

Analyse des données — Master 2 IMAFA

TD 1—pré-traitement des données

Andrea G. B. Tettamanzi
Université Nice Sophia Antipolis
andrea.tettamanzi@unice.fr

Jeudi 19 décembre 2017

Résumé

Au cours de cette séance, nous allons préparer les données (essentiellement des séries historiques de prix et parités entre devises) qui nous serviront par la suite. En faisant cela, nous allons toucher avec la main les principales problématiques de la phase de pré-traitement et nous allons appliquer une partie des techniques vues en cours.

1 Introduction

Nous allons analyser des séries historiques financières sur une longue période (plusieurs décennies) : taux de change entre différentes devises, prix de l'or, certaines actions européennes et américaines. En particulier, nous essayerons d'étudier, dans les prochaines séances, si ces séries présentent des patrons cachés qui rendraient possible des « arbitrages stochastiques ». De plus, nous irons tester si ces patrons deviennent plus faciles à trouver si l'on change de devise de référence.

Pour pouvoir faire ce type d'analyse, nous devons nous procurer les données brutes, puis les nettoyer, intégrer, transformer et consolider pour rendre possible l'application des méthodes que nous allons étudier dans les prochaines séances.

Nous nous servirons, pour cette phase, surtout du système R.

2 Ressources nécessaires

On commence par récupérer les données brutes en ligne.

Les prix historiques de l'or, fixés au marché des métaux de Londres, peuvent être téléchargés du site Web de la Federal Reserve Bank de St. Louis : <https://fred.stlouisfed.org/release/tables?rid=256&eid=20467>.

Les taux d'échange de référence historiques de l'euro peuvent être téléchargés du site Web de la Banque Centrale Européenne, à l'onglet Statistiques : <https://www.ecb.europa.eu/stats/exchange/eurofxref/html/index.en.html>. Les mêmes données peuvent être récupérées aussi à partir du site de la Banque de France : <https://www.banque-france.fr/nc/economie-et-statistiques/changes-et-taux/les-taux-de-change-salle-des-marches/parites-quotidien.html>.

Les parités USD/FRF sont disponibles sur le site Web de la Réserve Fédérale des États Unis : <https://www.federalreserve.gov/releases/h10/hist/default1999.htm>, pour la période 1977–1998.

Pour votre commodité, les données brutes des parités des devises (et de l'or, qui peut être considéré à l'instar d'une devise) sont disponibles sur le site Web du cours (ici).

Les séries historiques des prix des actions peuvent être téléchargées au format CSV depuis Yahoo! Finance. En général, pour une action dont le *ticker* est XYZ, les prix historiques se trouvent à l'URL `http://chart.finance.yahoo.com/table.csv?s=XYZ&a=...&b=...&c=...&d=...&e=...&f=...&g=d&ignore=.csv`, où les paramètres veulent dire :

- a* : mois de début (janvier = 0, ..., décembre = 11) ;
- b* : jour du mois de début ;
- c* : année de début (ex. : 1968) ;
- d* : mois de fin (janvier = 0, ..., décembre = 11) ;
- e* : jour du mois de fin ;
- f* : année de fin (ex. : 2016) ;
- g* : fréquence (*d* = *daily*, journalière).

Comme vous le pourrez constater, les données sont ordonnées de la date la plus récente à celle la plus éloignée.

Pour chaque jour d'ouverture du marché où l'action est traitée, vous avez six champs (ou colonnes), qui sont appelés, dans le fichier CSV : Date, Open, High, Low, Close, Volume et Adj Close. Ce dernier contient le prix de clôture (le même que Close), ajusté pour les dividendes, les fractionnements ou les regroupements

3 Nettoyage et intégration

1. Télécharger les données brutes des parités des devises et de l'or.
2. Télécharger les données historiques de l'indice Dow Jones (Industrial Average) (*ticker* `^DJI`), de l'indice FTSE 100 (`^FTSE`), de l'indice CAC 40 (`^FCHI`) et des actions General Electric (`GE`), Exxon Mobil (`XOM`), Procter & Gamble (`PG`), Lloyds (`LLOY.L`), Rolls Royce (`RR.L`), Schneider Electric (`SU.PA`) et Société Générale (`GLE.PA`).
3. Trouver comment charger correctement toutes ces séries historiques dans des *dataframes* de R (suggestion : utiliser les fonctions `read.table` et `read.csv`). Il existe des packages R qui facilitent le traitement des séries historiques (`zoo`, `xts` et `timeSeries`, par exemple : voir le tutoriel sur les séries temporelles financières en R d'Eric Zivot, de l'Université de Washington ; nous allons cependant utiliser des *dataframes* classiques).
4. Vérifier le chargement correct des séries en faisant des graphiques avec la fonction `plot.default` : si *x* est un *dataframe* avec une colonne Date pour les dates et Price pour les prix,

```
plot.default(xDate, xPrice, type="l")
```

produira un graphique de la série. Attention : veillez à ce que les données soient ordonnées par dates croissantes.
5. Nettoyer les séries : traiter les données manquantes, erronées (*outliers*), uniformiser les formats des dates, etc.
6. Transformer les séries des actions en ajustant tous les prix pour tenir compte des paiements des dividendes, les fractionnements et les regroupements (suggestion : comparer les colonnes Close et Adj Close).
7. On veut pouvoir convertir les séries des prix des actions dans des devises autres que celle originale, y compris en or. A ce but, il faut étendre dans le passé la série des prix de l'or en euro en prenant la série des parités du franc français comme référence ; il faut aussi étendre les parités de l'euro avec le dollar et la livre sterling.
8. Convertir les séries dans les différentes devises. Pour chaque série, il nous faudra la version en dollars, en livres, en euros et en or.