Chemical Classification / Defined Daily Dose

² AutoRegressive Integrated Moving Average

2. Présentation des données

Les informations de base sont fournies sous la forme d'un fichier qui contient une identification du mois (à partir du 1^{er} janvier 1996 jusqu'au dernier mois du trimestre disponible), le prix public (PP), le coût remboursé (CI) et le nombre d'unités de consommation (DDD). En tant que variable dérivée, le coût à charge du patient (CP) se calcule de la façon suivante :

$$CP = PP - CI.$$

À l'exception du nombre d'unités de consommation, les variables s'expriment en unité monétaire (€), en coûts courants.

Périodiquement, les données sont sujettes à révision en fonction des modifications tarifaires et des références en DDD, susceptibles d'influencer la série chronologique. Les médicaments qui ne furent plus remboursés depuis février 1997 sont exclus des données relatives à l'année 1996 et au mois de janvier 1997.

Par leur tendance croissante, les séries considérées indiquent que les consommations de médicaments augmentent dans le temps. Les oscillations périodiques, traduisant des hausses ou des baisses à intervalles réguliers, illustrent le caractère fluctuant des séries chronologiques de consommation médicamenteuse. La Figure 1 trace l'évolution mensuelle du nombre d'unités de consommation, du coût à charge du patient et du coût remboursé, de janvier 1996 à décembre 2002 (84 mois).

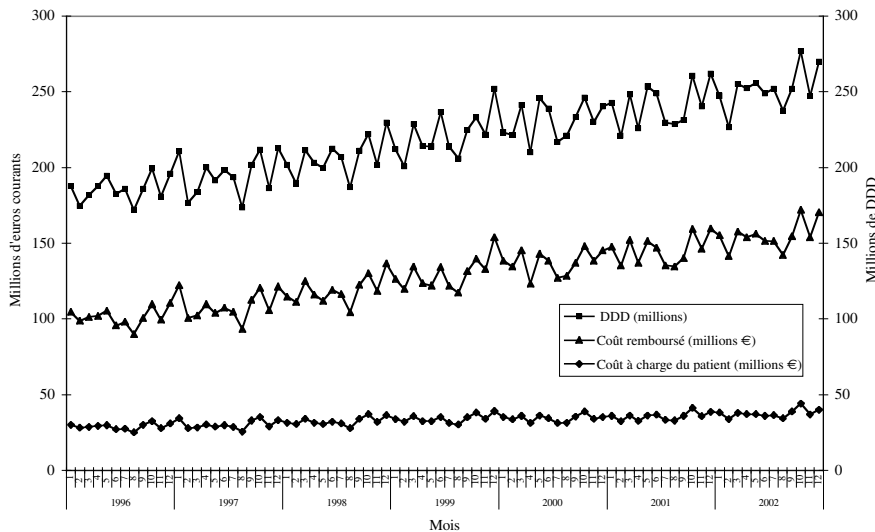


FIGURE 1

Série chronologique mensuelle observée du nombre d'unités de consommation (DDD), du coût à charge du patient et du coût remboursé des spécialités pharmaceutiques en Belgique, de janvier 1996 à décembre 2002

Les fluctuations périodiques ont tendance à se maintenir dans les mêmes limites, au cours du temps. La variance de la série est donc stable, avec une absence de valeurs extrêmes ou aberrantes.

3. Choix du modèle ARIMA

Parmi les divers modèles susceptibles de s'ajuster à une série chronologique, nous nous sommes orientés vers les modèles ARIMA. Dans ce contexte, la méthode de Box et Jenkins (BOX et JENKINS, 1976) vise, par une approche structurée, à choisir le modèle ARIMA le plus approprié, grâce auquel les prévisions de la série seront établies.

La première étape consiste à la rendre stationnaire. Sa variance stable indique qu'une transformation des données n'est pas nécessaire. La tendance croissante peut être annulée par une différentiation première, dont la nécessité est confirmée par le rejet du test de Dickey-Fuller (DICKEY *et al.*, 1984, 1986), appliqué au retard 1.

Si X_t représente la variable aléatoire associée à l'instant t , la différence première ordinaire s'exprimera de la façon suivante :

$$Y_t = X_t - X_{t-1} = (1 - B)X_t,$$

où B représente l'opérateur de retard.

Cette nouvelle série Y_t , de moyenne μ , compte une valeur de moins que la série de départ X_t .

La seconde étape consiste à préciser le choix du modèle ARIMA sur la base des autocorrélations simples et partielles, calculées sur la variable différenciée. Dans le cas présent, l'autocorrélation partielle la plus élevée de retard 2 nous a orientés vers un modèle autorégressif d'ordre 2. Lors des premiers essais, une transformation saisonnière fixe était appliquée aux données. Suite à une lecture critique, un terme autorégressif d'ordre 12 permet de prendre en compte les fluctuations saisonnières de manière souple et performante. Il faut encore ajouter un élément de moyenne mobile d'ordre 3, afin que la distribution du résidu (valeur observée – valeur prédite), obtenu après ajustement du modèle, puisse être considérée comme un bruit blanc, constitué de variables indépendantes entre elles et de moyenne nulle.

Le modèle ajusté, la critique de l'ajustement s'opère à partir des résidus et de leurs autocorrélations simples, par différents moyens. Par le test individuel de l'autocorrélation, l'hypothèse d'un bruit blanc ne sera pas rejetée pour un retard déterminé, si l'autocorrélation simple du résidu se situe à l'intérieur de ses limites de confiance. Le test global du bruit blanc (test de Ljung et Box) permet d'éprouver l'hypothèse d'un bruit blanc sur les premiers retards, sur base des autocorrélations simples. Enfin, le graphique des résidus ne doit plus montrer aucune structure dans le temps.

Ce processus peut exiger des retours en arrière; lorsqu'une étape n'est pas jugée satisfaisante, il faut reprendre la démarche à une phase antérieure, jusqu'à ce que les résultats obtenus soient valables. Lorsque le modèle est clairement validé, il peut donner lieu à des prévisions et à une interprétation.

Pour les données récoltées par l'IPhEB, l'ajustement d'un modèle ARIMA $(2, 1, 3)(1, 0, 0)_{12}$, dont les 2 premiers paramètres de moyenne mobile sont nuls, donne un résultat adéquat pour le nombre d'unités de consommation (DDD), considéré en premier lieu, et également pour le coût remboursé (CI) et celui à charge du patient (CP), analysés par la suite. Le modèle, estimé sur les données X_t de consommation médicamenteuse, est de la forme :

$$(1 - \phi_1 B - \phi_2 B^2)(1 - \phi_{12} B^{12})(Y_t - \mu) = (1 - \theta_3 B^3)\varepsilon_t. \quad (1)$$

B est l'opérateur de retard déjà rencontré, ε_t , un bruit blanc de moyenne nulle et de variance σ_ε^2 , égale à la variance résiduelle du modèle. Ce dernier comprend 5 paramètres. μ est la moyenne de la série différenciée Y_t . ϕ_1 , ϕ_2 et ϕ_{12} sont les paramètres autorégressifs d'ordre 1, 2 et 12, tandis que θ_3 est le paramètre de moyenne mobile d'ordre 3.

L'estimation des paramètres du modèle ARIMA est effectuée par la méthode du maximum de vraisemblance. Cette dernière fournit également l'écart-type de la valeur estimée du paramètre. Le test t de Student éprouve de façon approchée la signification de chacun des 5 paramètres du modèle, avec un nombre de degrés de liberté égal au nombre de résidus (nombre d'observations disponibles après différenciation) moins le nombre de paramètres estimés.

4. Prévisions mensuelles

Vu la longueur actuelle de la série (84 mois), il s'est avéré opportun de prévoir, après la dernière donnée disponible, 6 valeurs mensuelles, accompagnées de leur intervalle de prévision.

4.1. Valeur mensuelle prédite

Lorsqu'elles sont estimées au sens des moindres carrés ou par la méthode du maximum de vraisemblance avec l'hypothèse de normalité du bruit blanc, les prévisions $\hat{X}_T(h)$, réalisées à partir de T observations disponibles et pour l'horizon de prévision h , sont égales à l'espérance mathématique conditionnelle de X_{T+h} , connaissant les T valeurs actuellement observées :

$$\hat{X}_T(h) = E(X_{T+h} | X_T, X_{T-1}, X_{T-2}, \dots).$$

Le développement du modèle (1) s'écrit à l'horizon de prévision h , c'est-à-dire au temps $T + h$:

$$\begin{aligned} X_{T+h} = & (1 - \phi_1 - \phi_2 - \phi_{12} + \phi_1\phi_{12} + \phi_2\phi_{12})\mu + \\ & (1 + \phi_1)X_{T+h-1} + (\phi_2 - \phi_1)X_{T+h-2} - \phi_2 X_{T+h-3} + \\ & \phi_{12}X_{T+h-12} - (1 + \phi_1)\phi_{12}X_{T+h-13} + \\ & (\phi_1 - \phi_2)\phi_{12}X_{T+h-14} + \phi_2\phi_{12}X_{T+h-15} + \varepsilon_{T+h} - \theta_3\varepsilon_{T+h-3}. \end{aligned}$$

Dans la formule de la prévision de $\widehat{X}_T(h)$, l'espérance mathématique d'un bruit blanc futur est nulle, tandis que celle d'un bruit blanc passé ou présent est égale au résidu correspondant :

$$\begin{aligned} E(\varepsilon_{T+j} | X_T, X_{T-1}, X_{T-2}, \dots) &= \varepsilon_{T+j}, \quad \text{si } j \leq 0 \\ &= 0, \quad \text{si } j > 0. \end{aligned} \quad (2)$$

Les valeurs présente et passées sont remplacées par les observations et les valeurs futures, par leur prédiction la plus récente. Ainsi, la première valeur prédite, $\widehat{X}_T(1)$, s'écrit exclusivement en fonction des valeurs présente et passées de la série :

$$\begin{aligned} \widehat{X}_T(1) &= (1 - \phi_1 - \phi_2 - \phi_{12} + \phi_1\phi_{12} + \phi_2\phi_{12})\mu + \\ &\quad (1 + \phi_1)X_T + (\phi_2 - \phi_1)X_{T-1} - \phi_2X_{T-2} + \\ &\quad \phi_{12}X_{T-11} - (1 + \phi_1)\phi_{12}X_{T-12} + \\ &\quad (\phi_1 - \phi_2)\phi_{12}X_{T-13} + \phi_2\phi_{12}X_{T-14} - \theta_3\widehat{\varepsilon}_{T-2}. \end{aligned}$$

Les valeurs futures de la série chronologique interviennent à partir de la prévision d'horizon 2. Elles sont remplacées par les prévisions les plus récentes calculées au départ des T observations :

$$\begin{aligned} \widehat{X}_T(2) &= (1 - \phi_1 - \phi_2 - \phi_{12} + \phi_1\phi_{12} + \phi_2\phi_{12})\mu + \\ &\quad (1 + \phi_1)\widehat{X}_T(1) + (\phi_2 - \phi_1)X_T - \phi_2X_{T-1} + \\ &\quad \phi_{12}X_{T-10} - (1 + \phi_1)\phi_{12}X_{T-11} + \\ &\quad (\phi_1 - \phi_2)\phi_{12}X_{T-12} + \phi_2\phi_{12}X_{T-13} - \theta_3\widehat{\varepsilon}_{T-1}. \end{aligned}$$

À partir de l'horizon 4, le terme de moyenne mobile disparaît car la variable de bruit blanc à laquelle il est associé se situe dans le futur.

La valeur attendue, $\widehat{X}_{T-u}(1)$, correspondant au présent ou au passé de la série, est l'espérance mathématique de X_{T-u+1} , tout en connaissant les $T - u$ premières valeurs. Dans ce cas, la valeur attendue $\widehat{X}_{T-u}(1)$ se formule par une expression analogue à celle de la prévision d'horizon 1.

Il reste encore à estimer, à un instant déterminé, la précision de la valeur prédite par le modèle au moyen du résidu, qui est la différence entre les valeurs observée et attendue.

4.2. Erreur de prévision mensuelle

L'erreur de prévision mensuelle à l'horizon h , $e_T(h)$, représente l'écart entre X_{T+h} , la valeur théorique de la série au temps $T+h$, et $\widehat{X}_T(h)$, celle qui est prévue à l'horizon h :

$$e_T(h) = X_{T+h} - \widehat{X}_T(h).$$

Afin de calculer la variance de l'erreur de prévision, il est préférable d'écrire le modèle (1) sous la forme en moyenne mobile d'ordre infini (MA(∞)) correspondante. Sans nuire à la généralité du raisonnement, nous considérerons que μ , la moyenne théorique de Y_t , est nulle :

$$X_t = \frac{(1 - \theta_3 B^3)\varepsilon_t}{(1 - \phi_1 B - \phi_2 B^2)(1 - \phi_{12} B^{12})(1 - B)} = \psi(B)\varepsilon_t,$$

avec $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$, si $\psi_0 = 1$.

ε_t représente comme précédemment un bruit blanc, dont la variance est égale à la variance résiduelle du modèle ajusté. Les paramètres ψ_j peuvent être calculés en fonction de ϕ_1 , ϕ_2 , ϕ_{12} et θ_3 . L'auteur principal tient à la disposition des lecteurs intéressés la méthode de calcul et les expressions résultantes.

Pour le modèle MA(∞) précédemment défini, les valeurs futures s'écrivent pour T observations disponibles et un horizon h de prévision :

$$X_{T+h} = \sum_{j=0}^{\infty} \psi_j \varepsilon_{T+h-j}.$$

La prévision au sens des moindres carrés, obtenue à partir de ce modèle, s'exprime de la manière suivante :

$$\widehat{X}_T(h) = \sum_{j=0}^{\infty} \psi_j E(\varepsilon_{T+h-j} | X_T, X_{T-1}, X_{T-2}, \dots),$$

Après application des règles (2), la formule de la prévision se transforme comme suit :

$$\widehat{X}_T(h) = \sum_{j=h}^{\infty} \psi_j \varepsilon_{T+h-j}.$$

Pour un modèle MA(∞), la formule générale de l'erreur de prévision s'écrit :

$$e_T(h) = \sum_{j=0}^{\infty} \psi_j \varepsilon_{T+h-j} - \sum_{j=h}^{\infty} \psi_j \varepsilon_{T+h-j} = \sum_{j=0}^{h-1} \psi_j \varepsilon_{T+h-j}.$$

L'erreur de prévision à l'horizon 1 est simplement égale au bruit blanc d'instant $T + 1$:

$$e_T(1) = \varepsilon_{T+1}.$$

Si nous désirons calculer l'erreur de prévision, relative au passé ou au présent de la série chronologique, nous pouvons adopter un raisonnement analogue à celui utilisé pour la prévision (Point 4.1).

La connaissance de la forme de l'erreur de prévision mensuelle permet de calculer sa variance et son intervalle de prévision. Elle assure aussi le calcul de l'erreur de prévision agrégée.

4.3. Intervalle de prévision mensuelle

L'erreur de prévision du modèle $MA(\infty)$ est essentiellement constituée d'un bruit blanc; ce dernier a une moyenne nulle et une variance σ_ε^2 , égale à la variance résiduelle du modèle (1) ajusté; de plus, les variables du bruit blanc d'instant différents ne sont pas corrélées. Dans ces conditions, la variance de l'erreur de prévision s'écrit :

$$\sigma_{e_T(h)}^2 = \sigma_\varepsilon^2 \sum_{j=0}^{h-1} \psi_j^2$$

Dans cette expression, il faut encore remplacer les coefficients ψ_j du modèle $MA(\infty)$ par leur valeur en fonction des paramètres $\phi_1, \phi_2, \phi_{12}$ et θ_3 du modèle ajusté.

Sous l'hypothèse que le bruit blanc est distribué normalement, les erreurs de prévision mensuelles seront supposées de distribution asymptotiquement normale de moyenne nulle et de variance $\sigma_{e_T(h)}^2$, obtenue à l'aide de l'expression générale ci-dessus.

Cette propriété permet de bâtir l'intervalle de prévision mensuelle de niveau asymptotique $1 - \alpha$:

$$\left[\widehat{X}_T(h) \pm u_{1-\alpha/2} \widehat{\sigma}_{e_T(h)} \right],$$

où α représente le risque de première espèce, $u_{1-\alpha/2}$, le quantile de la distribution normale réduite correspondant à la probabilité $1 - \alpha/2$ et $\widehat{\sigma}_{e_T(h)}$, l'expression de $\sigma_{e_T(h)}$ où σ_ε et les coefficients ψ_j sont remplacés par leurs estimateurs.

5. Agrégation des données mensuelles

La série mensuelle contient à l'heure actuelle 84 observations. Les 28 données de la série trimestrielle seraient insuffisantes pour mettre en œuvre un modèle ARIMA. Si le modèle (1), estimé sur la base des données mensuelles, était complet (avec 3 paramètres de moyenne mobile), il deviendrait de type ARIMA $(2, 1, 4)(1, 0, 0)_4$, s'il était ajusté à la série trimestrielle correspondante (PINO *et al.*, 1987; ROSSANA et SEATER, 1995). Aux caractéristiques du modèle initial complet (différentiation

simple, 2 termes autorégressifs et 3 termes de moyenne mobile) s'ajouterait un terme de moyenne mobile.

Dans la situation présente, le modèle (1) est ajusté aux données mensuelles (Figure 2) et les prévisions obtenues subissent une agrégation trimestrielle (Figure 3) et semestrielle (Figure 4).

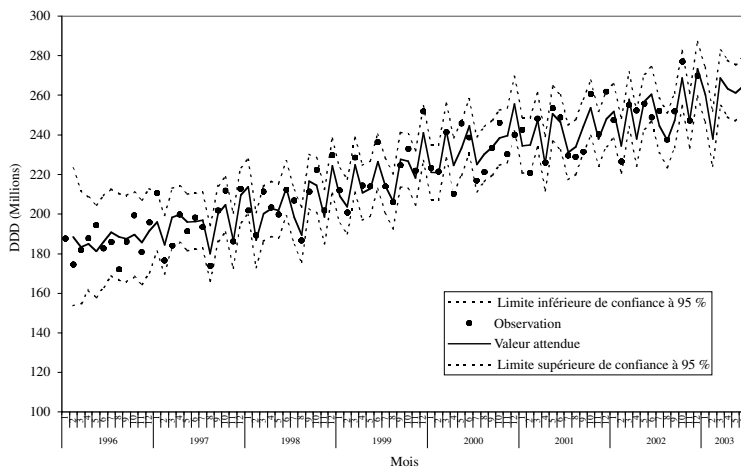


FIGURE 2

Modélisation de la série chronologique mensuelle du nombre d'unités de consommation (DDD), observée de janvier 1996 à décembre 2002, et prévisions mensuelles de janvier à juin 2003

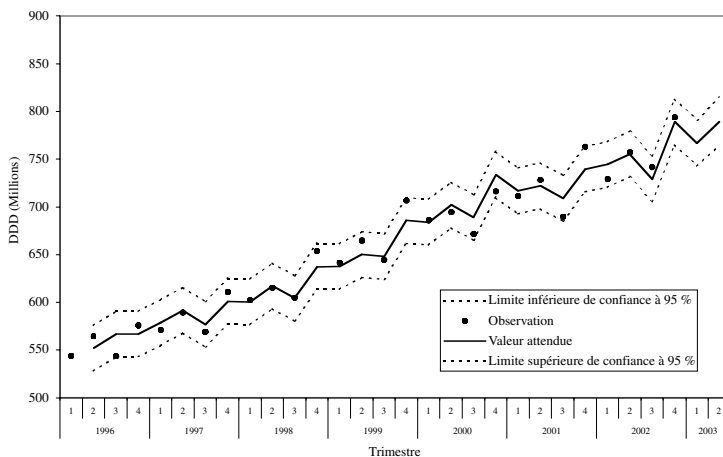


FIGURE 3

Agrégation trimestrielle de la série chronologique mensuelle du nombre d'unités de consommation (DDD), observée de janvier 1996 à décembre 2002, et prévisions trimestrielles de janvier à juin 2003

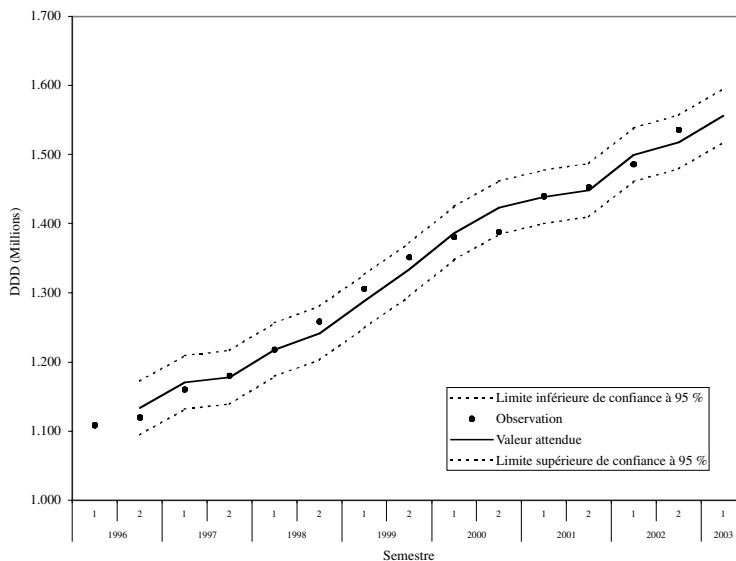


FIGURE 4

Agrégation semestrielle de la série chronologique mensuelle du nombre d'unités de consommation (DDD), observée de janvier 1996 à décembre 2002, et prévision semestrielle de janvier à juin 2003

5.1. Valeur agrégée théorique

À partir des données mensuelles X_t , la série agrégée se construit de la manière suivante

$$C'_g = \sum_{v=0}^{m-1} X_{t-v},$$

en considérant que m est l'ordre d'agrégation, c'est-à-dire le nombre de valeurs distinctes qu'il faut sommer. L'indice v permet de parcourir l'intervalle sur lequel s'effectue l'agrégation. Pour passer des valeurs mensuelles à des chiffres trimestriels et semestriels, m est choisi respectivement égal à 3 et 6. Une relation lie encore t et g :

$$t = mg.$$

5.2. Valeur agrégée prédite

De la même façon, les valeurs prédites agrégées $\hat{C}_G(r)$ sont obtenues par sommation des valeurs prédites mensuelles, définies au Point 4.1,

$$\hat{C}_G(r) = \sum_{v=0}^{m-1} \hat{X}_T(h-v),$$

si nous appelons r l'horizon agrégé et G , la dernière observation agrégée disponible. Des relations lient ces grandeurs à celles qui leur correspondent dans la série mensuelle :

$$T = mG \quad \text{et} \quad h = mr.$$

Dans le cas présent, nous pouvons écrire les 2 prévisions trimestrielles à partir du dernier trimestre disponible G :

$$\widehat{C}_G(1) = \widehat{X}_T(1) + \widehat{X}_T(2) + \widehat{X}_T(3)$$

$$\widehat{C}_G(2) = \widehat{X}_T(4) + \widehat{X}_T(5) + \widehat{X}_T(6).$$

De même, la prévision semestrielle, calculée à partir du dernier semestre disponible S , se formule :

$$\widehat{C}_S(1) = \widehat{X}_T(1) + \widehat{X}_T(2) + \widehat{X}_T(3) + \widehat{X}_T(4) + \widehat{X}_T(5) + \widehat{X}_T(6).$$

Les valeurs prédites mensuelles relatives au passé de la série peuvent être sommées de la même manière pour donner des prédictions trimestrielles et semestrielles.

Le résidu, différence entre les valeurs agrégées observée et attendue, permettra d'estimer l'erreur de prédiction agrégée.

5.3. Erreur de prévision agrégée

L'erreur de prévision agrégée à l'horizon r , $e_G(r)$, représente l'écart entre C_{G+r} , la valeur de la série agrégée au temps $G + r$, et $\widehat{C}_G(r)$, celle qui est prévue à l'horizon r :

$$e_G(r) = C_{G+r} - \widehat{C}_G(r).$$

Elle peut être exprimée à partir des erreurs de prévision mensuelles :

$$e_G(r) = \sum_{v=0}^{m-1} X_{T+h-v} - \sum_{v=0}^{m-1} \widehat{X}_T(h-v) = \sum_{v=0}^{m-1} e_T(h-v)$$

Afin de calculer l'erreur de prévision trimestrielle, il convient d'additionner les erreurs de prévision mensuelles correspondantes (Point 4.2). Ces dernières ont été obtenues à partir de la forme MA(∞) du modèle (1). Pour les deux premiers horizons

trimestriels, nous pouvons écrire, en rassemblant les termes relatifs au bruit blanc de même instant :

$$e_G(1) = e_T(1) + e_T(2) + e_T(3)$$

$$e_G(1) = \varepsilon_{T+3} + (1 + \psi_1)\varepsilon_{T+2} + (1 + \psi_1 + \psi_2)\varepsilon_{T+1}$$

$$e_G(2) = e_T(4) + e_T(5) + e_T(6)$$

$$e_G(2) = \varepsilon_{T+6} + (1 + \psi_1)\varepsilon_{T+5} + (1 + \psi_1 + \psi_2)\varepsilon_{T+4} + (\psi_1 + \psi_2 + \psi_3)\varepsilon_{T+3} + (\psi_2 + \psi_3 + \psi_4)\varepsilon_{T+2} + (\psi_3 + \psi_4 + \psi_5)\varepsilon_{T+1}$$

où les coefficients ψ_j sont ceux du modèle MA(∞) déjà calculés pour les prévisions mensuelles. L'addition des 6 erreurs de prévisions mensuelles permet de constituer l'erreur de prévision semestrielle :

$$e_S(1) = \varepsilon_{T+6} + (1 + \psi_1)\varepsilon_{T+5} + (1 + \psi_1 + \psi_2)\varepsilon_{T+4} + (1 + \psi_1 + \psi_2 + \psi_3)\varepsilon_{T+3} + (1 + \psi_1 + \psi_2 + \psi_3 + \psi_4)\varepsilon_{T+2} + (1 + \psi_1 + \psi_2 + \psi_3 + \psi_4 + \psi_5)\varepsilon_{T+1}.$$

5.4. Intervalle de prévision agrégée

En remplaçant dans l'erreur de prévision trimestrielle les coefficients par leur expression en fonction des paramètres ϕ_1 , ϕ_2 , ϕ_{12} et θ_3 du modèle (1), nous pouvons trouver la variance de l'erreur de prévision trimestrielle. Elle est constituée de variables du bruit blanc d'instant différents, donc non corrélées, de moyenne nulle et de variance σ_ε^2 , égale à la variance résiduelle du modèle ajusté :

$$\begin{aligned} \sigma_{e_G(1)}^2 &= \sigma_\varepsilon^2 [1 + (1 + \psi_1)^2 + (1 + \psi_1 + \psi_2)^2] \\ \sigma_{e_G(2)}^2 &= \sigma_\varepsilon^2 [1 + (1 + \psi_1)^2 + (1 + \psi_1 + \psi_2)^2 + (\psi_1 + \psi_2 + \psi_3)^2 + (\psi_2 + \psi_3 + \psi_4)^2 + (\psi_3 + \psi_4 + \psi_5)^2]. \end{aligned}$$

Ces expressions sont sensiblement plus complexes que celles qui résulteraient de l'agrégation de valeurs indépendantes, où la variance résiduelle σ_ε^2 serait seulement multipliée par 3.

En suivant la même logique, il est possible de déterminer la variance de l'erreur de prévision semestrielle :

$$\begin{aligned} \sigma_{e_S(1)}^2 &= \sigma_\varepsilon^2 [1 + (1 + \psi_1)^2 + (1 + \psi_1 + \psi_2)^2 + (1 + \psi_1 + \psi_2 + \psi_3)^2 + (1 + \psi_1 + \psi_2 + \psi_3 + \psi_4)^2 + (1 + \psi_1 + \psi_2 + \psi_3 + \psi_4 + \psi_5)^2]. \end{aligned}$$

Grâce à l'argument déjà invoqué pour les prévisions mensuelles (Point 4.1), la variance de l'erreur de prévision trimestrielle ou semestrielle, relative au passé de la série chronologique, sera respectivement égale à celle d'horizon trimestriel ou semestriel 1.

Par un raisonnement analogue à celui développé au Point 4.3, l'intervalle de prévision agrégée s'obtient facilement en remplaçant respectivement la prévision mensuelle et son écart-type par la valeur prédite agrégée et l'écart-type correspondant.

6. Application

Le modèle proposé ARIMA $(2, 1, 3)(1, 0, 0)_{12}$, dont les 2 premiers paramètres de moyenne mobile sont nuls (modèle (1)), est ajusté à la série chronologique mensuelle du nombre d'unités de consommation, du coût à charge du patient et du coût remboursé, relatifs à l'ensemble des délivrances médicamenteuses en Belgique, pendant la période de janvier 1996 à décembre 2002. Celles-ci contiennent 84 observations, dont 83 résidus sont effectivement calculés.

Dans l'examen des résultats, nous avons considéré un niveau de signification de 5 %.

À l'examen du Tableau 1, le modèle (1) s'adapte de façon satisfaisante aux différentes séries pendant la période mentionnée plus haut. La probabilité du test de Dickey et Fuller, inférieure au seuil de signification, indique que la différenciation première s'impose pour rendre les séries stationnaires. Le test de Ljung et Box, non significatif, montre que le résidu correspond à un bruit blanc.

TABLEAU 1

Résultats diagnostiques de l'ajustement du modèle ARIMA $(2, 1, 3)(1, 0, 0)_{12}$, pendant la période de janvier 1996 à décembre 2002

Variable	Écart-type résiduel	Probabilité du test de Dickey-Fuller*	Probabilité du test de Ljung et Box [#]
Nombre d'unités de consommation	7 059 848, 0	0, 0010	0,1591
Coût à charge du patient	1 469 400, 4 €	0, 0010	0, 3715
Coût remboursé	4 916 854, 5 €	0, 0010	0, 8261

* La probabilité du test F de Dickey-Fuller augmenté, dont l'hypothèse nulle suppose une différence première de moyenne égale à 0, correspond au retard 1.

[#] La probabilité du test de Ljung et Box est relative aux 24 premiers retards.

Tous les paramètres du modèle (Tableau 2) sont significativement différents de 0. Les probabilités correspondent à une distribution t à 78 degrés de liberté, soit le nombre de résidus (83) moins celui de paramètres estimés (5). À l'exception de la moyenne μ , les estimations des paramètres, pris séparément, ainsi que leur

écart-type sont relativement stables pour les 3 variables considérées. Les coefficients de corrélation entre les estimations des paramètres restent faibles et indiquent une absence de colinéarité entre eux.

La bonne adéquation du modèle peut être appréciée à l'examen des résidus. Leurs autocorrélations simples, présentées à la Figure 5, ne sont pas significatives. Les résidus, dont la Figure 6 montre l'évolution au cours du temps, ne possèdent pas de structure particulière. Leur distribution peut être donc considérée comme celle d'un bruit blanc.

Avant l'ajustement du modèle, les coefficients de corrélation croisée, correspondant aux séries initiales, sont maximaux pour un retard nul et possèdent une forte valeur (0,99 entre le nombre d'unités de consommation et le coût remboursé, 0,94 entre le nombre d'unités de consommation et le coût à charge du patient et 0,94 encore entre les deux coûts); ils s'estompent graduellement, lorsque les deux séries concernées se décalent dans le temps.

TABLEAU 2

Estimation du maximum de vraisemblance des paramètres du modèle, ajusté pendant la période de janvier 1996 à décembre 2002

Paramètre	Valeur estimée	Écart-type	Probabilité approchée du test de signification
<i>Nombre d'unités de consommation</i>			
μ	897 351, 3	201 183, 0	< 0, 0001
θ_3	0, 72665	0, 10041	< 0, 0001
ϕ_1	-1, 07880	0, 03783	< 0, 0001
ϕ_2	-0, 93687	0, 04550	< 0, 0001
ϕ_{12}	0, 74695	0, 07733	< 0, 0001
<i>Coût à charge du patient (€ courants)</i>			
μ	128 422, 3	52 257, 0	0, 0140
θ_3	0, 70302	0, 10682	< 0, 0001
ϕ_1	-1, 01871	0, 05183	< 0, 0001
ϕ_2	-0, 88836	0, 06561	< 0, 0001
ϕ_{12}	0, 78835	0, 06499	< 0, 0001
<i>Coût remboursé (€ courants)</i>			
μ	733 510, 1	125 428, 6	< 0, 0001
θ_3	0, 78080	0, 09192	< 0, 0001
ϕ_1	-1, 06492	0, 03917	< 0, 0001
ϕ_2	-0, 93049	0, 04765	< 0, 0001
ϕ_{12}	0, 77699	0, 07043	< 0, 0001

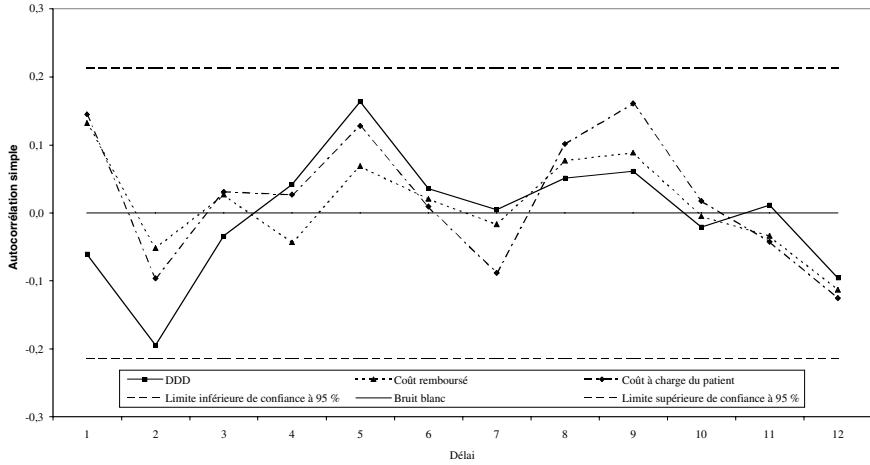


FIGURE 5

Corrélogramme des résidus du nombre d'unités de consommation (DDD), du coût remboursé et du coût à charge du patient, après ajustement du modèle sur les séries chronologiques mensuelles pendant la période de janvier 1996 à décembre 2002

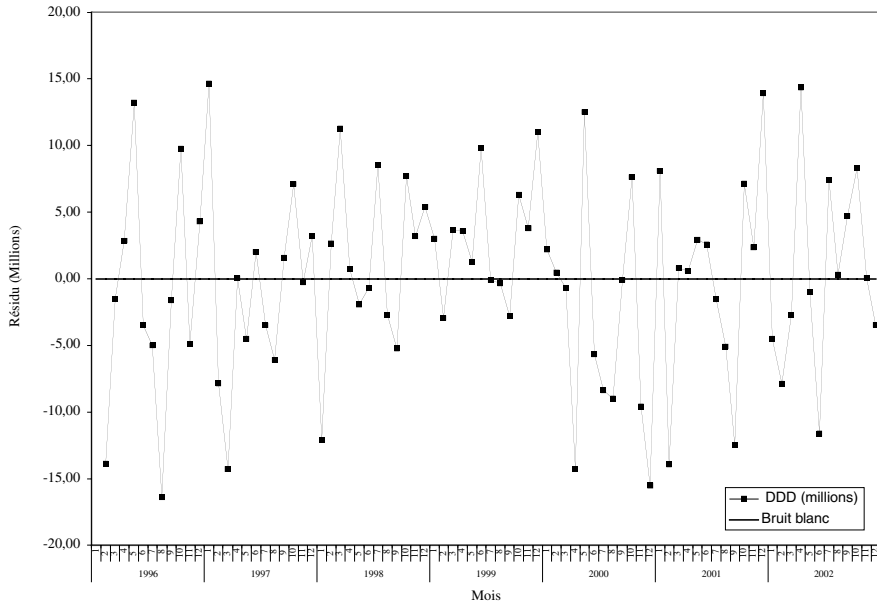


FIGURE 6

Évolution des résidus du nombre mensuel d'unités de consommation (DDD), après ajustement du modèle pendant la période de janvier 1996 à décembre 2002

Les coefficients de corrélation croisée de retard nul, calculés sur les résidus après ajustement du modèle de même type, restent élevés (0,91 entre le nombre d'unités de consommation et le coût remboursé, 0,85 entre le nombre d'unités de consommation et le coût à charge du patient et 0,90 entre les deux coûts). Maximaux pour ce retard nul, ils décroissent de façon brutale, lorsque les deux séries envisagées sont décalées dans le temps.

L'adéquation du modèle ARIMA de même type aux trois séries synchrones s'explique sans doute par le fait qu'elles sont liées par une variable commune, le nombre de boîtes délivrées, évoquée dans l'introduction, mais qui n'est pas présentée ici.

TABLEAU 3
*Prévisions mensuelles obtenues sur la base du modèle
ajusté pendant la période de janvier 1996 à décembre 2002*

Date	Limite inférieure de prévision à 95 %	Prévision	Limite supérieure de prévision à 95 %	Écart-type de l'erreur de prévision	Écart-type de prévision exprimé en pourcentage de l'écart-type résiduel	Écart-type de prévision exprimé en pourcentage de la prévision
<i>Nombre d'unités de consommation (10³)</i>						
01/01/2003	246 433, 6	260 270, 7	274 107, 7	7 059, 8	100,00	2,71
01/02/2003	224 006, 7	237 886, 6	251 766, 5	7 081, 7	100,31	2,98
01/03/2003	254 620, 5	268 651, 0	282 681, 5	7 158, 6	101,40	2,66
01/04/2003	249 217, 5	263 485, 5	277 753, 5	7 279, 7	103,11	2,76
01/05/2003	246 930, 3	261 228, 9	275 527, 5	7 295, 3	103,34	2,79
01/06/2003	250 347, 8	264 840, 3	279 332, 7	7 394, 2	104,74	2,79
<i>Coût à charge du patient (10³€ courants)</i>						
01/01/2003	37 016, 1	39 896, 1	42 776, 1	1 469, 4	100,00	3,68
01/02/2003	32 707, 2	35 587, 7	38 468, 1	1 469, 7	100,02	4,13
01/03/2003	36 951, 7	39 856, 7	42 761, 6	1 482, 2	100,87	3,72
01/04/2003	36 115, 2	39 066, 3	42 017, 4	1 505, 7	102,47	3,85
01/05/2003	35 540, 9	38 492, 0	41 443, 1	1 505, 7	102,47	3,91
01/06/2003	35 461, 2	38 439, 6	41 418, 0	1 519, 6	103,42	3,95
<i>Coût remboursé (10³€ courants)</i>						
01/01/2003	153 856, 7	163 493, 6	173 130, 4	4 916, 9	100,00	3,01
01/02/2003	140 920, 4	150 577, 5	160 234, 7	4 927, 2	100,21	3,27
01/03/2003	158 421, 3	168 170, 4	177 919, 6	4 974, 1	101,16	2,96
01/04/2003	152 789, 9	162 621, 6	172 453, 3	5 016, 3	102,02	3,08
01/05/2003	153 187, 3	163 030, 9	172 874, 6	5 022, 4	102,15	3,08
01/06/2003	153 691, 7	163 641, 0	173 590, 3	5 076, 2	103,24	3,10

TABLEAU 4
*Prévisions trimestrielles et semestrielles obtenues sur la base du modèle
ajusté pendant la période de janvier 1996 à décembre 2002*

Date	Limite inférieure de prévision à 95 %	Prévision	Limite supérieure de prévision à 95 %	Écart-type de l'erreur de prévision	Écart-type de prévision exprimé en pourcentage de l'écart-type résiduel	Écart-type de prévision exprimé en pourcentage de la prévision
<i>Nombre d'unités de consommation (10³)</i>						
Trimestre						
01/2003	742 873, 4	766 808, 3	790 743, 2	12 211, 9	172, 98	1, 59
02/2003	764 747, 5	789 554, 7	814 362, 0	12 657, 0	179, 28	1, 60
Semestre						
01/2003	1 517 632, 5	1 556 363, 0	1 595 093, 5	19 760, 8	279, 90	1, 27
<i>Coût à charge du patient (10³€ courants)</i>						
Trimestre						
01/2003	110 189, 0	115 340, 4	120 491, 8	2 628, 3	178, 87	2, 28
02/2003	110 625, 0	115 997, 9	121 370, 9	2 741, 4	186, 56	2, 36
Semestre						
01/2003	222 904, 8	231 338, 4	239 771, 9	4 302, 9	292, 83	1, 86
<i>Coût remboursé (10³€ courants)</i>						
Trimestre						
01/2003	465 474, 3	482 241, 5	499 008, 7	8 554, 9	173, 99	1, 77
02/2003	472 130, 9	489 293, 6	506 456, 2	8 756, 6	178, 09	1, 79
Semestre						
01/2003	945 108, 6	971 535, 1	997 961, 5	13 483, 1	274, 22	1, 39

La Figure 2 présente les valeurs mensuelles observées et prédites ainsi que l'intervalle de prévision du nombre d'unités de consommation. Pour chacune des trois variables analysées, le Tableau 3 reprend les informations relatives aux 6 prévisions mensuelles, tandis que le Tableau 4 présente les prévisions trimestrielles et semestrielles.

Entre les horizons 1 et 6, l'écart-type de l'erreur de la prévision mensuelle augmente respectivement de 4,7, 3,4 et 3,2 % pour le nombre d'unités de consommation, le coût à charge du patient et le coût remboursé. Le rapport de l'écart-type à la prévision mensuelle varie de 2,7 à 3,0 % pour le nombre d'unités de consommation, de 3,7 à 4,1 % pour le coût à charge du patient et de 3,0 à 3,3 % pour le coût remboursé.

Une certaine stabilisation résulte de l'agrégation de la série chronologique en trimestre (Figure 3). Selon les variables considérées, l'agrégation des valeurs mensuelles (Tableau 4) détermine un accroissement de l'écart-type de prévision (par rapport à l'écart-type résiduel) de 73,0 à 78,9 % pour le premier horizon trimestriel,

de 78,1 à 86,6 % pour le second. Le passage d'un trimestre à l'autre entraîne une augmentation respective de l'écart-type de 3,6, 4,3 et 2,4 %, pour le nombre d'unités de consommation, le coût à charge du patient et le coût remboursé.

Lors de l'agrégation semestrielle (Figure 4), les oscillations périodiques ont pratiquement disparu pour ne laisser apparente que la tendance générale. L'écart-type de l'erreur de prévision semestrielle (Tableau 4) s'accroît de 174,2 à 192,8 % par rapport à l'écart-type résiduel, selon les variables analysées.

TABLEAU 5

*Prévisions mensuelles du nombre d'unités de consommation (en milliers)
obtenues sur la base du modèle ajusté pendant les périodes
de janvier 1996 à juin 2001, décembre 2001 et juin 2002*

Date	Valeur observée	Limite inférieure de prévision à 95 %	Prévision	Limite supérieure de prévision à 95 %
<i>Ajustement de janvier 1996 à juin 2001</i>				
01/07/2001	229 635, 5	217 598, 4	231 188, 9	244 779, 4
01/08/2001	228 859, 8	220 302, 2	233 943, 7	247 585, 2
01/09/2001	231 448, 0	229 950, 4	243 771, 2	257 592, 0
01/10/2001	260 725, 1	239 274, 1	253 306, 0	267 337, 9
01/11/2001	240 306, 7	227 180, 3	241 245, 1	255 309, 9
01/12/2001	261 939, 4	235 046, 6	249 336, 7	263 626, 7
<i>Ajustement de janvier 1996 à décembre 2001</i>				
01/01/2002	247 441, 5	238 169, 8	251 906, 2	265 642, 6
01/02/2002	226 586, 2	219 935, 4	233 730, 7	247 525, 9
01/03/2002	255 181, 4	244 121, 4	258 113, 3	272 105, 2
01/04/2002	252 396, 2	225 075, 1	239 225, 6	253 376, 1
01/05/2002	255 799, 0	245 815, 4	260 001, 4	274 187, 4
01/06/2002	248 917, 5	244 719, 6	259 143, 9	273 568, 2
<i>Ajustement de janvier 1996 à juin 2002</i>				
01/07/2002	252 025, 0	229 854, 8	243 968, 6	258 082, 3
01/08/2002	237 570, 3	224 011, 1	238 165, 4	252 319, 8
01/09/2002	251 971, 8	231 946, 6	246 253, 1	260 559, 6
01/10/2002	276 988, 7	252 522, 1	267 002, 7	281 483, 3
01/11/2002	247 179, 7	233 338, 7	247 845, 8	262 352, 8
01/12/2002	269 889, 2	255 185, 6	269 878, 2	284 570, 7

L'agrégation trimestrielle provoque un gain de précision pour la prévision. Le rapport de l'écart-type à la prévision trimestrielle est de 1,6 % pour le nombre d'unités de consommation, de 2,3 à 2,4 % pour le coût à charge du patient et de 1,8 % pour le coût remboursé. Cette constatation est corroborée lors de l'agrégation semestrielle. Le rapport de l'écart-type à la prévision semestrielle est respectivement égal à 1,3,

1,9 et 1,4 % pour le nombre d'unités de consommation, le coût à charge du patient et le coût remboursé.

Pour juger de la validité de l'ajustement au cours du temps, le même modèle ARIMA (1) est ajusté sur les séries mentionnées, mais limitées successivement à juin 2001, décembre 2001 et juin 2002. Le Tableau 5 présente, à côté des nombres d'unités de consommation observés mensuels, les valeurs prédites obtenues à un horizon de 6 mois, ainsi que leur intervalle de prévision à 95 %. Les observations sont contenues dans ce dernier.

Ces valeurs mensuelles sont agrégées par trimestre et semestre au Tableau 6. Ici également, l'intervalle de prévision à 95 % contient les observations.

TABLEAU 6
Prévisions trimestrielles et semestrielles du nombre d'unités de consommation (en milliers) obtenues sur la base du modèle ajusté pendant les périodes de janvier 1996 à juin 2001, décembre 2001 et juin 2002

Date	Valeur observée	Limite inférieure de prévision à 95 %	Prévision	Limite supérieure de prévision à 95 %
<i>Ajustement de janvier 1996 à juin 2001</i>				
Trimestre				
03/2001	689 943, 3	685 391, 8	708 903, 8	732 415, 7
04/2001	762 971, 3	719 509, 7	743 887, 8	768 265, 9
Semestre				
02/2001	1 452 914, 6	1 414 720, 2	1 452 791, 5	1 490 862, 8
<i>Ajustement de janvier 1996 à décembre 2001</i>				
Trimestre				
01/2002	729 209, 2	720 020, 1	743 750, 2	767 480, 2
02/2002	757 112, 8	733 897, 7	758 370, 9	782 844, 0
Semestre				
01/2002	1 486 321, 9	1 464 102, 6	1 502 121, 1	1 540 139, 5
<i>Ajustement de janvier 1996 à juin 2002</i>				
Trimestre				
03/2002	741 567, 1	703 932, 3	728 387, 1	752 841, 9
04/2002	794 057, 6	759 545, 8	784 726, 6	809 907, 4
Semestre				
02/2002	1 535 624, 7	1 474 058, 7	1 513 113, 7	1 552 168, 7

7. Discussion

Les méthodes propres aux séries chronologiques, appliquées aux données de consommation médicamenteuse, permettent d'établir des prévisions. D'autres

modèles ont été proposés dans ce même contexte; ils ont été développés pour évaluer l'influence d'événements sur la série temporelle. WEATHERBY *et al.* (2001) étudient une variable dépendante binomiale, la prescription simultanée de médicaments contre-indiqués en association avec le cisapride, grâce à la régression logistique (procédure SAS LOGISTIC), incluant des éléments temporels (année, trimestre) comme facteurs explicatifs.

Afin d'explorer l'association temporelle entre le taux de suicide et la prise d'antidépresseurs en Suède, CARLSTEN *et al.* (2001) proposent d'ajuster à une série chronologique du nombre de suicides un modèle log-linéaire à deux pentes, avec identification d'un point de rupture correspondant ici à l'introduction d'une nouvelle classe de médicaments antidépresseurs. Ce modèle fournit une meilleure adaptation aux données que son correspondant à une seule pente. Si ces modèles permettent une mesure d'impact par une comparaison des caractéristiques de la série (la tendance, dans ces deux cas) avant et après l'événement, ils ne tiennent pas compte des autocorrélations propres aux séries temporelles qui peuvent influencer les prévisions.

Enfin, McMANUS *et al.* (1996) analysent une série chronologique du nombre de prescriptions avant et après adoption d'une participation du patient au paiement du médicament. Ils utilisent dans ce but un modèle de régression linéaire segmentée, contenant divers termes correspondant au niveau et à la tendance avant et après introduction de la mesure. Un contrôle de l'autocorrélation intervient par introduction d'un terme autorégressif saisonnier de période 12 (procédure SAS AUTOREG).

La régression linéaire segmentée s'adapte de façon plus satisfaisante à la série chronologique que le modèle log-linéaire à 2 pentes, mais l'ajustement et la prévision ne s'effectuent pas avec un nombre aussi réduit de paramètres qu'avec les modèles ARIMA. En outre, ces derniers présentent l'avantage d'associer à chaque valeur prédite un écart-type, qui intervient lui-même dans la construction d'un intervalle de prévision.

L'estimation des paramètres, le calcul des prévisions mensuelles et de leur intervalle sont réalisés de manière standard par des logiciels tels que SAS (SAS INSTITUTE, 2000) et Statistica (STATSOFT, 1999). L'agrégation de données demande une programmation spécifique.

Le modèle employé ici au départ sur des données mensuelles donne des prévisions assez précises; l'écart-type de prévision est de 3 à 4 % de la valeur prédite. Pour les horizons 2 à 6, l'écart-type de prévision mensuelle augmente de 3 à 5 % par rapport à l'écart-type résiduel. Il est possible d'agréger les données en trimestre et de déduire l'écart-type et l'intervalle de prévision trimestrielle, tout en se servant des estimations obtenues à partir de la série chronologique mensuelle. Cette agrégation peut être étendue à une période plus longue (semestre), pour autant que la longueur de la série initiale le permette.

La stabilité des observations et des prévisions s'accroît avec le degré d'agrégation. Si cette dernière est semestrielle, seule la tendance reste visible. Le fait d'agréger la série mensuelle en trimestre ou semestre donne des prévisions plus précises; l'écart-type de prévision agrégée est de 2 % de la valeur prédite trimestrielle et de 1 à 2 % de la prévision semestrielle. Par rapport à l'horizon 1, l'écart-type de prévision trimestrielle de l'horizon 2 augmente de 2 à 4 %. L'écart-type de prévision trimestrielle

s'accroît, par rapport à l'écart-type résiduel, de 73 à 79 % pour le premier horizon et de 78 à 87 % pour le second. L'agrégation semestrielle entraîne une augmentation de l'écart-type de prévision de 174 à 193 % par rapport à l'écart-type résiduel. Les intervalles de prévision sont d'une étendue acceptable pour l'objectif prévisionnel.

L'adéquation du modèle ARIMA est encore vérifiée au cours du temps par des ajustements limités à juin 2001, décembre 2001 et juin 2002. Les prévisions mensuelles, trimestrielles et semestrielles, obtenues sur ces périodes, sont comparées avec les observations, qui sont contenues dans l'intervalle de prévision correspondant.

Au modèle ARIMA tel qu'il est décrit ici, il est possible d'adjoindre des variables explicatives. Récemment, l'effet d'une campagne nationale d'information sur les antibiotiques a été testé à l'aide d'un de ces modèles, qui incluait une intervention (campagne d'information) et un indice d'épidémie (BAURAIND *et al.*, 2001, 2003).

Remerciements

Les auteurs voudraient particulièrement remercier le Professeur Guy Mélard et le Docteur Toufik Zahaf (Institut de Statistique et de Recherche Opérationnelle, Université Libre de Bruxelles) pour les conseils qu'ils leur ont dispensés lors de la mise en application des modèles ARIMA. Les remarques et suggestions de deux lecteurs anonymes ont permis d'améliorer la version initiale de cet article. Qu'ils soient ici aussi grandement remerciés.

Références

- ABRAHAM B.(1982), Temporal Aggregation and Time Series. *International Statistical Review*, 50 : 285-291.
- BAURAIND I., VANDEN BREM T., BOGAERT M., GOOSSENS H., MOUCHET P., TREFOIS P., MARCHAL J.-L., SEYS B., TULKENS P.M. and VERBIST L. (2001), Evaluation of the Impact of a Public Campaign for a More Rational Use of Antibiotics (AB) in Belgium. *Abstracts of the 41st Interscience Conference on Antimicrobial Agents and Chemotherapy (ICAAC)*, Chicago, USA, December 16 - 19, 2001; Addendum p. 11.
- BAURAIND I., GOOSSENS H., HENDRICKX E., YANE F., SEYS B., MARCHAL J.-L., TULKENS P.M., VERBIST L., LOPEZ-LOZANO J.-M. and BEYAERT A. (2003), Two Years of National Public Campaigns to Promote Appropriate Use of Antibiotics in the Community in Belgium. *Abstracts of the 43rd Interscience Conference on Antimicrobial Agents and Chemotherapy (ICAAC)*, Chicago, USA, September 14 – 17, 2003; Addendum p. 11.
- BOX G.E.P. and JENKINS G.M. (1976), *Time Series Analysis : Forecasting and Control*. 2nd Edition, Holden-Day, San Francisco.
- CARLSTEN A., WAERN M., EKEDAHL A. and RANSTAM J. (2001), Antidepressant Medications and Suicide in Sweden. *Pharmacoepidemiology and Drug Safety*, 10 : 525-530.

- DICKEY D.A., BELL W.R. and MILLER R.B. (1986), Unit Roots in Time Series Models : Tests and Implications. *The American Statistician*, 40 (1) : 12-26.
- DICKEY D.A., HASZA D.P. and FULLER W.A. (1984), Testing of Unit Roots in Seasonal Time Series. *Journal of American Statistical Association*, 79 (386) : 355-367.
- DIGGLE P.J. (1990), *Time Series. A biostatistical Introduction*. Oxford Statistical Science Series 5, New York, 257 pp.
- DROESBEKE J.-J., FICHET B. et TASSI P. (1989), *Séries chronologiques. Théorie et pratique des modèles ARIMA*. Éditions Economica, Paris, 299 pp.
- HOUBEN P., LISMONT J., MARCHAL J.-L., REZETTE I., SEYS B., SPAGO B. and VANSNICK L. (2001), The IPHEB-IFEB Data Base on reimbursed Drugs : An efficient Tool to monitor Drug Exposure. *Symposium on Drug Interaction in clinical Practice. Health and Economic Consequences*. Brussels, November 17th, 2001.
- INSTITUT NATIONAL D'ASSURANCE MALADIE-INVALIDITE. (1999), Service des soins de santé. Conseil scientifique. Comité d'évaluation de la pratique médicale en matière de médicaments. *Tableaux de bord pharmaceutiques. Délivrances pharmaceutiques dans le secteur ambulatoire. Année 1998*. Bruxelles, 26 pp.
- McMANUS P., DONNELLY N., HENRY D., HALL W., PRIMROSE J. and LINDNER J. (1996), Prescription Drug Utilization following Patient Co-payment Changes in Australia. *Pharmacoepidemiology Drug Safety*, 5 : 385-392.
- MELARD GUY (1990), *Méthodes de prévision à court terme*. Éditions de l'Université de Bruxelles, Bruxelles, 468 pp.
- MONTFORT A. (1989), Processus univariés ARMA. Notions de base et extensions, in DROESBEKE J.-J., FICHET B. et TASSI P., éditeurs, *Séries chronologiques. Théorie et pratique des modèles ARIMA*. Éditions Economica, Paris, 43-73.
- PINO F.A., MORETTIN P.A. and MENTZ R.P. (1987), Modelling and Forecasting Linear Combinations of Time Series. *International Statistical Review*, 55(3) : 295-313.
- ROSSANA R.J. and SEATER J.J. (1995), Temporal Aggregation and Economic Time Series. *Journal of Business and Economic Statistics*, 13(4) : 441-451.
- SAS INSTITUTE. (2000), *SAS/ETS User's Guide*, Version 8, Volumes 1 and 2, Cary, 1596 pp.
- SEYS B., HOUBEN P., MARCHAL J.-L., SPAGO B. et VANSNICK L. (2001), Qu'est-ce que la DDD? *Revue de la médecine générale*, 24 (188) : 457-458.
- SEYS B., SPAGO B., HOUBEN P., MARCHAL J.-L. and VANSNICK L. (1999), Description and Validation of a Data Base on Drug Dispensing in Belgium. Proceedings of the VIth Symposium of the Association of public Health. Brussels, November 27, 1999. *Arch. Public Health*, 57 (S1), 43.
- STATSOFT. (1999), *Electronic Statistics Textbook*, Tulsa.

- VANDER STICHELE R.H., BOSSENS M., DE FALLEUR M., SEYS B. and ELSEVIERS M. (2001), Utilization of Antibiotics in Belgium. *Proceedings of the EuroDURG Meeting 2001 Integrating Drug Utilization Studies in wider Europe*, Prague, Czech Republic, June 7-9, 2001.
- WEATHERBY L.B., WALKER A.M., FIFE D., VERVAET P. and KLAUSNER M.A. (2001), Contraindicated Medications dispensed with Cisapride : Temporal Trends in Relation to the Sending of «Dear Doctor» Letters. *Pharmacoepidemiology Drug Safety*, 10 : 211-218.
- WEI W.W.S. (1994), *Time Series Analysis. Univariate and Multivariate Methods*. Addison-Wesley, Redwood City, 478 pp.
- WORLD HEALTH ORGANIZATION (2001), *Guidelines for ATC Classification and DDD Assignment*. WHO Collaborating Centre for Drug Statistics Methodology. 4th Edition, Oslo, 262 pp.