

Estatística Aplicada

Projecto intercalar

Pedro Magalhães (200202298)

2022-01-03

Declaro que o presente relatório é de minha autoria e não foi utilizado previamente noutro curso ou unidade curricular, desta ou de outra instituição. As referências a outros autores (afirmações, ideias, pensamentos) respeitam escrupulosamente as regras da atribuição, e encontram-se devidamente indicadas no texto e nas referências bibliográficas, de acordo com as normas de referência. Tenho consciência de que a prática de plágio e auto-plágio constitui um ilícito académico.

Introdução

Definição da questão

Este projeto analisa a relação entre a evolução diária da cotação da Microsoft no mercado de valores de Nova York para os anos de 2015 a 2016 e os indicadores de análise técnica mais utilizados com vista a apresentar um modelo de previsão da evolução de curto prazo da cotação.

Este trabalho utiliza a informação de cotação diária das ações da Microsoft na bolsa de Nova York e na informação fundamental divulgada no final de cada ano em análise. Informação publicada no site da Kaggle.

Crítérios de selecção do modelo

Analisaremos o modelo que apresentar o melhor binómio, qualidade de ajustamento, calculado pelo R^2 e parsimónia dado pelo AIC ,

Pressupostos e limitações

Apesar deste projeto estudar a relação linear entre a cotação diária de encerramento e vários indicadores, do ponto de vista da teoria da análise de mercados e da avaliação de carteiras o uso de regressão constitui uma limitação e simplificação. Apesar de válida do ponto de vista académico, na realidade um modelo definido desta forma não permite retirar conclusões adequadas para suportar qualquer conclusão ou decisão.

De entre as limitações e pressupostos assumidos, o mais importante é o pressuposto de **independência entre as observações e consequentemente dos resíduos do modelo**. Dada a natureza dos dados utilizados é de esperar que as observações y_s representando a quotação de fecho do momento S e y_{s-h} representando a observação no momento $S-h$ **não são independentes** estando portante perante uma série temporal.

O estudo dos efeitos auto regressivos contidos na amostra da cotação está fora do âmbito deste projeto pelo que iremos desconsiderar este efeito no modelo e apenas levar em linha de conta sinais externos à série os quais serão dados pelos indicadores técnicos calculados.

Descrição dos dados

```
##          date          close          eps          ema_15
## Min.      :2015-01-02    Min.      :40.29    Min.      :2.12    Min.      :41.48
## 1st Qu.:2015-07-05    1st Qu.:46.23    1st Qu.:2.12    1st Qu.:45.99
## Median :2016-01-02    Median :51.33    Median :2.12    Median :51.42
## Mean     :2016-01-01    Mean     :50.99    Mean     :2.12    Mean     :50.78
## 3rd Qu.:2016-07-02    3rd Qu.:55.72    3rd Qu.:2.12    3rd Qu.:54.87
## Max.      :2016-12-30    Max.      :63.62    Max.      :2.12    Max.      :62.65
##          ema_30          evwma_30          rsi          roc
## Min.      :41.94    Min.      :42.59    Min.      :21.82    Min.      : -0.0970986
## 1st Qu.:45.76    1st Qu.:45.45    1st Qu.:45.19    1st Qu.: -0.0066319
## Median :51.71    Median :51.46    Median :52.76    Median : 0.0001881
## Mean     :50.57    Mean     :50.07    Mean     :53.03    Mean     : 0.0005774
## 3rd Qu.:54.17    3rd Qu.:53.01    3rd Qu.:61.38    3rd Qu.: 0.0076617
## Max.      :61.82    Max.      :60.35    Max.      :84.19    Max.      : 0.0994130
##          momentum          per_30          volatility          per_30_mean
## Min.      : -4.35000    Min.      :19.78    Min.      :0.04814    Length:504
## 1st Qu.: -0.34500    1st Qu.:21.58    1st Qu.:0.16838    Class :character
## Median : 0.01000    Median :24.39    Median :0.21645    Mode  :character
## Mean     : 0.03113    Mean     :23.86    Mean     :0.23926
## 3rd Qu.: 0.36250    3rd Qu.:25.55    3rd Qu.:0.27475
## Max.      : 4.84000    Max.      :29.16    Max.      :0.53066
## diff_close_open
## Length:504
## Class :character
## Mode  :character
##
##
##
```

A base de dados utilizada contém os seguintes dados:

date: índice diário entre 2015-01-02 e 2016-12-30

close: preço de fecho da sessão para a acção

```
## Warning in prettyNum(.Internal(format(x, trim, digits, nsmall, width, 3L, :
## 'big.mark' and 'decimal.mark' are both '.', which could be confusing
```

Cotação de fecho

Microsoft



Fig 1: evolução da cotação de fecho das ações da Microsoft no mercado de valores de NY em 2015 e 2016

```
## <ScaleContinuousDate>
## Range:
## Limits: 1.64e+04 -- 1.72e+04
```

low: mínimo diário a que a ação foi transaccionada

high: máximo diário a que a ação foi transaccionada

volume: volume de transacções diário

eps: Earning per Share anual conforme reportado pelas demonstrações financeiras de final do ano e usando o seguinte rácio

$$EPS = \frac{\text{Net Income}}{\text{Outstanding Shares}}$$

ema_15 e ema_30 Exponential Moving Average para a janela de 15 e 30 dias respectivamente.

```
## Warning in prettyNum(.Internal(format(x, trim, digits, nsmall, width, 3L, :
## 'big.mark' and 'decimal.mark' are both '.', which could be confusing
```

Cotação de fecho

Microsoft

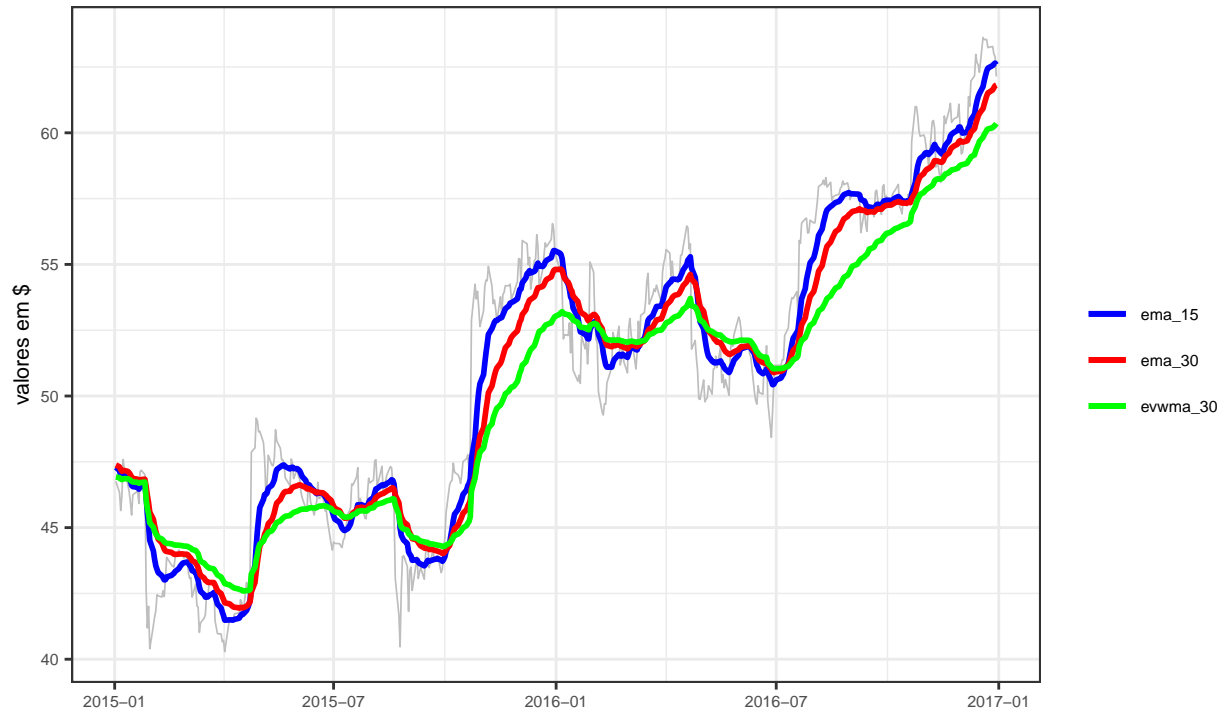


Fig 1: evolução da cotação de fecho das ações da Microsoft no mercado de valores de NY em 2015 e 2016

```
## <ScaleContinuousDate>
## Range:
## Limits: 1.64e+04 -- 1.72e+04
```

ewvma_30 Exponential Moving Average Volume Weigthed. Semelhante ao indicador anterior mas corrigido pelo volume de transacção diário

rsi: Relative Strength Index mensura a força do movimento de subida ou descida em face de um intervalo de evolução. Diferenças entre a evolução deste indicador e a cotação podem indiciar uma inversão da tendência

roc: Percentagem de diferença entre n observações

momentum: Diferença entre n observações (lag)

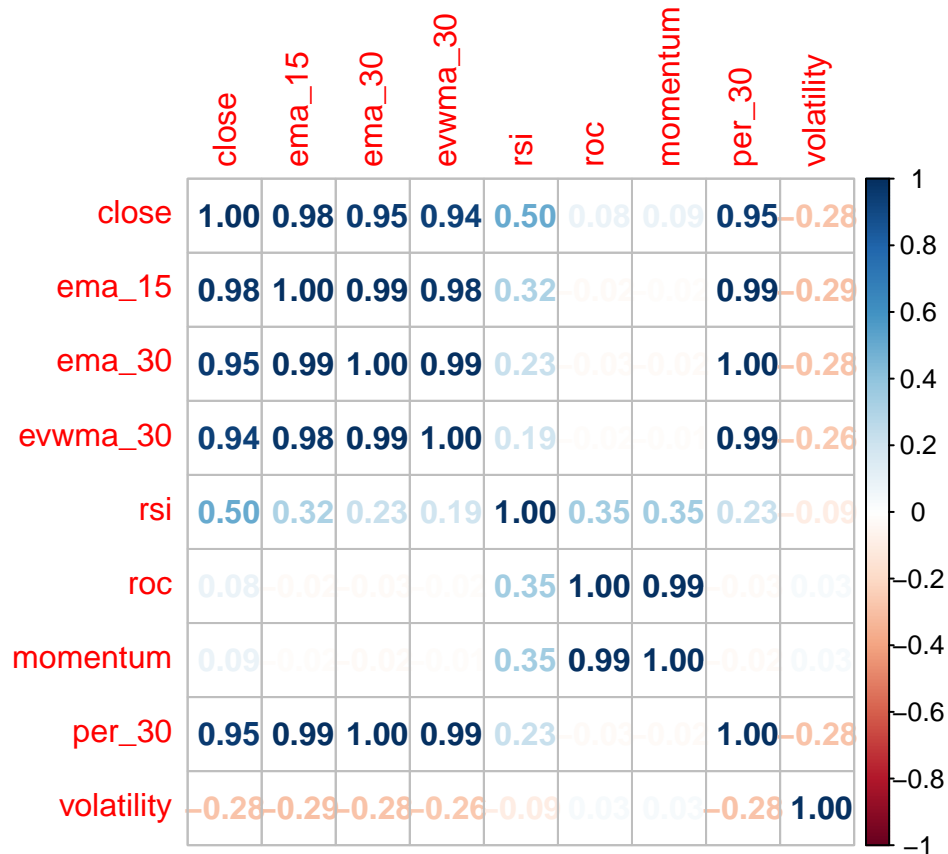
per_30: Price Earning Ratio 30 days trailing

per_30_mean: Variável categórica que assume valores se per_30 for superior à média dos dois anos então "Higher than average" caso contrário "Lower than average"

volatility: Close-to-Close Volatility,

diff_close_open: Variável categórica para se ocorreu um ganho intradia (Close > Open) ou uma perda (Close < Open)

Matriz de correlação



Modelos e avaliação

```
## Warning in predict.lm(model, data): prediction from a rank-deficient fit may be
## misleading
```

```
## # A tibble: 11 x 4
##   model                                model_name      r2    aic
##   <chr>                                <chr>          <dbl> <dbl>
## 1 close ~ ema_15 + ema_30 + evwma_30 + rsi + roc +~ complete    9.97e-1  361.
## 2 close ~ ema_15                        ema_15        9.52e-1 1700.
## 3 close ~ ema_30                        ema_30        9.08e-1 2022.
## 4 close ~ per_30                        per_30        9.08e-1 2022.
## 5 close ~ evwma_30                      evwma_30      8.78e-1 2167.
## 6 close ~ per_30_mean                    per_30_mean   6.78e-1 2656.
## 7 close ~ rsi                           rsi           2.52e-1 3080.
## 8 close ~ volatility                     volatility     7.91e-2 3185.
## 9 close ~ momentum                       momentum      7.97e-3 3222.
## 10 close ~ roc                           roc           7.19e-3 3223.
## 11 close ~ diff_close_open               diff_close_o~ 8.92e-4 3226.
```

O modelo mais completo é neste caso também o modelo mais parsimonio

```
##
## Call:
## lm(formula = ., data = msft_prices_2015_2016_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.19962 -0.19045  0.00417  0.18250  1.68574
##
## Coefficients: (1 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -6.694400   0.412921  -16.212 < 2e-16 ***
## ema_15          0.914772   0.051964   17.604 < 2e-16 ***
## ema_30         -0.425631   0.080679   -5.276 1.98e-07 ***
## evwma_30        0.530109   0.038779   13.670 < 2e-16 ***
## rsi             0.116420   0.002815   41.354 < 2e-16 ***
## roc             0.361779   9.085868    0.040 0.968254
## momentum        0.229818   0.184550    1.245 0.213616
## per_30           NA         NA         NA     NA
## volatility      -0.301320   0.152295   -1.979 0.048423 *
## per_30_meanLower than average 0.104737   0.068662    1.525 0.127800
## diff_close_openintraday loss 0.132513   0.036578    3.623 0.000322 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.342 on 494 degrees of freedom
## Multiple R-squared:  0.9967, Adjusted R-squared:  0.9967
## F-statistic: 1.664e+04 on 9 and 494 DF,  p-value: < 2.2e-16

##
## Call:
## lm(formula = close ~ ema_15 + per_30_mean, data = msft_prices_2015_2016_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.6686 -0.6776  0.0274  0.6901  5.3115
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -5.59595   1.12051  -4.994 8.18e-07 ***
## ema_15          1.10626   0.02034  54.377 < 2e-16 ***
## per_30_meanLower than average 0.96067   0.22934   4.189 3.31e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.28 on 501 degrees of freedom
## Multiple R-squared:  0.9533, Adjusted R-squared:  0.9531
## F-statistic: 5111 on 2 and 501 DF,  p-value: < 2.2e-16
```

Questões sobre Regressão Linear

Formula do modelo utilizado:

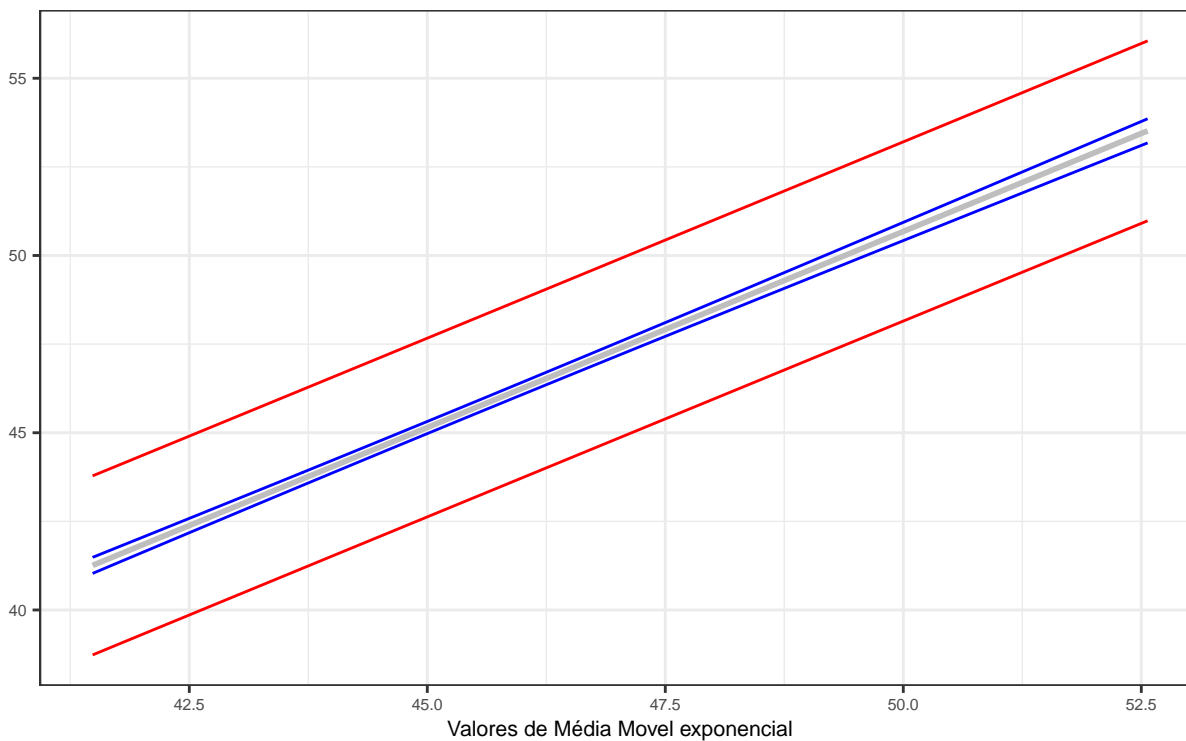
$$y = -5.59595 + 1.10626x_{ema15} + 0.96067x_{per30}$$

e.1) Mantendo tudo o resto constante, um aumento de 1 USD na média móvel exponencial a 15 dias representa um aumento de 1,11 USD no valor spot. Mantendo tudo o resto constante em média cotações spot superiores ao price earning ratio têm um acréscimo de 0,9 USD no preço.

e.2.)

Previsões modelo com banda de confiança e previsão

Azul banda de confiança, vermelha banda de estimação



e3)

e4)

```
## [1] 11.16853
```

```
##      1
## 7.720056
```

Um aumento de dois desvios padrão da média móvel exponencial implicaria o aumento de 7,72 USD no valor e fecho diário da cotação.

e5)

```
##
## Call:
## lm(formula = close ~ ema_15 * per_30_mean, data = msft_prices_2015_2016_df)
##
## Residuals:
```

```

##      Min      1Q  Median      3Q      Max
## -4.6473 -0.6491  0.0086  0.6922  5.2271
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   -3.70261     1.28912  -2.872  0.00425 **
## ema_15                        1.07181     0.02342  45.766 < 2e-16 ***
## per_30_meanLower than average -5.44975     2.21807  -2.457  0.01435 *
## ema_15:per_30_meanLower than average 0.13438     0.04625   2.905  0.00383 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.271 on 500 degrees of freedom
## Multiple R-squared:  0.9541, Adjusted R-squared:  0.9538
## F-statistic: 3461 on 3 and 500 DF, p-value: < 2.2e-16

```

Verificamos que a interacção entre as duas variáveis é estatisticamente significativa. Tendo a variável `per_30_mean` apenas duas categorias, para o caso de ser “Lower than average” a interacção é zero e um aumento de 1 USD de `ema_15` resulta num aumento de 1.1 USD no entanto, caso contrário o efeito de 1 USD em `ema_15` representa um aumento de $1.1 + 0.13$.