Should_predictors_be_differenced.R

georges

2025-03-02

Should the predictors be differenced?

Answer: No. The Arima model fits a model in which the the residuals of the regression have an ARIMA structure.

```r
library(cv)
```

```
## Loading required package: doParallel
```

```
## Loading required package: foreach
```

```
## Loading required package: iterators
```

```
## Loading required package: parallel
```

```r
dd <- within(
  data.frame(time = 1:10000),
  {
    x <- rnorm(10000)
    ygen <- arima.sim(model = list(order = c(1,0,1), ar = .9, ma = .7, sd =.1), n = 10000)
    y1 <- cumsum(ygen + x)

    y2 <- cumsum(ygen) + x

  }
)


head(dd)
```

```
##   time        y2        y1      ygen            x
## 1    1  3.507717  3.507717 4.061137 -0.5534198
## 2    2  5.213335  4.659916 2.139596 -0.9873980
## 3    3 10.166086  8.625269 2.776137  1.1892162
## 4    4 12.429070 12.077469 3.581381 -0.1291810
## 5    5 15.069470 14.588687 3.657758 -1.1465396
## 6    6 18.191879 16.564557 2.406858 -0.4309884
```

```r
fit1 <- Arima(y1 ~ x, dd, order = c(1,1,1))
```

```
## Note: 'data' coerced to 'ts_data_frame'
```

```r
fit2 <- Arima(y2 ~ x, dd, order = c(1,1,1))
```

```
## Note: 'data' coerced to 'ts_data_frame'
```

```r
str(fit1)
```

```
## List of 10
##  $ formula    :Class 'formula'  language y1 ~ x
##   .. ..- attr(*, ".Environment")=<environment: R_GlobalEnv>
##  $ data       :Classes 'ts_data_frame' and 'data.frame':   10000 obs. of  5 variables:
##   ..$ time: Time-Series [1:10000] from 1 to 10000: 1 2 3 4 5 6 7 8 9 10 ...
##   ..$ y2  : Time-Series [1:10000] from 1 to 10000: 3.51 5.21 10.17 12.43 15.07 ...
##   ..$ y1  : Time-Series [1:10000] from 1 to 10000: 3.51 4.66 8.63 12.08 14.59 ...
##   ..$ ygen: Time-Series [1:10000] from 1 to 10000: 4.06 2.14 2.78 3.58 3.66 ...
##   ..$ x   : Time-Series [1:10000] from 1 to 10000: -0.553 -0.987 1.189 -0.129 -1.147 ...
##  $ order      : num [1:3] 1 1 1
##  $ seasonal   :List of 2
##   ..$ order : int [1:3] 0 0 0
##   ..$ period: logi NA
##  $ call       : language Arima(formula = y1 ~ x, data = dd, order = c(1, 1, 1))
##  $ model      :'data.frame':   10000 obs. of  2 variables:
##   ..$ y1: num [1:10000] 3.51 4.66 8.63 12.08 14.59 ...
##   ..$ x : num [1:10000] -0.553 -0.987 1.189 -0.129 -1.147 ...
##   ..- attr(*, "terms")=Classes 'terms', 'formula'  language y1 ~ x
##   .. .. ..- attr(*, "variables")= language list(y1, x)
##   .. .. ..- attr(*, "factors")= int [1:2, 1] 0 1
##   .. .. .. ..- attr(*, "dimnames")=List of 2
##   .. .. .. .. ..$ : chr [1:2] "y1" "x"
##   .. .. .. .. ..$ : chr "x"
##   .. .. ..- attr(*, "term.labels")= chr "x"
##   .. .. ..- attr(*, "order")= int 1
##   .. .. ..- attr(*, "intercept")= int 1
##   .. .. ..- attr(*, "response")= int 1
##   .. .. ..- attr(*, ".Environment")=<environment: R_GlobalEnv>
##   .. .. ..- attr(*, "predvars")= language list(y1, x)
##   .. .. ..- attr(*, "dataClasses")= Named chr [1:2] "numeric" "numeric"
##   .. .. .. ..- attr(*, "names")= chr [1:2] "y1" "x"
##  $ dots       : list()
##  $ arima      :List of 14
##   ..$ coef     : Named num [1:3] 0.851 0.708 0.498
##   .. ..- attr(*, "names")= chr [1:3] "ar1" "ma1" "x"
##   ..$ sigma2   : num 1.41
##   ..$ var.coef : num [1:3, 1:3] 2.97e-05 -1.27e-05 -1.46e-08 -1.27e-05 7.41e-05 ...
##   .. ..- attr(*, "dimnames")=List of 2
##   .. .. ..$ : chr [1:3] "ar1" "ma1" "x"
##   .. .. ..$ : chr [1:3] "ar1" "ma1" "x"
##   ..$ mask     : logi [1:3] TRUE TRUE TRUE
##   ..$ loglik   : num -15907
##   ..$ aic      : num 31821
##   ..$ arma     : int [1:7] 1 1 0 0 1 1 0
##   ..$ residuals: Time-Series [1:10000] from 1 to 10000: 0.00378 0.43652 1.34289 0.80671 -0.97833 ...
##   ..$ call     : language stats::arima(x = y, order = order, seasonal = seasonal, xreg = x, include.r
##   ..$ series   : chr "y"
##   ..$ code     : int 0
##   ..$ n.cond   : int 0
##   ..$ nobs     : int 9999
##   ..$ model    :List of 10
##   .. ..$ phi  : num 0.851
##   .. ..$ theta: num 0.708
##   .. ..$ Delta: num 1
```

```
##    .. ..$ Z    : num [1:3] 1 0 1
##    .. ..$ a    : num [1:3] 2.635 0.131 257.084
##    .. ..$ P    : num [1:3, 1:3] 0.00 0.00 4.73e-22 0.00 0.00 ...
##    .. ..$ T    : num [1:3, 1:3] 0.851 0 1 1 0 ...
##    .. ..$ V    : num [1:3, 1:3] 1 0.708 0 0.708 0.502 ...
##    .. ..$ h    : num 0
##    .. ..$ Pn   : num [1:3, 1:3] 1.00 7.08e-01 -2.53e-21 7.08e-01 5.02e-01 ...
##    ..- attr(*, "class")= chr "Arima"
##  $ response     : 'tsp' Named num [1:10000] 3.51 4.66 8.63 12.08 14.59 ...
##    ..- attr(*, "names")= chr [1:10000] "1" "2" "3" "4" ...
##    ..- attr(*, "tsp")= num [1:3] 1 10000 1
##  $ model.matrix: num [1:10000, 1] -0.553 -0.987 1.189 -0.129 -1.147 ...
##    ..- attr(*, "dimnames")=List of 2
##    .. ..$ : chr [1:10000] "1" "2" "3" "4" ...
##    .. ..$ : chr "x"
##  - attr(*, "class")= chr "ARIMA"
```

fit1`$`arima

```
##
## Call:
## stats::arima(x = y, order = order, seasonal = seasonal, xreg = x, include.mean = has.intercept)
##
## Coefficients:
##          ar1      ma1        x
##       0.8512   0.7083   0.4983
## s.e.  0.0054   0.0086   0.0026
##
## sigma^2 estimated as 1.41:  log likelihood = -15906.75,  aic = 31821.5
```

summary(fit1)

```
##
## Call:
## Arima(formula = y1 ~ x, data = dd, order = c(1, 1, 1))
##
## Residuals:
##       Min      1st Q     Median      3rd Q        Max
## -4.443847 -0.789605 -0.001078   0.813760   4.568069
##
## Estimates:
##      Estimate Std. Error z value Pr(>|z|
## ar1 0.851163    0.005447  156.26  <2e-16 ***
## ma1 0.708335    0.008608   82.29  <2e-16 ***
## x   0.498264    0.002595  192.02  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard deviation: 1.187
## Log-likelhood = -15910
## AIC = 31820
```

summary(fit2)

```
##
## Call:
```

```
## Arima(formula = y2 ~ x, data = dd, order = c(1, 1, 1))
##
## Residuals:
##          Min       1st Q      Median      3rd Q         Max
## -3.630e+00 -6.812e-01 -6.926e-06  6.810e-01  3.854e+00
##
## Estimates:
##       Estimate Std. Error z value Pr(>|z|
## ar1 0.893515   0.004580   195.08  <2e-16 ***
## ma1 0.694947   0.007539    92.18  <2e-16 ***
## x   0.998253   0.002173   459.32  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard deviation: 0.9934
## Log-likelhood = -14120
## AIC = 28260
```

```r
dd$r1 <- residuals(fit1)
dd$r2 <- residuals(fit2)

fitr1 <- Arima(~ r1, order = c(1,1,1), dd)
```

```
## Note: 'data' coerced to 'ts_data_frame'
```

```r
fitr2 <- Arima(~ r2, order = c(1,1,1), dd)
```

```
## Note: 'data' coerced to 'ts_data_frame'
```

```r
summary(fitr1)
```

```
##
## Call:
## Arima(formula = ~r1, data = dd, order = c(1, 1, 1))
##
## Residuals:
##        Min     1st Q    Median     3rd Q       Max
## -4.31528 -0.79507 -0.00752   0.80410   4.53558
##
## Estimates:
##        Estimate Std. Error  z value  Pr(>|z|
## ar1 -0.081787   0.010020    -8.163 3.28e-16 ***
## ma1 -0.997788   0.001158 -861.845  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard deviation: 1.184
## Log-likelhood = -15880
## AIC = 31760
```

```r
summary(fitr2)
```

```
##
## Call:
## Arima(formula = ~r2, data = dd, order = c(1, 1, 1))
##
## Residuals:
```

```
##       Min     1st Q    Median    3rd Q       Max
## -3.64269 -0.68870 -0.01149  0.66962  3.85774
##
## Estimates:
##        Estimate Std. Error   z value Pr(>|z|
## ar1 -0.0096665  0.0100001    -0.967   0.334
## ma1 -0.9999995  0.0009887 -1011.385  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard deviation: 0.9934
## Log-likelhood = -14130
## AIC = 28260
```