

## **DETECTION D'ANOMALIES DANS DES SERIES CHRONOLOGIQUES ET MISE EN ŒUVRE EN R **CALIBRI 14****

Jean Palate **Calibri 12**

Retraité de la Banque Nationale de Belgique **Calibri 12**

[palatejean@gmail.be](mailto:palatejean@gmail.be) **Calibri 12**

**Mots-clés** : détection d'anomalies, séries temporelles, R **Calibri 12**

**Domaine concerné** : **Calibri 12** Séries temporelles – Analyse des données

---

### **Résumé **Calibri 12****

La détection d'anomalies est une étape importante dans tout processus de production statistique. Elle permet de trouver d'éventuelles erreurs et de voir si la méthode habituelle est adéquate ou s'il convient de l'adapter.

Pour chaque étape et chaque type de données, de multiples méthodes ont été développées. Dans cette présentation, on va se focaliser sur la détection d'anomalies sur des séries chronologiques, dans une approche univariée et en utilisant des méthodes basées sur des modèles. Ce type d'approche peut s'appliquer lors de toute étape de la chaîne de production, sur des données individuelles comme sur des séries agrégées finales. Le traitement univarié simplifie considérablement l'automatisation, pour autant que l'on utilise une modélisation suffisamment générale, tel un modèle ARIMA ou un modèle structurel.

Au cours des dernières décennies, de nombreux articles et développements d'algorithmes ont été réalisés dans ce domaine, en particulier dans le cadre des méthodes d'ajustements saisonniers, pour lesquels le traitement des anomalies est essentiel.

La solution développée dans le programme Tramo-Seats (Gomez et Maravall [1996]) est particulièrement efficace. Elle consiste en une procédure itérative, qui enrichit progressivement un modèle RegARIMA avec les outliers les plus pertinents. Basée sur des techniques linéaires qui fournissent d'excellentes approximations de la solution exacte, elle est remarquablement efficace et permet le traitement de séries longues, difficilement envisageables via des techniques plus classiques. En outre, elle peut facilement être adaptée à des modèles autres que les modèles de type « Box-Jenkins » standards, par exemple à des modèles structurels ou à d'autres périodicités (données journalières...)

On a développé un package R qui implémente cet algorithme et qui donne accès à ces principales sous-routines. Cette solution généralise ce qui a été développé dans Tramo et

permet à l'utilisateur d'élaborer en R des méthodes alternatives, basées par exemple sur des modèles particuliers. En outre, à titre de comparaison, le package fournit les solutions exactes par modélisation RegARIMA (style X13) ou par modèles structurels, en utilisant le filtre de Kalman (De Jong et Penzer [1998]).

Après avoir expliqué en détail l'algorithme de détection d'anomalies par méthode approchée (généralisation de l'algorithme de Tramo) et montré en quoi il diffère de solutions exactes (RegARIMA, filtre de Kalman), des exemples/comparaisons pratiques en R seront rapidement présentés. Ils mettront en évidence l'impact limité d'une approche approximative par rapport à une solution exacte, le gain en performance ainsi qu'une application à des données journalières.

### **Bibliographie** Calibri 12

[1] Victor Gomez and Agustin Maravall. Programs tramo (time series regression with arima noise, missing observations, and outliers) and seats (signal extraction in arima time series). instructions for the user. Documento de Trabajo, 9628:56, 1996.

[2] Stefano Grassi, Gian Luigi Mazzi, and Tommaso Proietti. Automatic outlier detection for the basic structural time series model. Handbook on Seasonal Adjustment, 2018. [ec.europa.eu/eurostat/web/products-manuals-and-guidelines/KS-GQ-18-001](http://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/KS-GQ-18-001).

[3] Piet de Jong and Jeremy Penzer. Diagnosing shocks in time series. J. Am. Stat. Assoc., 93(442):796–806, June 1998.

[4] Regina Kaiser and Agustín Maravall. Seasonal outliers in time series. Documentos de trabajo/Banco de España, 9915, 1999.

[5] Anna Smyk and Jean Palate. Outlier detection and seasonality breaks with JDemetra+ 3.0, NTTS 2021.