


工程統計期末報告

科系年級__資工大二__學號__109550178__姓名__黃昱翰__

(總分 105) = 摘要+動機+資料+方法+結果與討論+圖與表+參考文獻

與資料來源+課程建議

報告題目	歐洲股市和經濟
報告學生 (照片與姓名)	 黃昱翰
Abstract 摘要(10%)	<p>歷史上股票市場和整體經濟總是呈現相關性。當股市上漲時，我們往往可以期待經濟迎來好轉。但會一直如此嗎？股票市場將如何影響經濟，股市崩盤對整體經濟環境意味著什麼？對交易員而言關鍵要點是什麼？</p> <p>股票市場占據重要地位的原因很多，它為交易員和投資者從其波動獲益提供機會並創造個人財富，同時其為國家的商業的工業健康長度提供基準，也為企業提供擴大規模和步入繁榮昌盛提供機會進而令整體經濟受益。因此，一個運行良好的股票市場對企業、個人及國家都是非常有價值的。</p> <p>股票市場通常被視為可靠的經濟晴雨表。其揭露了主要公司的經營狀況進而讓市場洞察到經濟健康驅動因素如消費者支出。股票價格上漲意味著企業和消費者信心上升，反之亦然，股票價格下跌意味著企業和消費者信心下滑。</p> <p>倘若一支股指處於牛市之中例如以科技股為中心的納斯達克指數，這可能意味著一系列事情如：投資者對電子產品的需求充滿信心，同時投資者對如微軟和蘋果等對指數有較大影響力的科技巨頭的財務實力也充滿信心。</p> <p>以上都可能令指數出現波動，並且信心本身就能培育信心。不過，僅僅因為股市上漲就認定整體經濟正在改善，或者僅僅因為股市下跌就認定經濟正在萎縮，這並不全然正確。</p> <p>同樣地，經濟對股票市場有著意義重大的影響，重要的基本面驅動因素如非農就業報告、大選、利率水平、通脹和自然災害等都有能力影響股票價格。</p>

<p>Motivation 動機(10%)</p>	<p>藉由過去歐洲股票收盤價的歷史資料，我想看看實際的股票狀況是不是真的和整體經濟有很大的相關性，由於給定的資料只有從 1991-1998，而且也並未提供日期，所以我想是不是能夠蒐集到更多的歷史資料，來看看能不能從 1991-1998 的這段資料去預測到未來股票的走向，<u>並且看看股票價格劇烈浮動的時候，是不是正好有甚麼歷史事件發生，因此導致股票大跌。</u></p> <p>把資料視覺化之後，可以從圖(一)看出隨著時間的遞進，雖說價格一直有在浮動，但長期來看的話是有明顯的成長，當然可能跟當時的幣值有關，但這暫且不討論，單論價格和時間，我們是可以看出兩者是有正相關。假如我們可以找到符合資料走向的模型，就可以在的誤差內預測未來股票的價格。</p>
<p>Data 資料(15%)</p>	<p>原始資料使用的是歐洲主要股票指數 1991-1998 的平日收盤價（不包括假日或節慶日），四國有德國(DAX)、瑞士(SMI)、法國(CAC)、英國(FTSE)，資料由 Erste Bank AG, Vienna, Austria 所提供。由於歐洲國家一年工作天約 200 多天甚至更少，所以 9 年來總共有 1860 筆資料。</p> <p>之後我上網把四個股票指數從原本的資料後到 2022/1/14 的資料都載下來延伸做資料分析，資料取自 WSJ MARKET 和 yahoo finance，由於有些日期的資料並沒有紀錄，所以我最後只取四個國家都有資料的日期，共有 7735 筆資料。</p>
<p>Methodology 方法(10%)</p>	<p>觀察資料視覺化後劇烈浮動的時期，查找當時世界上是否有發生甚麼大事事件影響了股市，並透過視覺化資料來探討股市價格和時間的相關性</p> <p>透過回歸分析分別對四個變數做線性和非線性回歸來預測未來的資料，再透過從網路取得的資料進行比較，對比兩組資料回歸線的走向是不是大致相同、方程式係數差異大不大、回歸線和資料的擬合程度。</p>

Results &
Discussion
結果與討論
(30%)

從圖一可以看出四個股票的走向幾乎是相同的，只是幅度有所差別，其中瑞士的成長幅度最大，再來是德國、英國，法國最後。

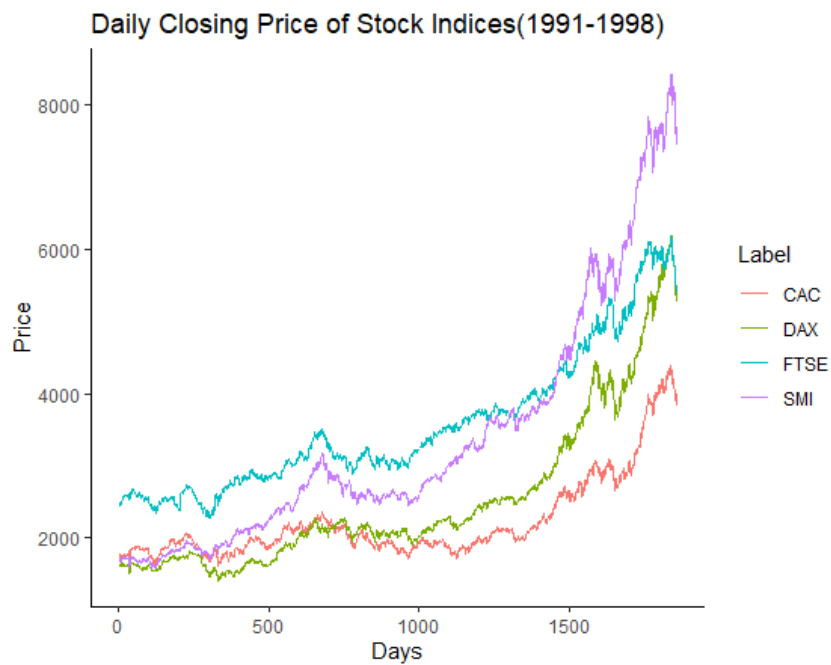
再來對比圖二，從長期來來看，價格依然是隨時間遞增。其中兩個比較劇烈下跌的時期差不多是 2000~2003 年和 2007~2009 年。正好是網路科技泡沫和次貸危機時期。靠近 8000 有一個劇烈下跌的時期就是 2020 疫情爆發的時期，相比 2000 年和 2008 年的經濟事件算是滿快就回復穩定的。後來對比網路上的資料發現也確實是如此，疫情帶來的恐慌以及石油價格暴跌，高點至下次漲回高點所費時間大約只花了 90 天，相比於網路泡沫的 6 年 9 個月和次貸危機 2 的 5 年 2 個月確實很快就穩定了。

從圖三可以看到原始資料四組的的線性回歸的擬合程度除了 CAC 其他都還算高，從數據上看也確實如此(數據在最後一頁)，除了 CAC 的 0.5 其他三個的 r^2 都有 0.7 以上。再來使用多項式回歸可以看出和真實資料非常貼近，4 個變數的 r^2 都有 0.9 以上。對比 4 個多項式方程式的係數除了 4 次方的係數有點差距，其他大致上都一樣，這不難理解，畢竟原始資料 4 組的資料價格幅度和走向並沒有太大的差異。

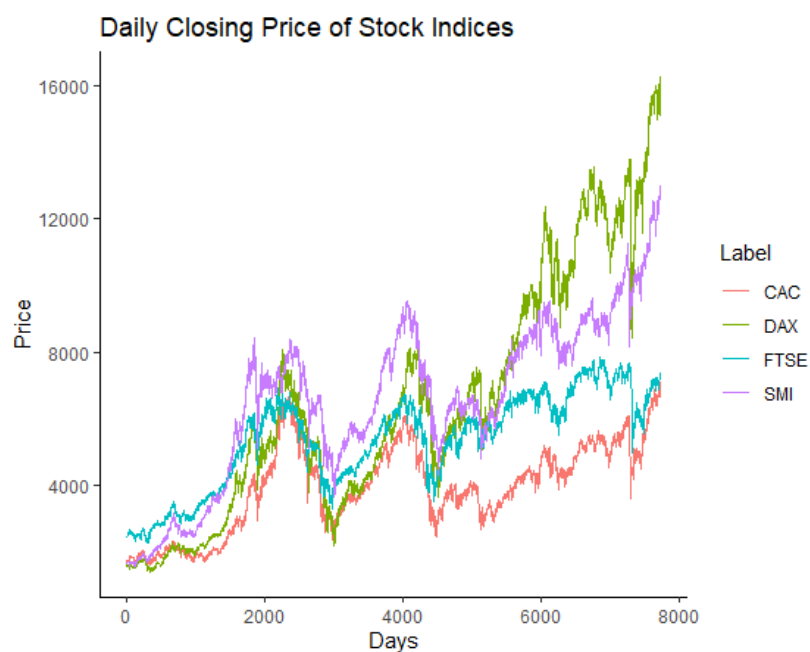
圖四是利用更新過後的資料做的回歸線，可以看出來多項式回歸也沒有辦法像原始資料一樣很好地擬合，透過最後一頁的數據可以看出更新過後的 4 個多項式方程式差異有點大，甚至正負號都有些不同。藉由兩組資料的對比我們可以看出雖然以長期來看兩組資料的走向，價格都是隨時間增加，但兩者中間的差異卻非常大，因此，用原始資料來預測未來的情況在短期內可能會不如預期，但把時間擴大到一定的程度，或許會滿接近預測的模型。

從前面的討論來看，經濟和股市似乎一定是正相關，但其實並不盡然，事實上股市和經濟經常會發生「脫鉤」的情況，很多宏觀統計數據，比如失業率、GDP 等，由於需要時間做統計，因此都有一定的滯後性，滯後的時間從一個月到一個季度不等。像 GDP 這樣的數據，甚至在幾年後還可能會被修正。因此當我們從新聞裡讀到這些數據時，其實只不過是對過去一段時間經濟活動的總結。但是股市可不會這麼運作。說到底，為了對一家公司股票進行合理的估值，投資者需要預測其未來的現金流，然後把這些未來預期的現金流折現到現在，然後算出其合理的內在價值。所以，股市有一定的前瞻性，往往能夠提早預測經濟衰退的到來，同時也會比經濟更早見底。

Figure &
Table
圖與表(20%)

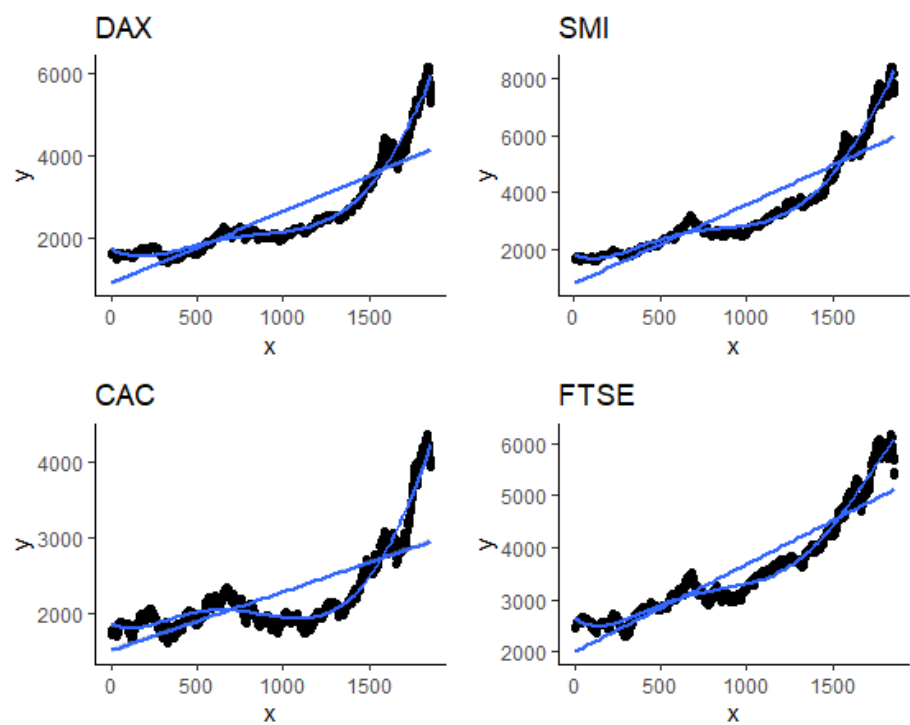


圖一. 將給定的資料視覺化

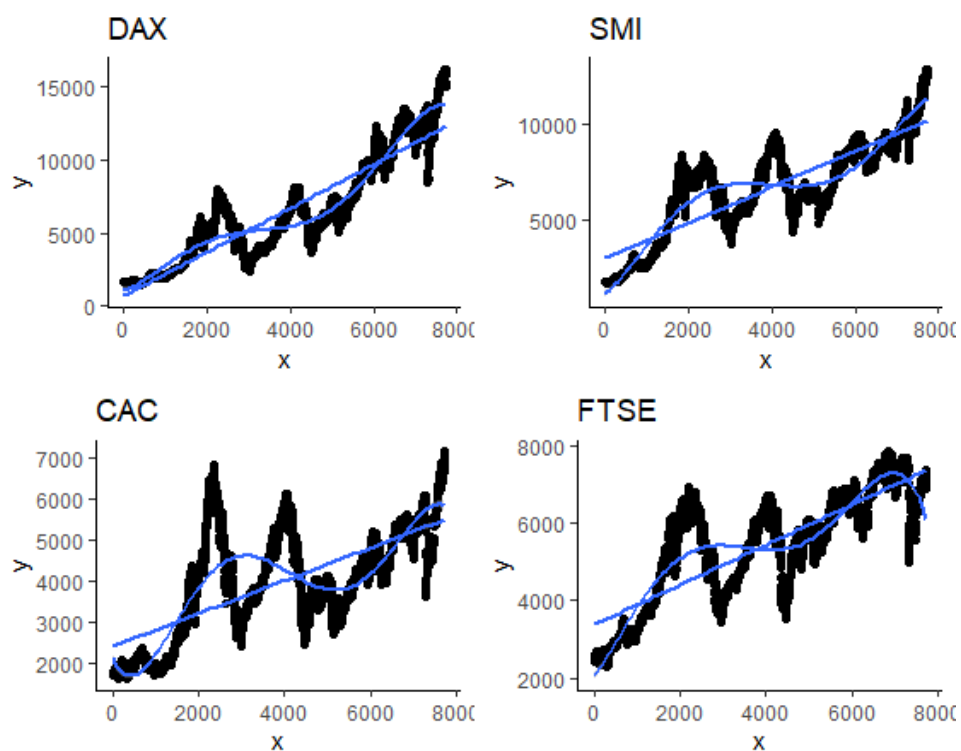


圖二. 從給定的資料(1988)到 2022/1/14 的資料視覺化

Figure &
Table
圖與表(20%)



圖三. 分別對 4 個變數做線性和非線性回歸



圖四. 分別對更新過後的 4 個變數做線性和非線性回歸

References & Resources 參考文獻與資料來源 (5%)	股市與經濟的關係-伍治堅 一文看懂股票市場如何影響經濟-Lisa Yin 20 世紀歷史上的股市大跌：回顧美股百年來重大的 7 次股災-Mr. Market 【股市大崩盤有什麼前兆？】5 個歷史上的重大股災和金融海嘯事件-慢活夫妻 Nonlinear Regression Essentials in R: Polynomial and Spline Regression Models-Alboukadel Kassambara Data Resource-WSJ Market and Yahoo Finance
Suggestion 課程建議 (5%)	我覺得對於統計方面的概念可以再更多一點，因為我在寫回歸分析的時候，發現自己對於統計的概念不是很熟，只是知道怎麼寫 code，雖然可能是我平時沒有練習的關係，謝謝助教們和教授。

最後附上 4 個變數原本的資料和更新後的資料做回歸線時的到的數值

DAX

original

update

<pre>[1] "DAX" [1] "Linear Regression" RMSE R2 1 531.8742 0.7304552 [1] "Polynomial Regression" RMSE R2 1 159.5502 0.9752398 Call: lm(formula = y ~ poly(x, 5, raw = TRUE), data = train.data) Residuals: Min 1Q Median 3Q Max -659.03 -97.26 -10.25 101.34 710.07 Coefficients: Estimate Std. Error t value Pr(> t) (Intercept) 1.769e+03 2.447e+01 72.30 <2e-16 *** poly(x, 5, raw = TRUE)1 -2.898e+00 2.653e-01 -10.93 <2e-16 *** poly(x, 5, raw = TRUE)2 1.201e-02 8.849e-04 13.58 <2e-16 *** poly(x, 5, raw = TRUE)3 -1.586e-05 1.207e-06 -13.14 <2e-16 *** poly(x, 5, raw = TRUE)4 8.602e-09 7.153e-10 12.03 <2e-16 *** poly(x, 5, raw = TRUE)5 -1.477e-12 1.530e-13 -9.66 <2e-16 *** --- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Residual standard error: 159.2 on 1482 degrees of freedom Multiple R-squared: 0.9792, Adjusted R-squared: 0.9791 F-statistic: 1.394e+04 on 5 and 1482 DF, p-value: < 2.2e-16</pre>	<pre>[1] "DAX" [1] "Linear Regression" RMSE R2 1 1542.313 0.831555 [1] "Polynomial Regression" RMSE R2 1 1109.746 0.91281 Call: lm(formula = y ~ poly(x, 5, raw = TRUE), data = train.data) Residuals: Min 1Q Median 3Q Max -5008.8 -753.9 -53.6 645.0 3388.4 Coefficients: Estimate Std. Error t value Pr(> t) (Intercept) 1.134e+03 8.373e+01 13.540 < 2e-16 *** poly(x, 5, raw = TRUE)1 7.868e-01 2.186e-01 3.598 0.000323 *** poly(x, 5, raw = TRUE)2 1.570e-03 1.750e-04 8.970 < 2e-16 *** poly(x, 5, raw = TRUE)3 -8.355e-07 5.739e-08 -14.557 < 2e-16 *** poly(x, 5, raw = TRUE)4 1.493e-10 8.187e-12 18.240 < 2e-16 *** poly(x, 5, raw = TRUE)5 -8.498e-15 4.218e-16 -20.149 < 2e-16 *** --- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Residual standard error: 1097 on 6186 degrees of freedom Multiple R-squared: 0.9104, Adjusted R-squared: 0.9103 F-statistic: 1.257e+04 on 5 and 6186 DF, p-value: < 2.2e-16</pre>
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

SMI

original

update

<pre>[1] "SMI" [1] "Linear Regression" RMSE R2 1 752.7553 0.7895758 [1] "Polynomial Regression" RMSE R2 1 181.3819 0.9877637 Call: lm(formula = y ~ poly(x, 5, raw = TRUE), data = train.data) Residuals: Min 1Q Median 3Q Max -773.62 -121.07 -22.69 111.87 781.88 Coefficients: Estimate Std. Error t value Pr(> t) (Intercept) 1.808e+03 3.237e+01 55.871 <2e-16 *** poly(x, 5, raw = TRUE)1 -2.936e+00 3.486e-01 -8.425 <2e-16 *** poly(x, 5, raw = TRUE)2 1.767e-02 1.157e-03 15.270 <2e-16 *** poly(x, 5, raw = TRUE)3 -2.649e-05 1.575e-06 -16.820 <2e-16 *** poly(x, 5, raw = TRUE)4 1.586e-08 9.328e-10 17.004 <2e-16 *** poly(x, 5, raw = TRUE)5 -3.082e-12 1.996e-13 -15.440 <2e-16 *** --- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Residual standard error: 203.5 on 1482 degrees of freedom Multiple R-squared: 0.9852, Adjusted R-squared: 0.9851 F-statistic: 1.971e+04 on 5 and 1482 DF, p-value: < 2.2e-16</pre>	<pre>[1] "SMI" [1] "Linear Regression" RMSE R2 1 1341.076 0.7068944 [1] "Polynomial Regression" RMSE R2 1 1008.755 0.8342668 Call: lm(formula = y ~ poly(x, 5, raw = TRUE), data = train.data) Residuals: Min 1Q Median 3Q Max -3154.2 -754.0 -80.3 750.4 2782.3 Coefficients: Estimate Std. Error t value Pr(> t) (Intercept) 1.075e+03 8.004e+01 13.427 < 2e-16 *** poly(x, 5, raw = TRUE)1 1.976e+00 2.071e-01 9.540 < 2e-16 *** poly(x, 5, raw = TRUE)2 1.199e-03 1.648e-04 7.275 3.88e-13 *** poly(x, 5, raw = TRUE)3 -6.926e-07 5.384e-08 -12.865 < 2e-16 *** poly(x, 5, raw = TRUE)4 1.128e-10 7.658e-12 14.735 < 2e-16 *** poly(x, 5, raw = TRUE)5 -5.783e-15 3.935e-16 -14.696 < 2e-16 *** --- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Residual standard error: 1025 on 6186 degrees of freedom Multiple R-squared: 0.8308, Adjusted R-squared: 0.8307 F-statistic: 6075 on 5 and 6186 DF, p-value: < 2.2e-16</pre>
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

CAC

original

update

```
[1] "CAC"
[1] "Linear Regression"
      RMSE      R2
1 401.2764 0.5430769
[1] "Polynomial Regression"
      RMSE      R2
1 136.1628 0.9475365

Call:
lm(formula = y ~ poly(x, 5, raw = TRUE), data = train.data)

Residuals:
    Min       1Q   Median       3Q      Max
-473.47  -87.63    0.57   94.70  319.46

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.855e+03  2.060e+01  90.056 < 2e-16 ***
poly(x, 5, raw = TRUE)1 -1.219e+00  2.252e-01  -5.411 7.29e-08 ***
poly(x, 5, raw = TRUE)2  7.321e-03  7.578e-04   9.660 < 2e-16 ***
poly(x, 5, raw = TRUE)3 -1.169e-05  1.040e-06 -11.243 < 2e-16 ***
poly(x, 5, raw = TRUE)4  6.958e-09  6.187e-10  11.246 < 2e-16 ***
poly(x, 5, raw = TRUE)5 -1.291e-12  1.327e-13  -9.722 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 136.8 on 1482 degrees of freedom
Multiple R-squared:  0.944,    Adjusted R-squared:  0.9438
F-statistic: 4995 on 5 and 1482 DF,  p-value: < 2.2e-16
```

```
[1] "CAC"
[1] "Linear Regression"
      RMSE      R2
1 980.2306 0.4559801
[1] "Polynomial Regression"
      RMSE      R2
1 751.9382 0.6799537

Call:
lm(formula = y ~ poly(x, 5, raw = TRUE), data = train.data)

Residuals:
    Min       1Q   Median       3Q      Max
-2170.41  -437.39    -7.22   352.19  2563.87

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.114e+03  5.716e+01  36.99 < 2e-16 ***
poly(x, 5, raw = TRUE)1 -2.229e+00  1.493e-01 -14.93 < 2e-16 ***
poly(x, 5, raw = TRUE)2  3.428e-03  1.194e-04  28.71 < 2e-16 ***
poly(x, 5, raw = TRUE)3 -1.266e-06  3.909e-08 -32.40 < 2e-16 ***
poly(x, 5, raw = TRUE)4  1.812e-10  5.565e-12  32.56 < 2e-16 ***
poly(x, 5, raw = TRUE)5 -8.914e-15  2.861e-16 -31.16 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 742.1 on 6186 degrees of freedom
Multiple R-squared:  0.6815,    Adjusted R-squared:  0.6813
F-statistic: 2648 on 5 and 6186 DF,  p-value: < 2.2e-16
```

FTSE

original

update

```
[1] "FTSE"
[1] "Linear Regression"
      RMSE      R2
1 376.4158 0.8483714
[1] "Polynomial Regression"
      RMSE      R2
1 144.9473 0.9775705

Call:
lm(formula = y ~ poly(x, 5, raw = TRUE), data = train.data)

Residuals:
    Min       1Q   Median       3Q      Max
-694.47  -95.21    6.77   91.05  434.31

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.654e+03  2.357e+01  112.62 < 2e-16 ***
poly(x, 5, raw = TRUE)1 -2.733e+00  2.560e-01 -10.68 < 2e-16 ***
poly(x, 5, raw = TRUE)2  1.352e-02  8.519e-04  15.87 < 2e-16 ***
poly(x, 5, raw = TRUE)3 -1.941e-05  1.160e-06 -16.73 < 2e-16 ***
poly(x, 5, raw = TRUE)4  1.165e-08  6.871e-10  16.95 < 2e-16 ***
poly(x, 5, raw = TRUE)5 -2.367e-12  1.469e-13 -16.11 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 149.2 on 1482 degrees of freedom
Multiple R-squared:  0.9769,    Adjusted R-squared:  0.9768
F-statistic: 1.251e+04 on 5 and 1482 DF,  p-value: < 2.2e-16
```

```
[1] "FTSE"
[1] "Linear Regression"
      RMSE      R2
1 838.2864 0.6542519
[1] "Polynomial Regression"
      RMSE      R2
1 675.1198 0.7757501

Call:
lm(formula = y ~ poly(x, 5, raw = TRUE), data = train.data)

Residuals:
    Min       1Q   Median       3Q      Max
-2069.91  -427.48   37.68  405.07  1694.21

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.068e+03  5.006e+01  41.299 < 2e-16 ***
poly(x, 5, raw = TRUE)1  1.424e+00  1.305e-01  10.915 < 2e-16 ***
poly(x, 5, raw = TRUE)2  7.929e-04  1.046e-04   7.581 3.92e-14 ***
poly(x, 5, raw = TRUE)3 -5.596e-07  3.430e-08 -16.315 < 2e-16 ***
poly(x, 5, raw = TRUE)4  1.060e-10  4.890e-12  21.676 < 2e-16 ***
poly(x, 5, raw = TRUE)5 -6.316e-15  2.516e-16 -25.101 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 654.6 on 6186 degrees of freedom
Multiple R-squared:  0.786,    Adjusted R-squared:  0.7858
F-statistic: 4543 on 5 and 6186 DF,  p-value: < 2.2e-16
```