

Note de lecture rédigée par Jean-Pierre RAOULT¹

SÉRIES TEMPORELLES AVEC R MÉTHODES ET CAS

Yves ARAGON

Livre (265 pages)

Édition : Springer-Verlag France – Pratique R – 2011

Sur la page 4 de couverture de cet ouvrage, on lit : *Ce livre à destination des étudiants, des professionnels et des chercheurs sera particulièrement utile à toute personne ayant reçu une bonne formation théorique sur les séries temporelles mais pour qui la mise en pratique reste source d'embarras.*

Cette phrase décrit parfaitement la difficulté à laquelle se heurtent, singulièrement en France, de nombreux praticiens, à un niveau élevé, des séries temporelles : ceux issus de formations universitaires en Statistique se trouvent souvent désarmés devant la mise en œuvre de la culture, fortement mathématisée, qu'ils ont acquise ; ou bien ceux ayant suivi des formations en Ecoles d'Ingénieurs ou de Gestion sont trop dépourvus de substrats théoriques pour pouvoir aisément s'adapter à des situations nouvelles. Longtemps ces obstacles ont été accrus par l'inadéquation des logiciels statistiques les plus pratiqués « industriellement » à des études de données nécessitant recul et imagination ; la percée de R, depuis quelques années, a fait notablement évoluer cet état de fait : bien maîtrisé, il permet d'aborder avec lucidité et pertinence de nombreuses études statistiques.

La collection « Pratique R », lancée cette année (2011) chez Springer-Verlag France par Pierre-André Cornillon et Eric Matzner-Leber, correspond donc à un réel besoin. Après des ouvrages sur l'emploi de R pour la régression (par les directeurs de la collection) et pour les méthodes de Monte-Carlo (par Christian Robert et George Casella), voici le troisième livre publié dans cette série. Celui-ci nous paraît atteindre parfaitement son but, alliant, comme l'indique le préfacier Dominique Haughton (Bentley University) « efficacité et pragmatisme ».

Cet ouvrage ne prétend servir de premier contact ni avec la théorie statistique des séries temporelles (quoique l'essentiel soit dit dans les chapitres 1, titré *Démarche de base en séries temporelles*), et 4, titré *Modèles de base en séries temporelles*, ni avec l'emploi de R (pour lequel sont recommandées des références). En revanche sa lecture, couplée avec les mises en œuvre qui y sont indiquées, permettra au lecteur de s'initier à la démarche à suivre, dès les « explorations » qui précèdent les modélisations, devant des séries de natures différentes et, surtout, impliquant des questionnements divers en fonction des besoins des fournisseurs et utilisateurs.

Cette diversité d'approches est mise en évidence par les 7 cas qui sont étudiés en détail et occupent une centaine de pages sur les 257 de l'ouvrage (hors préface, avant-propos, table des

¹ Professeur des Universités émérite, Laboratoire d'Analyse et de Mathématiques Appliquées, Université Paris-Est-Marne-la-Vallée (France), jean-pierre.raoult@univ-mlv.fr

Note de lecture : « Séries temporelles avec R – Méthodes et cas » (Y. Aragon, 2011)

matières, index et bibliographie) : trafic d'aéroport, température mensuelle moyenne, consommation d'électricité, production de lait, marché boursier en crise, rendement d'une entreprise (2 cas).

De cet ouvrage ressort en particulier une « déontologie » dans l'approche d'une série temporelle, faite d'opiniâtreté dans la recherche du (ou des) modèle(s) pertinent(s) et de lucidité devant ce que les modèles « captent » ou ne captent pas. On en retiendra ici un exemple, celui du chapitre 11, *Production de lait*, introduit par : *Nous étudions la collecte mensuelle de lait en France de janvier 1980 à janvier 2010. Cette série a été affectée par l'introduction de quota laitiers en janvier 1984.* Nous reproduisons la fin du chapitre (que les lecteurs de cette notice qui ne connaîtraient pas le vocabulaire utilisé se reportent au livre s'ils le désirent, mais ce qui nous intéresse ici, c'est « l'esprit » de ce passage) : *La série montre une forte saisonnalité. On y observe un changement de comportement à partir de l'introduction des quota. Nous avons d'abord modélisé la série de 60 observations avant quota et avons obtenu un SARMA, modèle stationnaire. Ensuite nous avons essayé de modéliser la fin de la série et avons obtenu un SARIMA, modèle non stationnaire. Nous avons alors changé de stratégie. Nous avons ajusté : (1) un SARIMA à l'ensemble de la série, sans obtenir de modèle très satisfaisant en termes de blancheur des résidus, puis (2) un ARMAX à la même série. Dans ce dernier modèle, une indicatrice distinguait les situations avant et après quota. Après quelques tâtonnements, simplifications et éliminations de variables explicatives, nous avons obtenu un modèle satisfaisant pour toute la série, modèle où les quota n'ont pas d'effet quantifiable sur le niveau moyen de la collecte. La comparaison des AIC du SARIMA et de l'ARMAX nous a conduit à retenir l'ARMAX pour modéliser la série. L'introduction des quota semble avoir eu comme effet de diminuer très progressivement le niveau de la collecte, sans que ceci puisse être capté par un de nos modèles, en diminuant le niveau du pic de la collecte mais pas celui de son « étiage ».*

Pareille étude « fouillée » d'une série nécessite de disposer d'un outil bien adapté, à la fois puissant et maniable. Par ses talents d'exposition, sa pratique étendue du logiciel et sa vaste expérience de secteurs divers de production de séries temporelles, Yves Aragon sait convaincre ses lecteurs que tel est bien le cas avec R.