Imputation for AR Time Series

Junyan Liu and Daniel P. Palomar
Hong Kong University of Science and Technology (HKUST)
2019-01-29

Contents

This vignette illustrates the usage of the package `imputeFin` for imputation of missing Values in incomplete AR time series.

1 Installation

The package can currently be installed from GitHub:

```r
# install.packages("devtools")
devtools::install_github("dppalomar/imputeFin")
# Getting help
library(imputeFin)
help(package = "imputeFin")
package?imputeFin
?imputeFin
```

2 Usage of the package

2.1 Estimation and imputation for univariate incomplete time series

We start by loading the package.

```r
library(MASS)
library(imputeTS)
library(xts)
library(imputeFin)
```

We download the adjusted price of S&P 500 index from 2012-01-01 to 2015-07-08, and compute the log prices. We delete some log prices to generate missing values.

```r
library(quantmod)
y_orig <- log(Ad(getSymbols("^GSPC", from = "2012-01-01", to = "2015-07-08", auto.assign = FALSE)))
T <- nrow(y_orig)
miss_pct <- 0.2 # the percentage of missing values
n_miss <- floor(miss_pct*T)
idx_miss <- round(T/2) + 1:n_miss
y_miss <- y_orig
y_miss[idx_miss] <- NA
idx_obs <- setdiff(1:T, idx_miss)
```

Then we use the function estimateAR1() to estimate the parameters of AR(1) model from this incomplete time series.

```
library(tictoc)
tic("parameter estimation")
estimation_result <- estimateAR1(y_miss)
toc()
#> parameter estimation: 15.56 sec elapsed
estimation_result$phi0
#> [1] 0.01272511
estimation_result$phi1
#> [1] 0.998373
estimation_result$sigma2
#> [1] 3.799748e-05
estimation_result$nu
#> [1] 5.481531
```

Next, we use the function imputeAR1() to impute the missing values in the incomplete time series.

```
# we generate n_sample = 2 imputed time series
tic("imputation of missing values with unknown parameters")
y_imputed_unknown <- imputeAR1(y_miss, n_sample = 2, n_burn = 200, n_thin = 100)
toc()
#> imputation of missing values with unknown parameters: 21.6 sec elapsed

idx_miss_bis <- (min(idx_miss)-1):(max(idx_miss)+1)
{plot(y_miss, main="original")
lines(y_orig[idx_miss_bis, 1], col="blue", lwd=2)}#plot the original time series
```
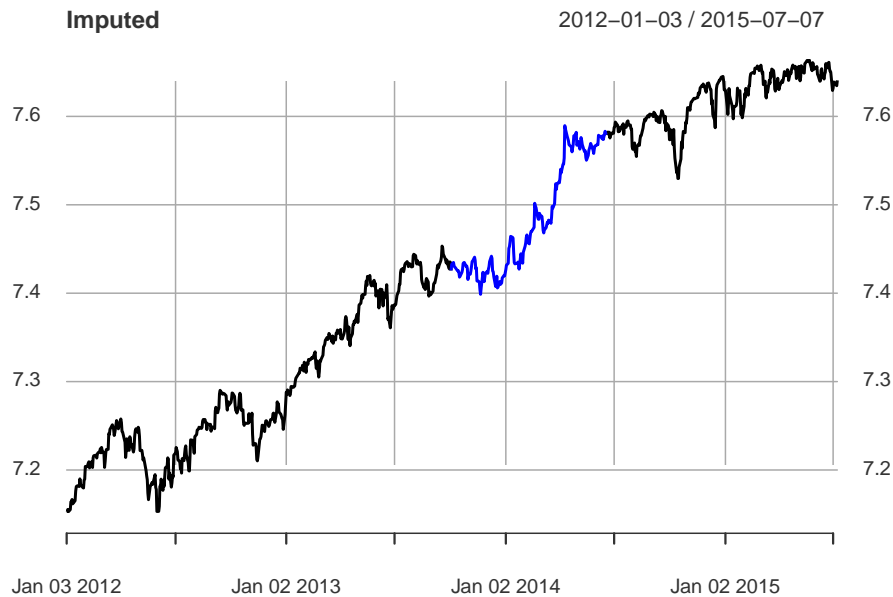

original   2012−01−03 / 2015−07−07
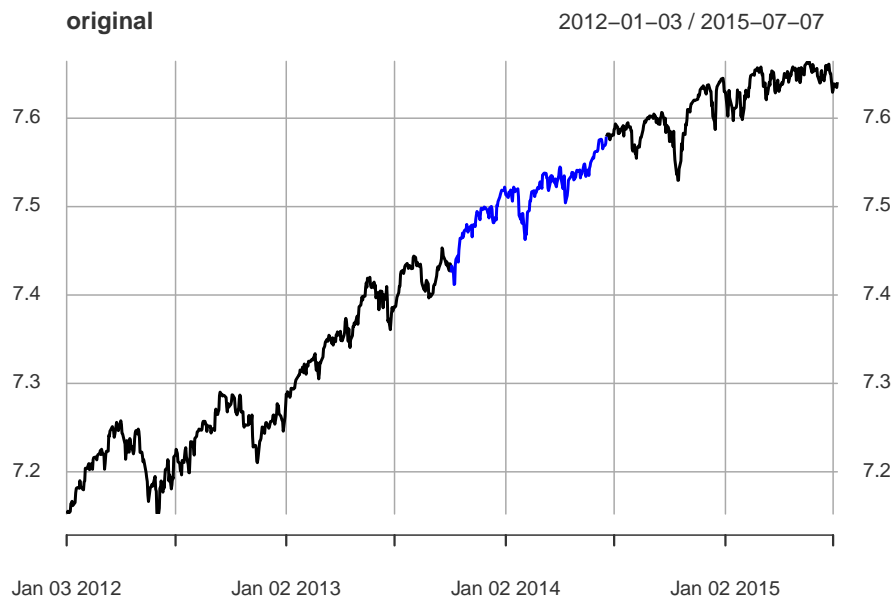
```
{plot(y_miss, main="Imputed")
lines(y_imputed_unknown[idx_miss_bis, 1], col="blue", lwd=2)}#plot the first imputed time series
```

Imputed   2012−01−03 / 2015−07−07
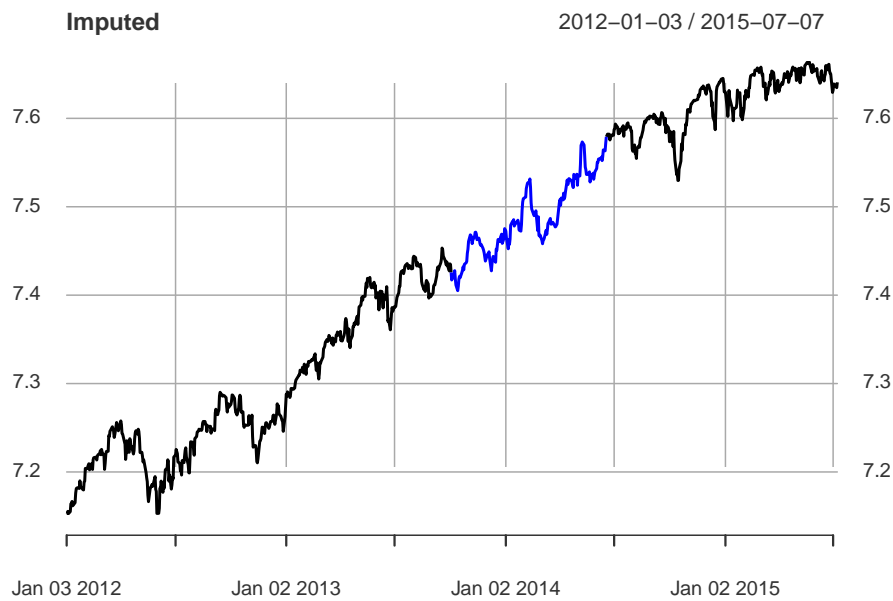
```
{plot(y_miss, main="Imputed")
lines(y_imputed_unknown[idx_miss_bis, 2], col="blue", lwd=2)}#plot the second imputed time series
```


Imputed   2012−01−03 / 2015−07−07

```
# if we have already known the parameters, then we can use this information when doing the imputation.
parameters_estimate = c(estimation_result$phi0, estimation_result$phi1,
                        estimation_result$sigma2, estimation_result$nu)
tic("imputation of missing values with known parameters")
y_imputed_known <- imputeAR1(y_miss, n_sample = 2, n_burn = 200, n_thin = 100,
                             parameters = parameters_estimate)
toc()
#> imputation of missing values with known parameters: 6.22 sec elapsed
```