

# R: $\alpha$ , and $\beta$

justina colmena

April 14, 2018

## Contents

<b>1 Prerequisites</b>	<b>1</b>
<b>2 Getting started</b>	<b>2</b>
<b>3 Fetching the data</b>	<b>2</b>
3.1 Time period for analysis . . . . .	2
3.2 S&P 500: benchmark $\beta = 1$ . . . . .	3
3.3 Fed Funds rate: benchmark $\beta = 0$ . . . . .	3
3.4 Stock ticker data . . . . .	4
<b>4 Analyzing the data</b>	<b>4</b>
4.1 Fitting a linear model . . . . .	4
4.2 Free $\alpha$ model . . . . .	6
4.3 Forced $\alpha = 0$ model . . . . .	6
<b>5 Interpreting the results</b>	<b>7</b>
<b>References</b>	<b>10</b>

## List of Tables

## List of Figures

1 Plot of price level data . . . . .	5
2 Scatter plot with lines of best fit from linear model . . . . .	8

## 1 Prerequisites

Before starting this project, we were somewhat familiar with R [13] and L<sup>A</sup>T<sub>E</sub>X[10], as well as with Donald Knuth’s philosophy of “literate programming” [9], but we had not yet used the “Sweave” function [5] of R, whose official documentation seems to have disappeared along

with its author [11]. Fortunately for our purposes, we found an easy Sweave template [8] online, which we adapted into this paper.

## 2 Getting started

We are going to calculate the beta of a stock using free data [4, 12] in R [13]. Nowadays it is fashionable to use APIs [1, 3] to fetch the data for analysis, which is very convenient once it is set up, even though it is somewhat technical and a little bit of work to get it working.

Once the API packages are installed, we must load them from the library in order to be able to use them in our R session. We also need to load our API account keys for access credentials to use the data.

```
> library('dplyr')
> library('xts')
> library('fredr')
> library('Quandl')
> source('my_api_keys.R')
> fredr_key(my_fred_api_key)
> Quandl.api_key(my_quandl_api_key)
```

## 3 Fetching the data

To calculate the beta of a stock, we need a major market index, an index of returns on a so-called “risk-free interest rate” and returns for our stock of interest. Let us fetch some data for the S&P 500 Index, the Federal Funds rate, and our stock of interest [2].

### 3.1 Time period for analysis

Assign an arbitrary `start_date` and an `end_date`.

```
> today <- Sys.Date()
> start_date <- today - 184
> end_date <- today
> start_date
```

```
[1] "2017-10-12"
```

```
> end_date
```

```
[1] "2018-04-14"
```

### 3.2 S&P 500: benchmark $\beta = 1$

The S&P 500 is a common performance benchmark for  $\beta = 1$  in the stock market.

```
> fetch_fred_sp500 <- fredr_series(series_id = 'SP500',
+   observation_start=start_date, observation_end=end_date)
> sp500 <- data.frame(date=index(fetch_fred_sp500),
+   coredata(fetch_fred_sp500))
> sp500[1:3,]

      date    SP500
1 2017-10-12 2550.93
2 2017-10-13 2553.17
3 2017-10-16 2557.64
```

### 3.3 Fed Funds rate: benchmark $\beta = 0$

We also need a so-called “risk-free” rate of interest as a benchmark for  $\beta = 0$ .

```
> fetch_fred_ffr <- fredr_series(series_id = 'DFF',
+   observation_start = start_date, observation_end = end_date)
> ffr <- data.frame(date=index(fetch_fred_ffr),
+   coredata(fetch_fred_ffr))
> ffr[1:3,]

      date  DFF
1 2017-10-12 1.16
2 2017-10-13 1.16
3 2017-10-14 1.16
```

After fetching the data, we construct an estimated risk-free index RFI by making an imaginary deposit of \$1,000.00 and investing it each banking day at the reported Daily Federal Funds rate actual/365 simple interest <sup>1</sup> until the next banking day.

```
> ffr$RFI[1] <- 1000.00
> for(i in 2:nrow(ffr))
+ {
+   ffr$RFI[i] <- lag(ffr$RFI)[i] *
+     (1 + lag(ffr$DFF)[i] / 100 *
+       (ffr$date[i]-lag(ffr$date)[i]) / 365)
+ }
> ffr[1:3,]

      date  DFF    RFI
1 2017-10-12 1.16 1000.000
2 2017-10-13 1.16 1000.032
3 2017-10-14 1.16 1000.064
```

---

<sup>1</sup>Actual/365 simple interest is only a guess: as we all know, the Fed is highly secretive about its operations, and we are only *estimating* for this purpose.

### 3.4 Stock ticker data

```
> fetch_quandl_p      <- Quandl.datatable('WIKI/PRICES',
+   ticker='CDE', date.gte=start_date, date.lte=end_date,
+   qopts.columns=c("date", "adj_close"))
> fetch_quandl_p[1:3,]
```

```
      date adj_close
1 2017-10-12      9.08
2 2017-10-13      8.94
3 2017-10-16      8.85
```

We use an “adjusted close” price of the stock we are analyzing; namely a daily price ticker series based on actual closing price, but conveniently “adjusted” for dividends and share splits.

## 4 Analyzing the data

The first step in analyzing the data is to “splice” it together — we have price level series for a major market index, a risk-free index, and the stock of our interest. We will use an *inner\_join*, because there are possibly some banking days on which the stock market is closed or vice versa.

```
> j <- inner_join(
+   inner_join(ffr, sp500, by="date"),
+   fetch_quandl_p, by="date")
> j$r0 <- log(j$RFI) - log(lag(j$RFI))
> j$r1 <- log(j$SP500) - log(lag(j$SP500))
> j$r2 <- log(j$adj_close) - log(lag(j$adj_close))
> j$x <- j$r1 - j$r0
> j$y <- j$r2 - j$r0
```

### 4.1 Fitting a linear model

The next step is to fit a linear model to the data,

$$y_i = \alpha + \beta x_i + \epsilon_i, \quad 1 \leq i \leq n, \quad (1)$$

where the “Sum of Squared Error”

$$SSE = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y_i - (\alpha + \beta x_i))^2 \quad (2)$$

is *minimized*.

This process, called **linear regression**, is very easy to do in R.

### price level performance

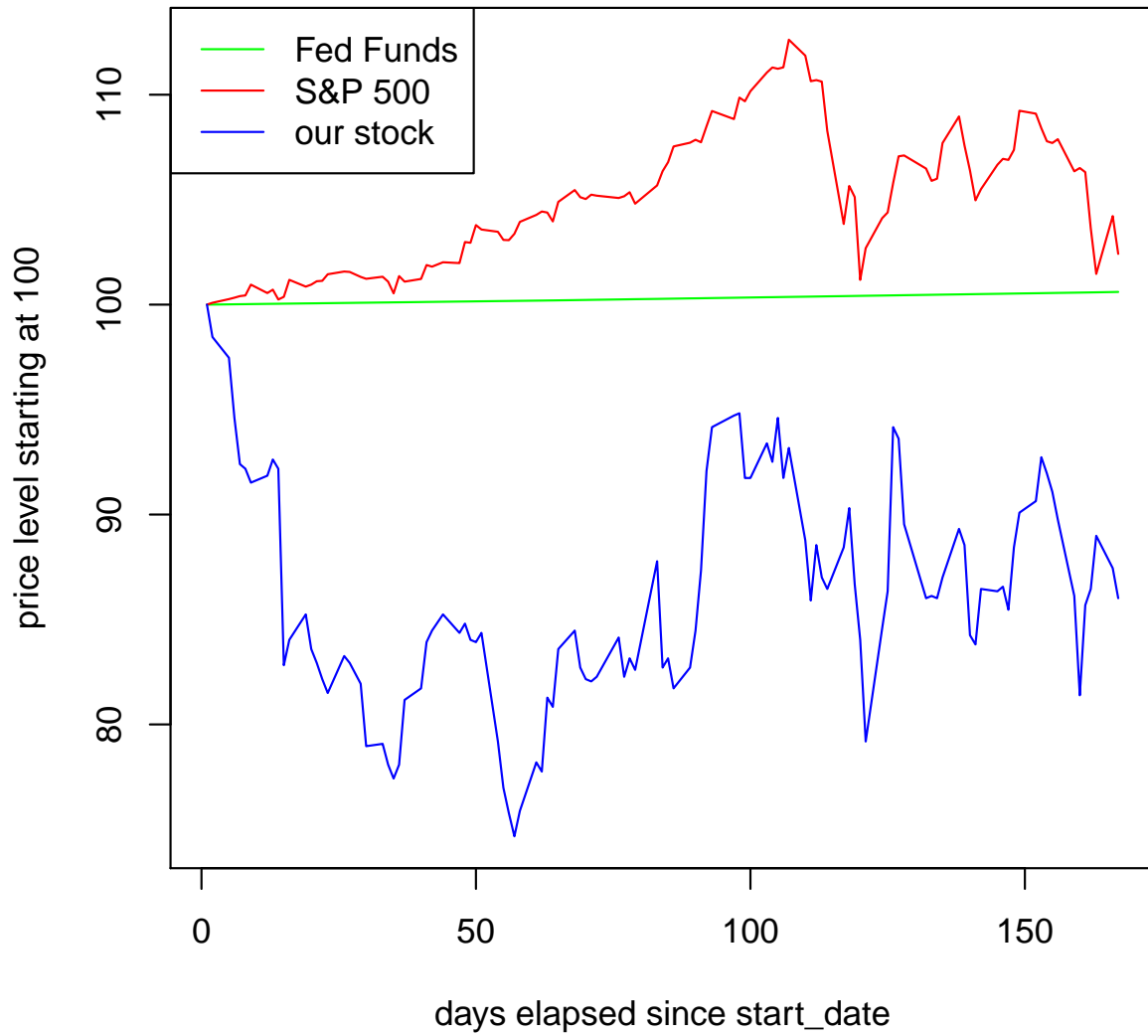


Figure 1: Plot of price level data

## 4.2 Free $\alpha$ model

This is a basic linear regression of daily logarithmic returns of the particular stock or portfolio that we are analyzing against returns of the general market.

```
> model_1 <- lm(formula = y ~ x, data = j[2:nrow(j),])
> summary(model_1)
```

Call:

```
lm(formula = y ~ x, data = j[2:nrow(j), ])
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.106360	-0.012160	-0.000778	0.012969	0.080280

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-0.001494	0.002650	-0.564	0.5741
x	0.592660	0.279848	2.118	0.0364 *

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02804 on 110 degrees of freedom

Multiple R-squared: 0.03918, Adjusted R-squared: 0.03044

F-statistic: 4.485 on 1 and 110 DF, p-value: 0.03645

The coefficient labeled “(Intercept)” is known as  $\alpha$  (“alpha”) — in this case a *daily*  $\alpha$  which would have to be multiplied by 252<sup>2</sup> to yield an annualized “stock market  $\alpha$ ” on a logarithmic basis.

The coefficient labeled “x” is the estimated  $\beta$  (“beta”) of the stock whose returns we are modeling against those of the market in general. A red line is plotted in Figure 2 with the slope and intercept of this model

## 4.3 Forced $\alpha = 0$ model

Now it would be rare, exceptional, and not particularly meaningful if the reported coefficient for the (Intercept) of model\_1 were statistically different from zero. The profound lack of statistical significance of estimates of  $\alpha$  is strong evidence for the common saying or disclaimer: “Past performance is no guarantee of future returns.”

Some readers may be of the philosophy that “There Ain’t No Such Thing As A Free Lunch” [7, 6],  $\alpha = 0$ , and therefore one free parameter may be eliminated from the model without affecting its validity.

```
> model_0 <- lm(formula = y ~ x - 1, data = j[2:nrow(j),])
> summary(model_0)
```

---

<sup>2</sup>Or, the number of trading days in a year.

```

Call:
lm(formula = y ~ x - 1, data = j[2:nrow(j), ])

Residuals:
    Min       1Q   Median       3Q      Max
-0.107850 -0.013629 -0.002275  0.011482  0.078822

Coefficients:
    Estimate Std. Error t value Pr(>|t|)
x    0.5900     0.2789   2.115  0.0367 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02795 on 111 degrees of freedom
Multiple R-squared:  0.03874,    Adjusted R-squared:  0.03008
F-statistic: 4.474 on 1 and 111 DF,  p-value: 0.03666

```

## 5 Interpreting the results

The summary reports generated by R yield a lot of useful information on the statistical significance of the estimated coefficients. The estimates of  $R^2$  are particularly significant —  $R^2$  is the fraction of the variation of the price level of a particular stock or portfolio which can be explained by the performance of the market as a whole.

For many stocks, an excessive  $\beta$  together with an excessive correlation with the market in general,  $R^2$ , is a serious problem, and one that discourages us from investing especially in large-cap stocks.

This is not at all the problem we face with the stock we are now analyzing, (i.e., shortly after the end of the first quarter 2018.) The stock has experienced a significant drop and high general volatility in price, which are very poorly explained by our model of regression against the S&P 500.

Therefore we need to look for other factors which may better explain day-to-day movements in the price of the stock which we are analyzing. A glance at the front page of the website [2] of the company whose stock we are analyzing suggests that a possible next step for further research would be to fetch some commodity price data for gold and silver, and do a multiple regression.

```

> m <- inner_join(data.frame(date=j$date, price=j$adj_close,
+   rfi=j$RFI, sp500=j$SP500),
+   inner_join(Quandl('COM/AG_USD', date.gte=start_date, date.lte=end_date),
+     Quandl('COM/AU_LPM', date.gte=start_date, date.lte=end_date),
+     by="date", suffix=c("_Ag", "_Au")), by="date")
> # natural logarithms of relative price levels
> m$l_price <- log(m$price/m$rfi)
> m$l_sp500 <- log(m$sp500/m$rfi)

```

### daily log returns

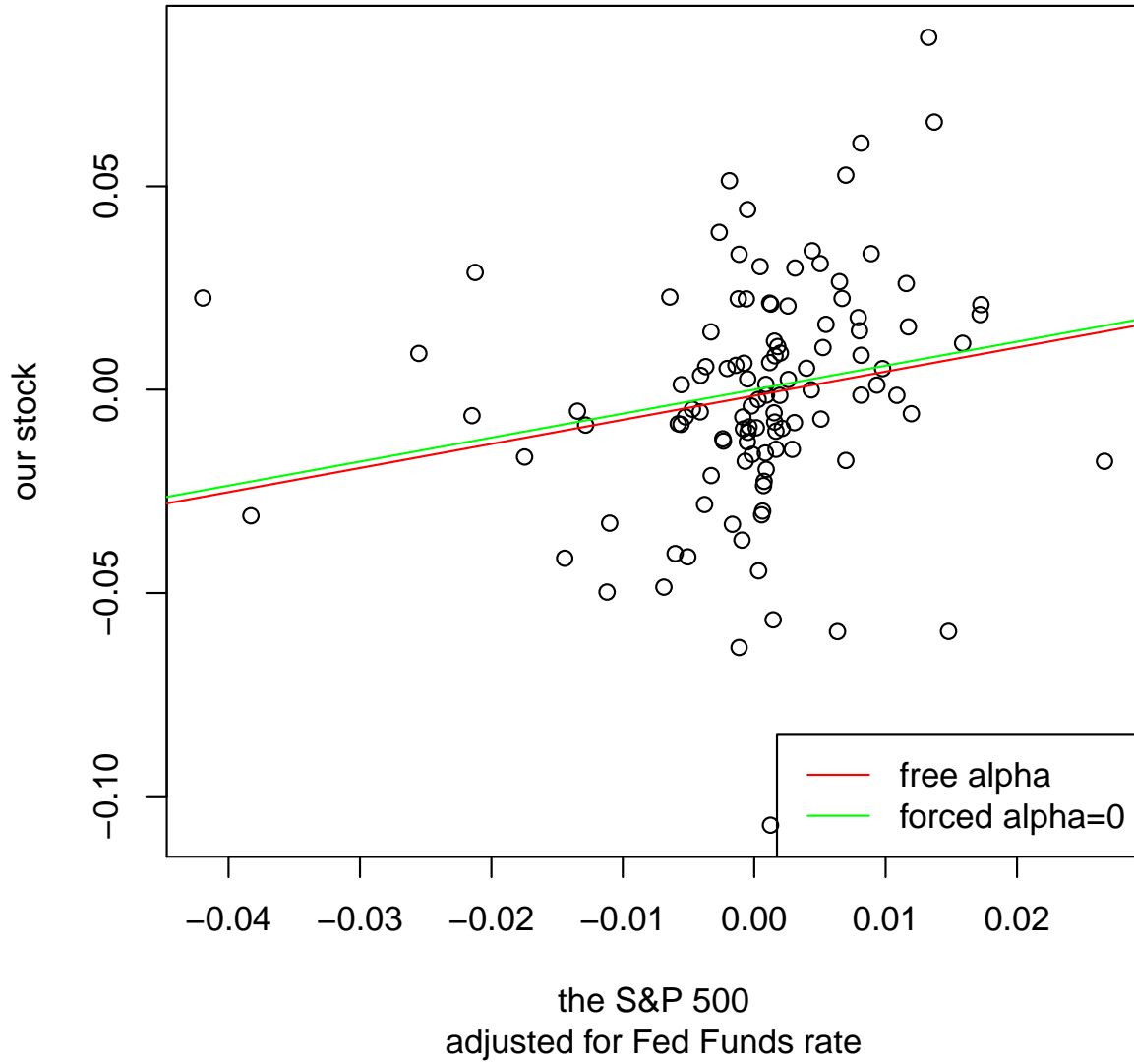


Figure 2: Scatter plot with lines of best fit from linear model



```

> m$l_ag <- log(m$value_Ag/m$rfi)
> m$l_au <- log(m$value_Au/m$rfi)
> # diff the logarithms from the previous day
> m$r_price <- m$l_price - lag(m$l_price)
> m$r_sp500 <- m$l_sp500 - lag(m$l_sp500)
> m$r_ag <- m$l_ag - lag(m$l_ag)
> m$r_au <- m$l_au - lag(m$l_au)

```

The data sets COM/AG\_USD and COM/AU\_LPM from [12] are so-called “London fix” prices for silver and gold, respectively, in U.S. dollars. The term “fix” suggests that the prices of these commodities are excessively regulated by — to be frank — a criminal **cartel**, and that in reality no true free market exists for such precious metals. The fact that these “fixes” are reported in the tables as “values” rather than “prices” should set off further alarm bells. We have heard the “value-vs.-price” story from brokers<sup>3</sup> before, and we do not like what we hear.

```

> model_M <- lm(data = m[2:nrow(m),], formula = r_price ~ r_sp500 + r_ag + r_au)
> summary(model_M)

```

Call:

```

lm(formula = r_price ~ r_sp500 + r_ag + r_au, data = m[2:nrow(m),
  ])

```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.103810	-0.012133	0.000738	0.015666	0.057138

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-0.003243	0.002676	-1.212	0.2286
r_sp500	0.631960	0.271118	2.331	0.0219 *
r_ag	-0.894446	0.440660	-2.030	0.0452 *
r_au	2.819952	0.625746	4.507	1.91e-05 ***

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.026 on 93 degrees of freedom

Multiple R-squared: 0.2513, Adjusted R-squared: 0.2272

F-statistic: 10.41 on 3 and 93 DF, p-value: 5.728e-06

We do not like the results of this multiple regression. As we do this analysis shortly after the end of the first quarter 2018, the company’s stock price shows an extremely high positive correlation with the price of gold, and a strong negative correlation with the price

---

<sup>3</sup>Technically, they are dealers rather than brokers: yes, you got that right, *dealers* selling a daily or twice-a-day “fix” like they run some sort of pawn shop on the side in a really bad part of London where they buy jewelry, melt it down, and sell coins and such.

of silver. Such correlations<sup>4</sup> sometimes expose a publicly traded mining company's hedge position in futures markets. This company may have hedged against more silver than the market believes it has in reserve, and on the other hand be inadequately hedged against a fall in the price of gold. Further analysis is warranted.

This company's stock seems to be a strong bet for gold against silver. There is not really a satisfactory explanation for this.

Nor is there an explanation for the lack of accounting for platinum, palladium, and other precious metals that tend to be extracted along with gold and silver. Nor can we explain the contemptuous zeros in the columns for "lead" and "zinc" in company's reports of "proven" and "probable" reserves or the total lack of mention of copper.

The ore may be not be of a grade that it is profitable to mine it solely for the sake of other metals than gold and silver, but at the same time, when the company has gone to the expense of digging the ore, and extracting the metals, we do not believe that other precious and valuable metals, invariably also extracted as a by-product of mining silver and gold, should simply be thrown away without accounting.

## References

- [1] Sam Boysel. `fredr` : Interact with the FRED API in the R workspace, 2018. URL <https://github.com/sboysel/fredr>.
- [2] CDE. Coeur Mining, Inc., 2018. URL <http://www.coeur.com/>.
- [3] CRAN. Package Quandl, 2016. URL <https://cran.r-project.org/web/packages/Quandl/>.
- [4] Federal Reserve Bank of St. Louis. Federal Reserve Economic Data, 2018. URL <https://fred.stlouisfed.org/>.
- [5] Charlie Geyer. An Sweave demo: Literate programming in R: Reproducible research, 2018. URL <http://users.stat.umn.edu/~geyer/Sweave/>.
- [6] Robert Heinlein. *The Moon is a Harsh Mistress*. Tom Doherty Associates, Inc., 175 Fifth Ave, New York, NY 10010, 1966. ISBN 0-312-86355-1.
- [7] David R. Henderson. TANSTAAFL, there ain't no such thing as a free lunch. *Library of Economics and Liberty*, March 3, 2014. URL <http://www.econlib.org/library/Columns/y2014/Hendersontanstaaf1.html>.
- [8] G. Jay Kerns. Easiest Sweave Template Ever, 2011. URL [http://gkerns.people.ysu.edu.user/tutorial/useR-B/document/EasySweaveTemplate.Rnw](http://gkerns.people.ysu.edu/user/tutorial/useR-B/document/EasySweaveTemplate.Rnw). Author's home page URL <http://gkerns.people.ysu.edu/>.

---

<sup>4</sup>O.K., we are not using the term "correlation" with its precise statistical meaning  $\text{Corr}(x, y) = \text{Covar}(x, y) / (\sigma_x \sigma_y)$ . We are talking about the coefficients of the regression, i.e., the linear *dependency* of the day-to-day change in price of the stock on that of the S&P 500, silver, or gold, respectively.

- [9] Donald E. Knuth. *Literate programming*. Cambridge University Press, 1992. URL <http://press.uchicago.edu/ucp/books/book/distributed/L/bo3629677.html>.
- [10] LaTeX. A document preparation system, 2018. URL <https://www.latex-project.org/>.
- [11] Friedrich Leisch. Not found, 2018. URL <http://www.ci.tuwien.ac.at/~leisch/>, URL <http://www.stat.uni-muenchen.de/~leisch/>.
- [12] Quandl. Financial, Economic and Alternative Data, 2018. URL <https://www.quandl.com/>.
- [13] R. The R project for statistical computing, 2018. URL <https://www.r-project.org/>.