

Les packages autour de JDemetra+ (rjd3) : une boîte à outils complète pour l’analyse des séries temporelles

Tanguy BARTHELEMY*

Résumé

JDemetra+ (Smyk et al. (2024)) est un logiciel open source d’analyse des séries temporelles dont les algorithmes sont accessibles via une interface graphique ou un écosystème de packages R (rjd3 Palate et al. (2024a)). Cet abstract a pour objectif de donner un aperçu des possibilités de ces packages, couvrant tous les domaines clef de l’analyse des séries temporelles (ajustement saisonnier, détection d’outliers, nowcasting, analyse des révisions, benchmarking et désagrégation temporelle). Nous détaillerons plus particulièrement l’ajustement saisonnier des séries hautes fréquences (infra-mensuelles) qui requiert une adaptation des algorithmes classiques.

Mots-clefs (3 à 5) : Statistique - Package

Développement

Introduction

JDemetra+ (Smyk et al. (2024)) est un logiciel open source d’analyse des séries temporelles (ajustement saisonnier, détection d’outliers, nowcasting...). Ce logiciel est recommandé par Eurostat et à tous les instituts du système statistique européen (ESS) depuis 2015. Le développement de JDemetra+ a été motivé par l’envie de proposer aux utilisateurs une interface et un support stable de production et d’études des séries temporelles. Une galerie de packages rjd3 (Palate et al. (2024a)) basé sur la version 3 de JDemetra+ a été développé. Ces packages permettent de faire le lien entre R et JDemetra+. Nous allons présenter ces outils, leur implémentation en R et leur avantage comparatif par rapport aux packages et outils disponibles sous R. L’intérêt essentiel des packages R est l’intégration des fonctionnalités de JDemetra+ dans l’environnement R (à travers la variété des traitements, des objets et des visualisations). Aujourd’hui la diversité des fonctionnalités proposées par les nouveaux packages R en version 3 rendent cet outil polyvalent permettant à la fois du travail d’étude one shot et de test mais aussi la mise en place de structure complexe de production, stable et reproductibles.

Méthode

Les méthodes d’ajustement saisonnier disponibles dans JDemetra+ relèvent des algorithmes classiques X13-Arima (Reg-Arima et décomposition à l’aide de moyennes mobiles) et TramoSeats (décomposition en modèles Arima). Ces algorithmes ont été étendus à des portées plus larges comme l’étude de fréquences non classiques (non mensuelle trimestrielle ou semestrielles) et des hautes fréquences (infra mensuelles). Pour cela deux packages dédiés (**rjd3x11plus** (Palate et al. (2024c)) et **rjd3highfreq** (Palate et al. (2024b))) implémentent des extensions de ces algorithmes.

Ces algorithmes sont écrits en Java. C’est pourquoi les packages R relatifs à JDemetra+ ont été développés à partir de cette structure et de ces routines de base pour ajouter une surcouche de connexion de JDemetra+ à l’environnement R. Cette connexion est effectuée grâce au Protocol buffer (le protocole de sérialisation de Google). Le Protobuf permet de faire le lien entre les programmes Java et l’environnement R. Il est nécessaire de décrire la structure de classe des programmes Java et d’explicitier un moyen de convertir chaque élément

*Insee, tanguy.barthelemy@insee.fr

en objet R. Les objets R seront nativement écrits en S4 mais pourront être réécrits en S3 pour être plus facilement manipulables.

JDemetra+ propose des outils d'estimations performants (modélisation Arima) et des nombreux tests statistiques. Le package **rjd3sts** est une boîte à outils polyvalente pour les modélisations espace état. Elle permet notamment de créer des modèles structurels de base, offrant une procédure d'ajustement saisonnier avec une décomposition explicite et une correction des jours ouvrables variables dans le temps.

Approche

Nous présentons un cas d'application : l'ajustement saisonnier de données haute fréquence. Comme expliqué précédemment, nous utiliserons les packages **rjd3x11plus** et **rjd3highfreq**. Les données haute fréquence sont désormais largement disponibles et de plus en plus utilisées dans la production statistique : données de cartes bancaires, validation des titres de transport, données journalières de suivi d'une épidémie. Ces données sont souvent saisonnières et leur analyse nécessite donc une désaisonnalisation préalable.

Conclusion

Les packages rjd3 offrent au datascientist de nouveaux outils pour l'analyse des séries temporelles en R. Ces packages sont en cours de développement et en cours de consolidation notamment en ce qui concerne les algorithmes les plus novateurs comme ceux consacrés à l'ajustement saisonnier des données haute fréquence.

Références

- Palate, Jean, Tanguy Barthelemy, Alain Quartier-la-Tente, Philippe Charles, Anna Smyk, and Corentin Lemasson. 2024a. "Organisation rjdemetra." GitHub. <https://github.com/rjdemetra>.
- . 2024b. "Package rjd3highfreq." GitHub. <https://github.com/rjdemetra/rjd3highfreq>.
- . 2024c. "Package rjd3x11plus." GitHub. <https://github.com/rjdemetra/rjd3x11plus>.
- Smyk, Anna, Tanguy Barthelemy, Alain Quartier-la-Tente, and Karsten Webel. 2024. *JDemetra+ Documentation*. jdemetra-new-documentation.netlify.app. <https://jdemetra-new-documentation.netlify.app/>.