Lab 2 Report

Alvin Ng, Matthew Rubino, and Kamaljeet Ghotra

3/27/2022

#Introduction Investment gurus and stock market junkies have long-attempted to predict underlying equity returns of constituent stocks with varying levels of success. With the mixed success of predictions, investors turned to employing strategies and philosophies to guide their investment decisions as well as keep them disciplined in the long run. Warren Buffet, the Oracle of Omaha, is arguably one of the best investors in generations and has been transparent about his investment strategy and philosophy with the investment community.

Warren Buffett has shared his investment wisdom with his shareholders over the years through his shareholder letters. For example, in Berkshire Hathaway's 1981 shareholder letter, Buffet notes that the company's "acquisition preferences run toward businesses that generate cash, not those that consume it" and that what Berkshire Hathaway "really want(s) do is buy a business that's a great business, which means that business is going to earn a high return on capital employed for a very long period of time, and where (they) think the management will treat (them) right."

Over the years Buffett has suggested that inflation is a "particularly ironic punishment for bad businesses" and that those businesses that thrive in such environments typically have the capacity to "increase prices rather easily, even when product demand is flat and capacity is not full utilized" and "accommodate large dollar volume increases in business with only minor additional investment of capital." In his 1979 shareholder letter, Buffet notes that the team has achieved its phenomenal performance during the Great Inflation period while also using a low amount of leverage.

In this study, we aim to identify the characteristics of Warren Buffet's approach that served him well throughout his investment career while seeking to explore these characteristics in environment's most similar to today's. To this end, we highlight our research question below.

##Research Question How do fundamental metrics identified by Warren Buffet such as Net Debt to EBITDA, Free Cash Flow Yield, and Return on Invested Capital relate to the total return of underlying S&P 500 constituents in inflationary environments.

###(i) Conceptualization of an Inflationary Environment: We define an inflationary environment as one whereby the Personal Consumption Expenditures Index percentage change is larger than 2% on an annual basis or roughly 0.49% on a quarterly basis (((1+0.02)^(1/4))-1). We selected such a threshold as 2% remains the Federal Reserve's long run stated goal for annual price increases; this goal has been set in accordance with one leg of their dual mandate - (i) to minimize unemployment and (ii) ensure stable prices.

```r
# install.packages("readxl")
install.packages("formatR")
```

```
## Installing package into '/opt/r'
## (as 'lib' is unspecified)
```

```r
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.8
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()    masks stats::lag()
```

```
library(lmtest)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
library(sandwich)
library(readxl)
library(dplyr)
library(readr)
library(ggplot2)
library(patchwork)
library(stargazer)
```

```
##
## Please cite as:
```

```
##  Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
##  R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

```
multiplesheets <- function(fname) {

  # getting info about all excel sheets
  sheets <- readxl::excel_sheets(fname)
  tibble <- lapply(sheets, function(x) readxl::read_excel(fname, sheet = x))
  data_frame <- lapply(tibble, as.data.frame)

  # assigning names to data frames
  names(data_frame) <- sheets
  return(data_frame)
}
```

#Data and Methodology ##(i) Data Our team gathered the following four categories of time series data by leveraging Bloomberg's Application Program Interface (API) functionality in Microsoft Excel:

###(1) Company Descriptive Data: Descriptive S&P 500 Index constituent data including company name, date or year of incorporation, number of employees, GICS sector names, GICS industry name, country, current market capitalization, MSCI value inclusion factor, MSCI growth inclusion factor, is the CEO a founder, % of shares held outstanding by CEO, CEO's age, total salaries and bonuses paid to CEO and equivalent, the percentage of independent directors, average Board of Directors tenure, average Board of Directors Age, Board of Directors size, total compensation of Board of Directors awarded, and registered state location. This represents a snapshot of the descriptive data for each of the companies as of December 2021.

###(2) Fundamental Data: Fundamental S&P 500 Index Constituent data on a quarterly basis from December 1999 to December 2022 including:

(i) Free cash flow yield
(ii) Leverage (Net Debt to EBITDA)
(iii) Return on Invested Capital (ROIC)

###(3) Total Return Data: (i) The total quarterly return of every S&P 500 Index constituent and the index from January 2000 to February 2022. The quarterly total return data was also converted into quarterly total returns by geometrically chain-linking the quarterly returns.

###(4) Inflation Data: Personal consumption expenditures month-over-month percentage change, quarterly percentage change, and rolling twelve-month percentage change from January 1999 to January 2022 from the Bureau of Economic Analysis within the U.S. Department of Commerce.

The Personal Consumption Expenditures Index (PCE) was selected as this Index is the Federal Reserve's preferred measure of inflation. The PCE includes how much is spent on durable and non-durable goods, as well as services; it uses prices from all households, corporations, and governments into account, along with GDP.

##(ii) Methodology The purpose of the analysis is to examine the effect S&P 500 companies' fundamental metrics on respective companies' equity total returns. Accordingly we have designed a multiple regression model below as a starting point to capture the implication of the trends in quarterly company fundamentals on corresponding quarterly total returns. The outcome variable of interest, Y, is represented by each company's monthly total return. Note, that the monthly total return was converted to a quarterly total return by geometrically chain-linking the underlying monthly returns to match the time window for the underlying fundamental data.

##Model A $Y = \beta 0 + \beta 1 LEV + \beta 2 FCFY + \beta 3 ROIC + \epsilon$

We show each of the variables and their respective representations in the dataframe. The underlying data spans from January 2000 to December 2021 on a quarterly basis.

```
variable_names <- data.frame(Variable=c('Y', 'LEV', 'FCFY', 'ROIC'),
                             Representation=c('Respective Quarterly Total Return of Constituent Stock', 'N
variable_names
```

```
##   Variable                                         Representation
## 1        Y Respective Quarterly Total Return of Constituent Stock
## 2      LEV                                 Net Debt to EBITDA Ratio
## 3     FCFY                                     Free Cash Flow Yield
## 4     ROIC                                Return on Invested Capital
```

Reading in data:

```
SP500 <- multiplesheets("Data v3.xlsx")
PCE <- read.csv("PCE v3.csv")
```

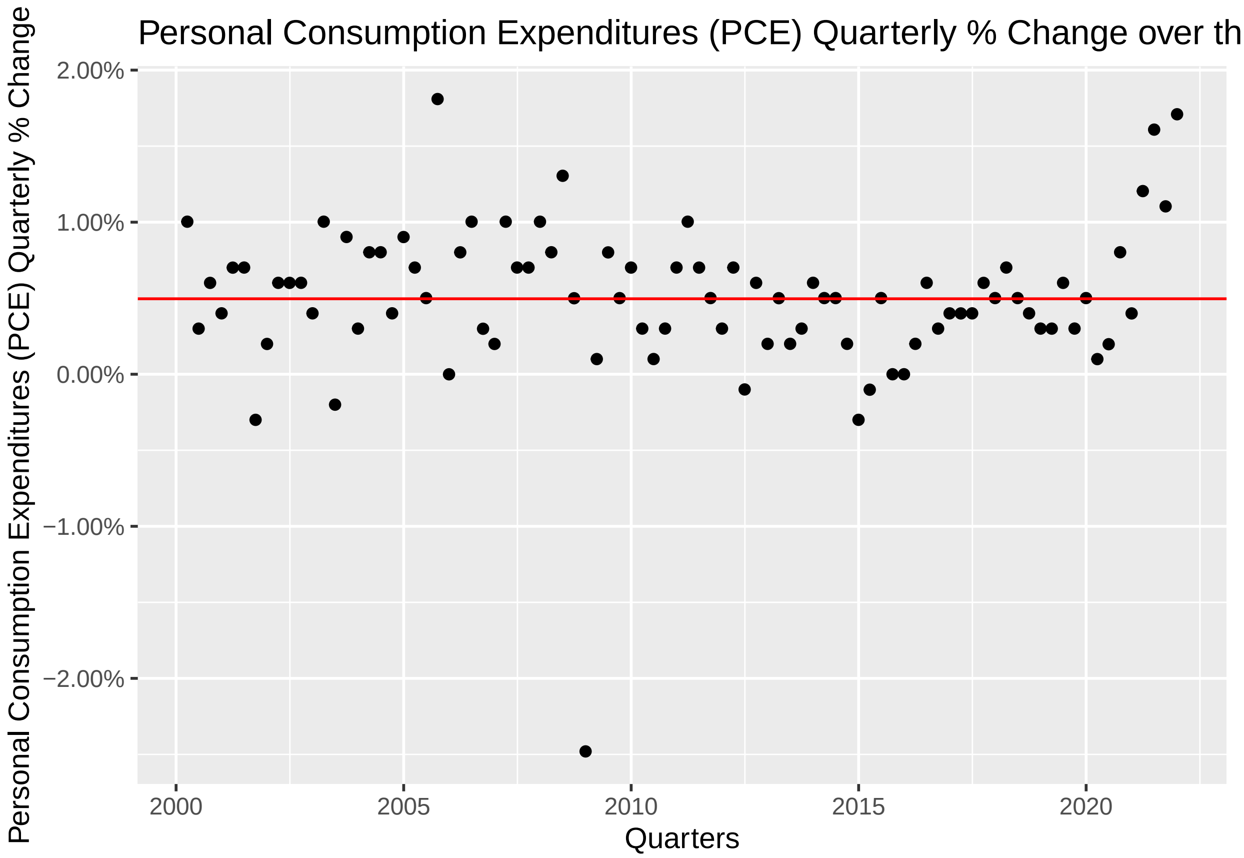
Cleaning the data column name:

```
PCE <- PCE %>%
  rename(
    Personal_Consumption_Exp_PCE_in_perc = Personal.consumption.expenditures..PCE...in...,
    Rolling_12_Month_PCE_in_perc = Rolling.12.Month.PCE..in...,
    PCE_Quarterly_Perc_Change = PCE.Quarterly...Change
  ) %>%
  mutate(
    Date = as.Date(parse_datetime(Quarter, "%m/%d/%Y"))
  )
```

#(iia) A Model Building Process We began by cleaning and exploring the underlying data. We first removed all null values from the dataset, which removed 10,000 observations from our initial 40,000 observations. Next we began performing exploratory data analysis by creating a scatter plot to compare quarterly fundamentals against the respective quarterly total returns.

```
PCE_line <- PCE %>%
  ggplot +
  aes(x = Date, y = PCE_Quarterly_Perc_Change) +
  geom_point() +
  labs(
    title = "Personal Consumption Expenditures (PCE) Quarterly % Change over the Quarters of 1999 - 2022",
    x = "Quarters",
    y = "Personal Consumption Expenditures (PCE) Quarterly % Change"
  ) +
  scale_y_continuous(labels = scales::percent) +
  geom_hline(yintercept=((1+0.02)^(1/4))-1, color = "red")
```

```
PCE_line
```



De-annualizing the rate of inflation to a quarterly basis..See Research Question section (i) above for the description of how we define inflationary environments.

```
inflation_Quarter <- PCE %>%
  filter(PCE_Quarterly_Perc_Change >= (((1+0.02)^(1/4))-1)) %>%
  select(Date)
```

```
unique_sectors <- SP500$Descriptive %>%
  select(`Sector Name`) %>%
  distinct(`Sector Name`)

max_rows <- 0
for (name in unique_sectors$`Sector Name`) {
  filtered <- SP500$Descriptive %>% filter(`Sector Name` == name) %>% select(`Company Name`)
  if (max_rows < length(filtered$`Company Name`)) {
    max_rows <- length(filtered$`Company Name`)
  }
}
```

```
companies_under_sector <- data.frame(matrix(ncol = length(unique_sectors$`Sector Name`), nrow = max_rows))
colnames(companies_under_sector) <- unique_sectors$`Sector Name`
for (name in names(companies_under_sector)) {
  filtered <- SP500$Descriptive %>% filter(`Sector Name` == name) %>% select(`Company Name`)
  companies_under_sector <- companies_under_sector %>%
    mutate(
      "{name}" := c(filtered$`Company Name`, rep(NA, max_rows - length(filtered$`Company Name`)))
    )
```

```r
}

companies_and_sector <- SP500$Descriptive %>% select(c("Company Name", "Sector Name"))
rm(max_rows)
rm(filtered)
```

```r
print(unique_sectors)
```

```
##               Sector Name
## 1               Materials
## 2              Financials
## 3  Communication Services
## 4  Information Technology
## 5             Industrials
## 6                  Energy
## 7        Consumer Staples
## 8             Health Care
## 9             Real Estate
## 10 Consumer Discretionary
## 11              Utilities
```

Building the dataset:

```r
columns <- c("Date","Total Return (%)","Lagged Total Return (%)","Net Debt to EBITDA","Free Cash Flow Yiel
df <- data.frame(matrix(nrow = 0, ncol = length(columns)))
colnames(df) <- columns

for (name in names(SP500$`Net Debt to EBITDA`)) {
  if (name != 'Date') {

    total_return_df <- SP500[["Quarterly Total Returns (%)"]] %>%
      select(c("Date", name)) %>%
      rename(
        "Total Return (%)" := {{name}}
      ) %>%
      filter(!is.na(`Total Return (%)`))

    lagged_total_return_df <-
      SP500[["Quarterly Total Returns (%)"]] %>%
      select(c("Date", name)) %>%
      rename(
        "Lagged Total Return (%)" := {{name}}
      ) %>%
      mutate(
        "Lagged Total Return (%)" = lead(`Lagged Total Return (%)`)
      ) %>%
      filter(!is.na(`Lagged Total Return (%)`))

    net_debt_df <- SP500[["Net Debt to EBITDA"]] %>%
      select(c("Date", name)) %>%
      rename(
        "Net Debt to EBITDA" := {{name}}
      ) %>%
      filter(!is.na(`Net Debt to EBITDA`))

    cash_flow_df <- SP500[["Free Cash Flow Yield (%)"]] %>%
      select(c("Date", name)) %>%
```

```
    rename(
      "Free Cash Flow Yield" := {{name}}
    ) %>%
    filter(!is.na(`Free Cash Flow Yield`))

  ROIC_df <- SP500[["ROIC (%)"]] %>%
    select(c("Date", name)) %>%
    rename(
      "ROIC" := {{name}}
    ) %>%
    filter(!is.na(`ROIC`))

  temp_df <- total_return_df %>%
    full_join(lagged_total_return_df, by = "Date") %>%
    full_join(net_debt_df, by = "Date") %>%
    full_join(cash_flow_df, by = "Date") %>%
    full_join(ROIC_df, by = "Date") %>%
    mutate(
      "Company Name" := {{name}}
    )

  df <- rbind(df, temp_df)
 }
}
```

```
## Note: Using an external vector in selections is ambiguous.
## i Use `all_of(name)` instead of `name` to silence this message.
## i See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.
## This message is displayed once per session.
```

```
rm(total_return_df)
rm(lagged_total_return_df)
rm(net_debt_df)
rm(cash_flow_df)
rm(ROIC_df)
rm(temp_df)
rm(name)
rm(PCE)
```

Adding the dummy variables:

```
df <- df %>%
  mutate(
    "Inflationary Environment" = ifelse(as.Date(Date) %in% inflation_Quarter$Date, 1, 0),
    "Materials" = ifelse(`Company Name` %in% companies_under_sector$Materials, 1, 0),
    "Financials" = ifelse(`Company Name` %in% companies_under_sector$Financials, 1, 0),
    "Communication Services" = ifelse(`Company Name` %in% companies_under_sector$`Communication Services`,
    "Information Technology" = ifelse(`Company Name` %in% companies_under_sector$`Information Technology`,
    "Industrials" = ifelse(`Company Name` %in% companies_under_sector$Industrials, 1, 0),
    "Energy" = ifelse(`Company Name` %in% companies_under_sector$Energy, 1, 0),
    "Consumer Staples" = ifelse(`Company Name` %in% companies_under_sector$`Consumer Staples`, 1, 0),
    "Health Care" = ifelse(`Company Name` %in% companies_under_sector$`Health Care`, 1, 0),
    "Real Estate" = ifelse(`Company Name` %in% companies_under_sector$`Real Estate`, 1, 0),
    "Consumer Discretionary" = ifelse(`Company Name` %in% companies_under_sector$`Consumer Discretionary`,
    "Utilities" = ifelse(`Company Name` %in% companies_under_sector$Utilities, 1, 0)
  ) %>%
  left_join(companies_and_sector, by = "Company Name")
```

```r
rm(companies_and_sector)
rm(companies_under_sector)
rm(inflation_Quarter)
rm(columns)
```

Removing NAs

```r
length_of_df <- df %>%
  filter(is.na(`Total Return (%)`))

print(dim(length_of_df))
```

```
## [1] 2576   20
```

```r
# we have 2,501 sample observations that has NA total returns
```

```r
# If we drop all NAs

drop_all_na_df <- df %>% drop_na()
print(dim(drop_all_na_df))
```

```
## [1] 29726    20
```

```r
# Then we will have 29881 sample observations, down from 39676
```

We begin by performing EDA on each of the underlying input variables (LEV, FCFY, ROIC). We plotted graphs to see if any of the variables need transformation.

0.0.1 Net Debt to EBITDA Scatter Plot

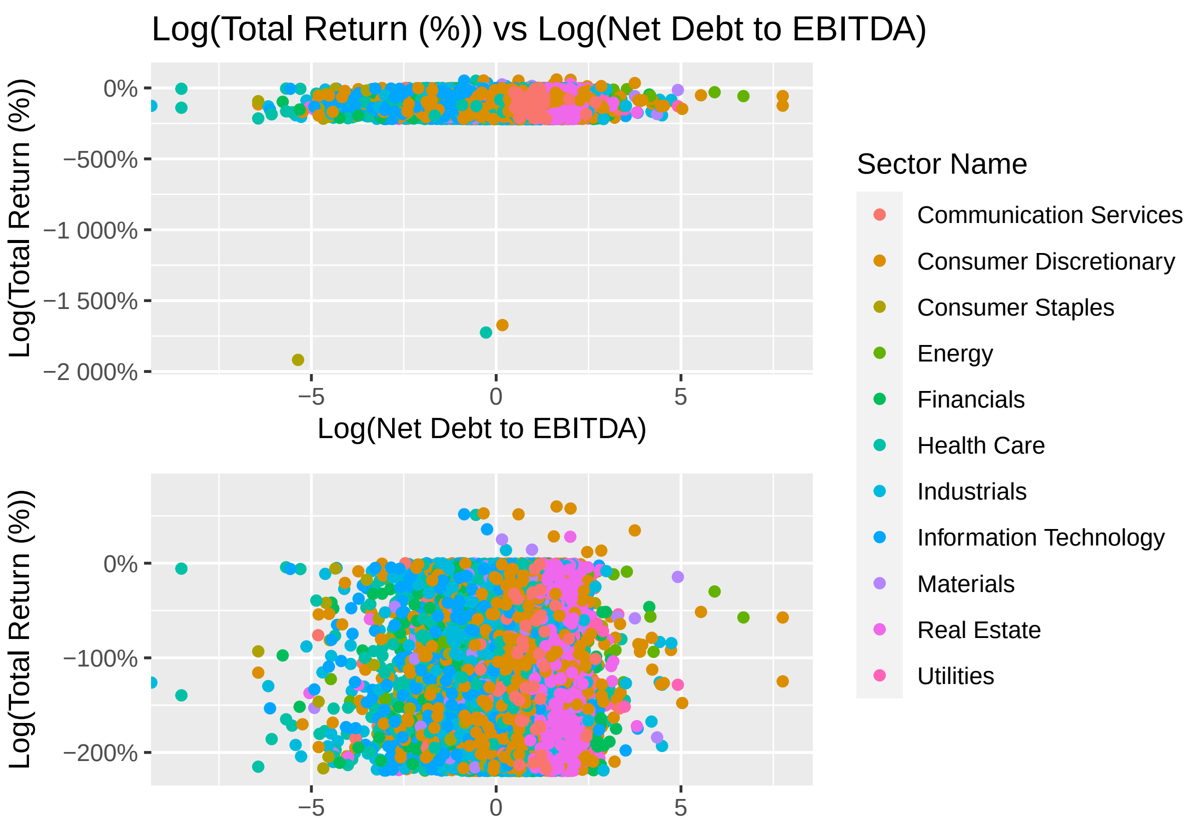
Even though the range of quarterly total returns and net debt to ebitda are large and typically, we would apply a log transformation, we decided to not apply a log transformation because there are a lot of negative numbers-to which we cannot apply a log. Because we are not able to apply a log to negative numbers, we decided not to pursue a log transformation.

```r
return_net_debt_scatter_w_outliers <- drop_all_na_df %>%
  ggplot(aes(x = log(`Net Debt to EBITDA`), y = log(`Total Return (%)`), color = `Sector Name`)) +
  geom_point() +
  labs(
    title = "Log(Total Return (%)) vs Log(Net Debt to EBITDA)",
    x = "Log(Net Debt to EBITDA)",
    y = "Log(Total Return (%))"
  ) +
  scale_y_continuous(labels = scales::percent)

return_net_debt_scatter_wo_outliers <- drop_all_na_df %>%
  filter(log(`Total Return (%)`) > -5) %>%
  ggplot(aes(x = log(`Net Debt to EBITDA`), y = log(`Total Return (%)`), color = `Sector Name`)) +
  geom_point() +
  labs(
    y = "Log(Total Return (%))"
  ) +
  theme(
    axis.title.x = element_blank()
  ) +
  scale_y_continuous(labels = scales::percent)
```

```
## Warning in log(`Total Return (%)`): NaNs produced
```

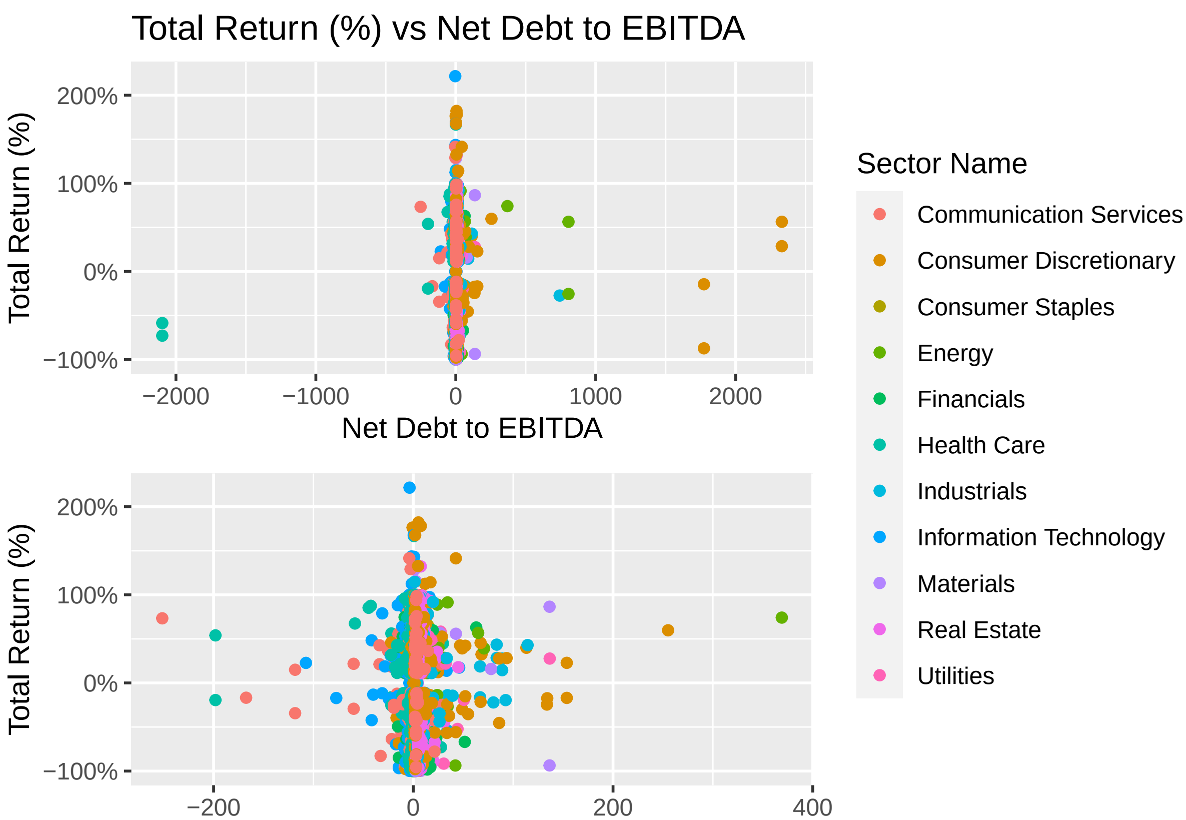
```
(return_net_debt_scatter_w_outliers + return_net_debt_scatter_wo_outliers +
  plot_layout(ncol = 1)) + plot_layout(guides = 'collect')
```

```
## Warning in log(`Net Debt to EBITDA`): NaNs produced

## Warning in log(`Net Debt to EBITDA`): NaNs produced

## Warning in log(`Total Return (%)`): NaNs produced

## Warning: Removed 15705 rows containing missing values (geom_point).

## Warning in log(`Net Debt to EBITDA`): NaNs produced

## Warning in log(`Net Debt to EBITDA`): NaNs produced

## Warning: Removed 4866 rows containing missing values (geom_point).
```




```
return_net_debt_scatter <- drop_all_na_df %>%
  ggplot(aes(x = `Net Debt to EBITDA`, y = `Total Return (%)`, color = `Sector Name`)) +
  geom_point() +
  labs(
    title = "Total Return (%) vs Net Debt to EBITDA",
    x = "Net Debt to EBITDA",
    y = "Total Return (%)"
  ) +
  scale_y_continuous(labels = scales::percent)

return_net_debt_scatter_w_filter <- drop_all_na_df %>%
  filter(`Net Debt to EBITDA` <= 500, `Net Debt to EBITDA` >= -500) %>%
  ggplot(aes(x = `Net Debt to EBITDA`, y = `Total Return (%)`, color = `Sector Name`)) +
  geom_point() +
```

```
  labs(
    y = "Total Return (%)"
  ) +
  scale_y_continuous(labels = scales::percent) +
  theme(
    axis.title.x = element_blank()
  )

(return_net_debt_scatter + return_net_debt_scatter_w_filter +
  plot_layout(ncol = 1)) + plot_layout(guides = 'collect')
```



Looking at the plot, Net Debt to EBITDA appear to correlate with Quarterly Total Return (%).

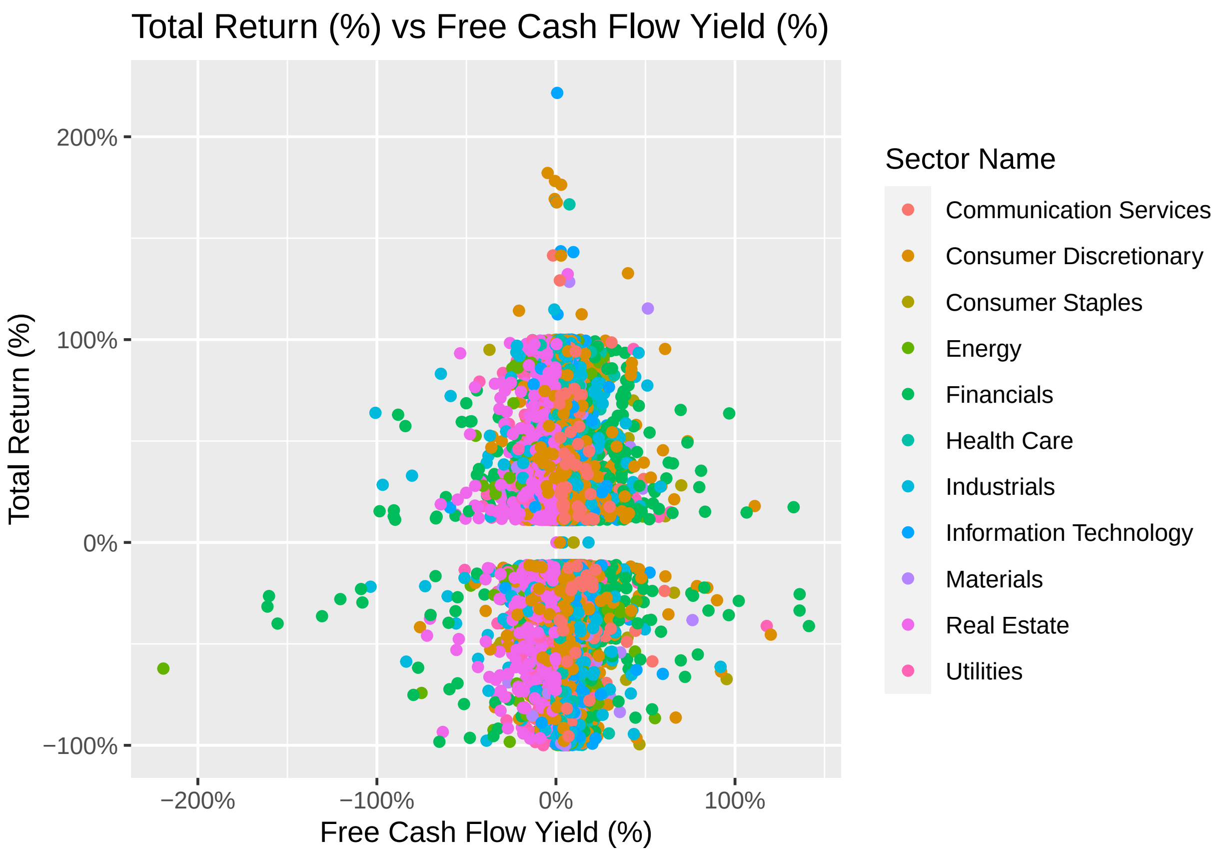
As seen above, the outliars have considerable weighing factor relative to the data points clustered together when it comes to explaining the variation in quarterly total return; hence we decided to remove them from the dataset since they only represent a small proportion of the dataset and are unlikely to drive meaningfully different results.

```
drop_all_na_df <- drop_all_na_df %>%
  filter(`Net Debt to EBITDA` >= -300, `Net Debt to EBITDA` <= 300) %>%
  filter(`ROIC` >= -5) %>%
  filter(`Free Cash Flow Yield` >= -2.5, `Free Cash Flow Yield` <= 2.5)
```

0.1 Free Cash Flow Yield Scatter Plot (%)

```
return_cash_flow_scatter <- drop_all_na_df %>%
  ggplot(aes(x = `Free Cash Flow Yield`, y = `Total Return (%)`, color = `Sector Name`)) +
  geom_point() +
  labs(
    title = "Total Return (%) vs Free Cash Flow Yield (%)",
```

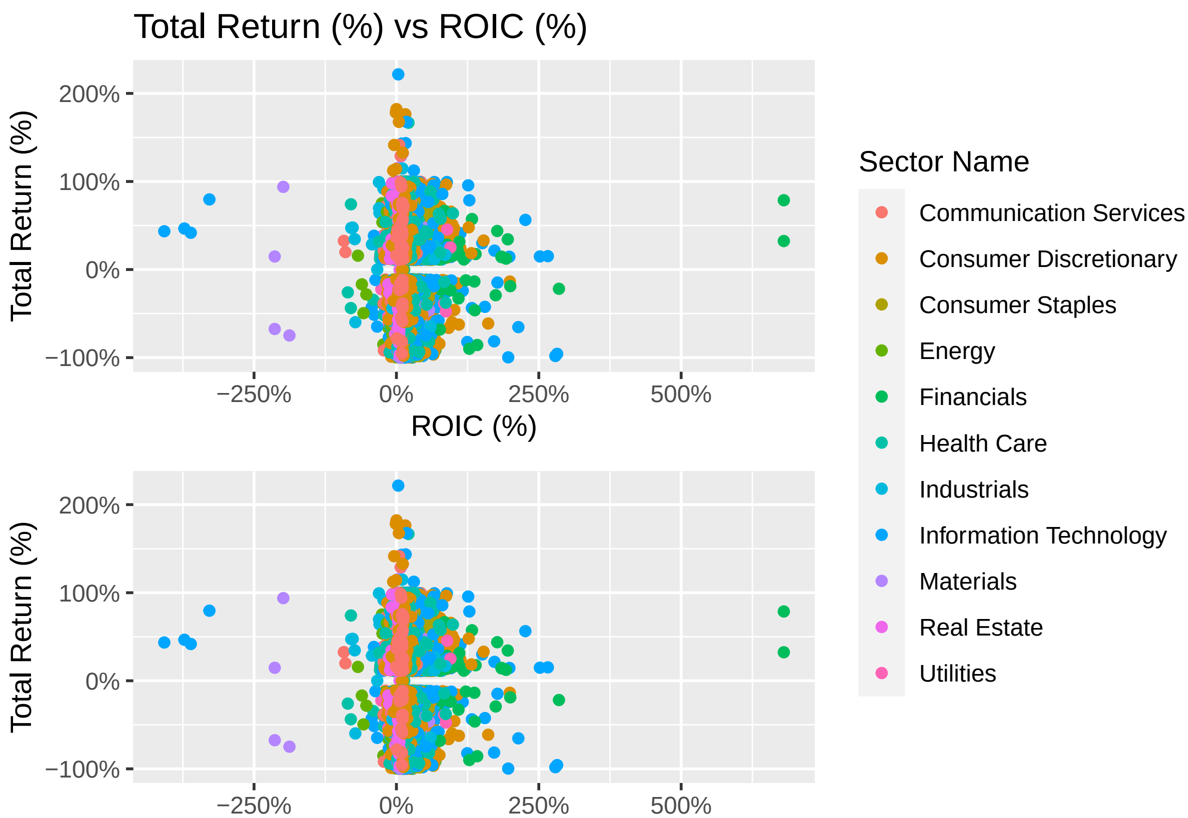
```
    x = "Free Cash Flow Yield (%)",
    y = "Total Return (%)"
  ) +
  scale_y_continuous(labels = scales::percent) +
  scale_x_continuous(labels = scales::percent)

return_cash_flow_scatter
```



0.2 ROIC Scatter Plot (%)

```
return_ROIC_scatter <- drop_all_na_df %>%
  ggplot(aes(x = `ROIC`, y = `Total Return (%)`, color = `Sector Name`)) +
  geom_point() +
  labs(
    title = "Total Return (%) vs ROIC (%)",
    x = "ROIC (%)",
    y = "Total Return (%)"
  ) +
  scale_y_continuous(labels = scales::percent) +
  scale_x_continuous(labels = scales::percent)

return_ROIC_scatter_w_filter <- drop_all_na_df %>%
  filter(`ROIC` >= -5) %>%
  ggplot(aes(x = `ROIC`, y = `Total Return (%)`, color = `Sector Name`)) +
  geom_point() +
  labs(
    y = "Total Return (%)"
  ) +
```

```
  scale_y_continuous(labels = scales::percent) +
  scale_x_continuous(labels = scales::percent) +
  theme(
    axis.title.x = element_blank()
  )

(return_ROIC_scatter + return_ROIC_scatter_w_filter +
  plot_layout(ncol = 1)) + plot_layout(guides = 'collect')
```



As seen above, the outliars have considerable weighing factor relative to the data points clustered together when it comes to explaining the variation in quarterly total return; hence we decided to remove them from the dataset since they only represent a small proportion of the dataset and are unlikely to drive meaningfully different results.

```
drop_all_na_df <- drop_all_na_df %>%
  filter(`ROIC` >= -5) %>%
  filter(`Free Cash Flow Yield` >= -2.5, `Free Cash Flow Yield` <= 2.5)
```

Then we tested our base model (Model A), which is discussed above and is shown below for convenience.

##Model A

$Y = \beta 0 + \beta 1 LEV + \beta 2 FCFY + \beta 3 ROIC + \epsilon$

(*See variable definitions above)

- Model A represents a Classical Linear Model (CLM) which was leveraged to build out the causal model. Note that CLM Assumptions are evaluated in the limitations section below.

```
short_model <- lm(`Total Return (%)` ~
               `Net Debt to EBITDA` +
               `Free Cash Flow Yield` +
```

```
              ROIC
            , data = drop_all_na_df)
```

We iterated on this model by adding a column which indicated whether or not the quarterly period of interest was in an inflationary environment. In doing so, we leveraged the F-test to ensure this incremental model provided additional information that was of statistical significance. Note that each of the underlying month-over-month inflationary periods were rolled up into quarterly periods by geometrically chain-linking the underlying data to reflect a quarterly timeframe. We show the refined model below for reference.

##Model B

$Y = \beta 0 + \beta 1 LEV + \beta 2 FCFY + \beta 3 ROIC + \beta 4 INF + \epsilon$

$$INF = \begin{cases} 0, QPCE < 0.0049 \\ 1, Otherwise \end{cases}$$

The dummy variable, INF, is such that the variable takes on a value of 1 if the observation occurred during the timeframe whereby the quarter over quarter percentage change of the Personal Consumption Expenditures Index (QPCE) was greater than or equal to 0.49%, and 0 otherwise.

In deciding between Model A and Model B [refer to the results panels], we decided that Model B provided a cleaner interpretation and a high R-squared. Further, we observed that in Model B the FCFY and INF variables were statistically significant, with t-tests beyond respective critical values.

```
short_model2 <- lm(`Total Return (%)` ~
              `Net Debt to EBITDA` +
              `Free Cash Flow Yield` +
              ROIC +
              `Inflationary Environment`
            , data = drop_all_na_df)
```

Observing Model B, we found FCFY and INF are significant variables and would like to observe if there is any interaction between FCFY and INF. With all three variables tested significant and being easy to interpret, we decided to select this model.

##Model C

$Y = \beta 0 + \beta 1 FCFY + \beta 2 INF + \beta 3 FCFY * INF + \epsilon$

$$INF = \begin{cases} 0, QPCE < 0.0049 \\ 1, Otherwise \end{cases}$$

```
short_model3 <- lm(`Total Return (%)` ~
              `Free Cash Flow Yield` +
              `Inflationary Environment` +
              `Free Cash Flow Yield`*`Inflationary Environment`
            , data = drop_all_na_df)
```

Then we iterated on this model further by creating three additional models. Again, we leveraged the F-test to ensure that each incremental model provided additional information that was of statistical significance. In Model D, shown below, we added sector dummy variables to Model B.

```
get_robust_se <- function(model) {
    sqrt(diag(sandwich::vcovHC(model)))
}

labels <- c("Net Debt", "FCFY", "ROIC", "INF", "FCFY:INF")

star = stargazer(short_model, short_model2, short_model3, covariate.labels = labels,
```

```
    title = "Results with current Total Return - short models",
    align = TRUE, type = "latex", header = FALSE, se = list(get_robust_se(short_model),
        get_robust_se(short_model2), get_robust_se(short_model3)),
    single.row = TRUE)
```

##Model D

In Model d, we created 11 additional dummy variables; one for each of the 11 sectors in the S&P 500 as defined by the S&P Dow Jones Indices Global Industry Classification Standards (GICS). We show each of the dummy variables and the corresponding GICS sector in the below table.

$Y = \beta 0 + \beta 1 LEV + \beta 2 FCFY + \beta 3 ROIC + \beta 4 INF + \beta 5 EN + \beta 6 MAT + \beta 7 IND + \beta 8 CD + \beta 9 CS + \beta 10 HC + \beta 11 FIN + \beta 12 IT + \beta 13 COM + \beta 14 UTIL + \beta 15 RE + \epsilon$

```
sector_names <- data.frame(Dummy_Variable=c('EN', 'MAT', 'IND', 'CD', 'CS', 'HC', 'FIN', 'IT', 'COM', 'UTI
sector_names
```

```
##    Dummy_Variable            GICS_Sector
## 1              EN                 Energy
## 2             MAT              Materials
## 3             IND            Industrials
## 4              CD Consumer Discretionary
## 5              CS       Consumer Staples
## 6              HC            Health Care
## 7             FIN             Financials
## 8              IT Information Technology
## 9             COM  Commication Services
## 10           UTIL              Utilities
## 11             RE            Real Estate
```

The sector dummy variables (EN, MAT, IND, CD, CS, HC, FIN, IT, COM, UTIL, RE) take the form of 0 if the constituent company of interest is not in the specified sector, and 1 otherwise.

$$EN, MAT, etc. = \begin{cases} 0, Company\ Not\ in\ GICS\ Sector \\ 1, Otherwise \end{cases}$$

```
model <- lm(`Total Return (%)` ~
              `Net Debt to EBITDA` +
              `Free Cash Flow Yield` +
              ROIC +
              `Inflationary Environment` +
              Materials +
              Financials +
              `Communication Services` +
              `Information Technology` +
              `Industrials` +
              `Energy` +
              `Consumer Staples` +
              `Health Care` +
              `Real Estate` +
              `Consumer Discretionary`
            , data = drop_all_na_df)
```

Our observed results suggested that Model D output a similar adjusted R-squared to Model B, but with the additional 11 dummy variables, Model D became incrementally more difficult to interpret. However, referencing the results panel, we can see that the energy sector's fundamentals do influence the underlying stock's quarterly total returns in inflationary environments on a statistically significant basis.

##Model E

In our next model iteration, we added interaction variables to Model D. That is, we added interaction terms for each fundamental metric and the inflationary environment dummy variable. We show the formulaic form of Model D below.

$Y = \beta0 + \beta1LEV + \beta2FCFY + \beta3ROIC + \beta4INF + \beta5EN + \beta6MAT + \beta7IND + \beta8CD + \beta9CS + \beta10HC + \beta11FIN + \beta12IT + \beta13COM + \beta14UTIL + \beta15RE + \beta16(INF * LEV) + \beta17(INF * FCFY) + \beta18(INF * ROIC) + \epsilon$

Again, the sector dummy variables (EN, MAT, IND, CD, CS, HC, FIN, IT, COM, UTIL, RE) take the form of 0 if the constituent company of interest is not in the specified sector, and 1 otherwise. The sector codes can be referenced in the table above.

$$EN, MAT, etc. = \begin{cases} 0, Company\ Not\ in\ GICS\ Sector \\ 1, Otherwise \end{cases}$$

```
model2 <- lm(`Total Return (%)` ~
               `Net Debt to EBITDA` +
               `Free Cash Flow Yield` +
               ROIC +
               `Inflationary Environment` +
               `Inflationary Environment`*`Net Debt to EBITDA` +
               `Inflationary Environment`*`Free Cash Flow Yield` +
               `Inflationary Environment`*ROIC +
               Materials +
               Financials +
               `Communication Services` +
               `Information Technology` +
               `Industrials` +
               `Energy` +
               `Consumer Staples` +
               `Health Care` +
               `Real Estate` +
               `Consumer Discretionary`
             , data = drop_all_na_df)
```

While the results panel suggested that Model E had a higher adjusted R-squared in comparison to Model C, the additional variables would cause for more challenging interpretability and thus, we decided against the trade-off.

##Model F Then we iterated on top of Model E by adding interaction variables between the sector dummy variables and the underlying fundamentals. We show the formula of the multiple regression below.

$Y = \beta0 + \beta1LEV + \beta2FCFY + \beta3ROIC + \beta4INF + \beta5EN + \beta6MAT + \beta7IND + \beta8CD + \beta9CS + \beta10HC + \beta11FIN + \beta12IT + \beta13CS + \beta14UTIL + \beta15RE + \beta16(INF * LEV) + \beta17(INF * FCFY) + \beta18(INF * ROIC) + \beta19(LEV * EN) + \beta20(LEV * MAT) + \beta21(LEV * IND) + \beta22(LEV * CD) + \beta23(LEV * CS) + \beta24(LEV * HC) + \beta25(LEV * FIN) + \beta26(LEV * IT) + \beta27(LEV * COM) + \beta28(LEV * UTIL) + \beta29(LEV * RE) + \beta30(FCFY * EN) + \beta31(FCFY * MAT) + \beta32(FCFY * IND) + \beta33(FCFY * CD) + \beta34(FCFY * CS) + \beta35(FCFY * HC) + \beta36(FCFY * FIN) + \beta37(FCFY * IT) + \beta38(FCFY * COM) + \beta39(FCFY * UTIL) + \beta40(FCFY * RE) + \beta41(ROIC * EN) + \beta42(ROIC * MAT) + \beta43(ROIC * IND) + \beta44(ROIC * CD) + \beta45(ROIC * CS) + \beta46(ROIC * HC) + \beta47(ROIC * FIN) + \beta48(ROIC * IT) + \beta49(ROIC * COM) + \beta50(ROIC * UTIL) + \beta51(ROIC * RE) + \epsilon$

```
model3 <- lm(`Total Return (%)` ~
               `Net Debt to EBITDA` +
               `Free Cash Flow Yield` +
               ROIC +
               `Inflationary Environment` +
               `Inflationary Environment`*`Net Debt to EBITDA` +
               `Inflationary Environment`*`Free Cash Flow Yield` +
               `Inflationary Environment`*ROIC +
               Materials +
               Materials*`Net Debt to EBITDA` +
```

```
              Materials*`Free Cash Flow Yield` +
              Materials*ROIC +
              Financials +
              Financials*`Net Debt to EBITDA` +
              Financials*`Free Cash Flow Yield` +
              Financials*ROIC +
              `Communication Services` +
              `Communication Services`*`Net Debt to EBITDA` +
              `Communication Services`*`Free Cash Flow Yield` +
              `Communication Services`*ROIC +
              `Information Technology` +
              `Information Technology`*`Net Debt to EBITDA` +
              `Information Technology`*`Free Cash Flow Yield` +
              `Information Technology`*ROIC +
              `Industrials` +
              `Industrials`*`Net Debt to EBITDA` +
              `Industrials`*`Free Cash Flow Yield` +
              `Industrials`*ROIC +
              `Energy` +
              `Energy`*`Net Debt to EBITDA` +
              `Energy`*`Free Cash Flow Yield` +
              `Energy`*ROIC +
              `Consumer Staples` +
              `Consumer Staples`*`Net Debt to EBITDA` +
              `Consumer Staples`*`Free Cash Flow Yield` +
              `Consumer Staples`*ROIC +
              `Health Care` +
              `Health Care`*`Net Debt to EBITDA` +
              `Health Care`*`Free Cash Flow Yield` +
              `Health Care`*ROIC +
              `Real Estate` +
              `Real Estate`*`Net Debt to EBITDA` +
              `Real Estate`*`Free Cash Flow Yield` +
              `Real Estate`*ROIC +
              `Consumer Discretionary` +
              `Consumer Discretionary`*`Net Debt to EBITDA` +
              `Consumer Discretionary`*`Free Cash Flow Yield` +
              `Consumer Discretionary`*ROIC
            , data = drop_all_na_df)
```

Again, Model F showed a higher adjusted R-squared, but the additional variables and lack of interpretability made this model not worth the trade off in comparison to Model C. However, in this iteration, we did find that the free cash flow yield can be seen as a very important indicator for high quarterly total returns in the information technology sector.

We show the F-tests of each of the models mentioned above below.

```
#  f-test
anova(short_model, model, test = "F")
```

```
## Analysis of Variance Table
##
## Model 1: `Total Return (%)` ~ `Net Debt to EBITDA` + `Free Cash Flow Yield` +
##     ROIC
## Model 2: `Total Return (%)` ~ `Net Debt to EBITDA` + `Free Cash Flow Yield` +
##     ROIC + `Inflationary Environment` + Materials + Financials +
##     `Communication Services` + `Information Technology` + Industrials +
##     Energy + `Consumer Staples` + `Health Care` + `Real Estate` +
```

```
##     `Consumer Discretionary`
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1  29706 6105.9
## 2  29695 6097.1 11    8.7891 3.8914 1.185e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(model, model2, test = "F")
```

```
## Analysis of Variance Table
##
## Model 1: `Total Return (%)` ~ `Net Debt to EBITDA` + `Free Cash Flow Yield` +
##     ROIC + `Inflationary Environment` + Materials + Financials +
##     `Communication Services` + `Information Technology` + Industrials +
##     Energy + `Consumer Staples` + `Health Care` + `Real Estate` +
##     `Consumer Discretionary`
## Model 2: `Total Return (%)` ~ `Net Debt to EBITDA` + `Free Cash Flow Yield` +
##     ROIC + `Inflationary Environment` + `Inflationary Environment` *
##     `Net Debt to EBITDA` + `Inflationary Environment` * `Free Cash Flow Yield` +
##     `Inflationary Environment` * ROIC + Materials + Financials +
##     `Communication Services` + `Information Technology` + Industrials +
##     Energy + `Consumer Staples` + `Health Care` + `Real Estate` +
##     `Consumer Discretionary`
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1  29695 6097.1
## 2  29692 6094.6  3    2.5601 4.1575 0.005935 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(model2, model3, test = "F")
```

```
## Analysis of Variance Table
##
## Model 1: `Total Return (%)` ~ `Net Debt to EBITDA` + `Free Cash Flow Yield` +
##     ROIC + `Inflationary Environment` + `Inflationary Environment` *
##     `Net Debt to EBITDA` + `Inflationary Environment` * `Free Cash Flow Yield` +
##     `Inflationary Environment` * ROIC + Materials + Financials +
##     `Communication Services` + `Information Technology` + Industrials +
##     Energy + `Consumer Staples` + `Health Care` + `Real Estate` +
##     `Consumer Discretionary`
## Model 2: `Total Return (%)` ~ `Net Debt to EBITDA` + `Free Cash Flow Yield` +
##     ROIC + `Inflationary Environment` + `Inflationary Environment` *
##     `Net Debt to EBITDA` + `Inflationary Environment` * `Free Cash Flow Yield` +
##     `Inflationary Environment` * ROIC + Materials + Materials *
##     `Net Debt to EBITDA` + Materials * `Free Cash Flow Yield` +
##     Materials * ROIC + Financials + Financials * `Net Debt to EBITDA` +
##     Financials * `Free Cash Flow Yield` + Financials * ROIC +
##     `Communication Services` + `Communication Services` * `Net Debt to EBITDA` +
##     `Communication Services` * `Free Cash Flow Yield` + `Communication Services` *
##     ROIC + `Information Technology` + `Information Technology` *
##     `Net Debt to EBITDA` + `Information Technology` * `Free Cash Flow Yield` +
##     `Information Technology` * ROIC + Industrials + Industrials *
##     `Net Debt to EBITDA` + Industrials * `Free Cash Flow Yield` +
##     Industrials * ROIC + Energy + Energy * `Net Debt to EBITDA` +
##     Energy * `Free Cash Flow Yield` + Energy * ROIC + `Consumer Staples` +
##     `Consumer Staples` * `Net Debt to EBITDA` + `Consumer Staples` *
##     `Free Cash Flow Yield` + `Consumer Staples` * ROIC + `Health Care` +
##     `Health Care` * `Net Debt to EBITDA` + `Health Care` * `Free Cash Flow Yield` +
```

```
##     `Health Care` * ROIC + `Real Estate` + `Real Estate` * `Net Debt to EBITDA` +
##     `Real Estate` * `Free Cash Flow Yield` + `Real Estate` *
##     ROIC + `Consumer Discretionary` + `Consumer Discretionary` *
##     `Net Debt to EBITDA` + `Consumer Discretionary` * `Free Cash Flow Yield` +
##     `Consumer Discretionary` * ROIC
##   Res.Df    RSS Df Sum of Sq     F    Pr(>F)
## 1  29692 6094.6
## 2  29662 6080.1 30     14.47 2.353 4.108e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
get_robust_se <- function(model) {
    sqrt(diag(sandwich::vcovHC(model)))
}

long_labels <- c("Net Debt", "FCFY", "ROIC", "INF", "Net Debt:INF",
    "FCFY:INF", "ROIC:INF", "Net Debt:Materials", "FCFY:Materials",
    "ROIC:Materials", "Net Debt:Financials", "FCFY:Financials",
    "ROIC:Financials", "Net Debt:Comm", "FCFY:Comm", "ROIC:Comm",
    "Net Debt:Info Tech", "FCFY:Info Tech", "ROIC:Info Tech",
    "Net Debt:Industrials", "FCFY:Industrials", "ROIC:Industrials",
    "Net Debt:Energy", "FCFY:Energy", "ROIC:Energy", "Net Debt:Con. S.",
    "FCFY:Con. S.", "ROIC:Con. S.", "Net Debt:Health Care", "FCFY:Health Care",
    "ROIC:Health Care", "Net Debt:Real Estate", "FCFY:Real Estate",
    "ROIC:Real Estate", "Net Debt:Con. D.", "FCFY:Con. D.", "ROIC:Con. D.",
    "Materials", "Financials", "Comm", "Info Tech", "Industrials",
    "Energy", "Con. S.", "Health Care", "Real Estate", "Con. D.")

star = stargazer(model, model2, model3, covariate.labels = long_labels,
    title = "Results with current Total Return - long models",
    align = TRUE, type = "latex", header = FALSE, se = list(get_robust_se(model),
        get_robust_se(model2), get_robust_se(model3)), single.row = TRUE)
```

##Models G, H, I, J, K, and L Finally, we created five additional models, which corresponded to each of the prior five models (A:G, B:H, C:I, D:J, E:K, and F:L), but with a different outcome variable, Y2. In each of these models, we tested the relationships against lagged quarterly total returns. We chose to explore this relationship because economically, the fundamentals, which drive each stock's intrinsic value, might take time to be reflected in the underlying stock prices and therefore total returns. In doing so, our aim was to explore if any of the underlying company fundamentals (LEV, FCFY, and ROIC) may have been statistically significantly linked to future total quarterly returns, specifically, one quarter ahead.

##Model G:

$Y2 = \beta 0 + \beta 1 LEV + \beta 2 FCFY + \beta 3 ROIC + \epsilon$

```
lagged_short_model <- lm(`Lagged Total Return (%)` ~
              `Net Debt to EBITDA` +
              `Free Cash Flow Yield` +
              ROIC
            , data = drop_all_na_df)
```

##Model H $Y2 = \beta 0 + \beta 1 LEV + \beta 2 FCFY + \beta 3 ROIC + \beta 4 INF + \epsilon$

$$INF = \begin{cases} 0, QPCE < 0.0049 \\ 1, Otherwise \end{cases}$$

The dummy variable, INF, is such that the variable takes on a value of 1 if the observation occurred during the timeframe whereby the quarter over quarter percentage change of the Personal Consumption Expenditures Index

(QPCE) was greater than or equal to 0.49%, and 0 otherwise.

```
lagged_short_model2 <- lm(`Lagged Total Return (%)` ~
              `Net Debt to EBITDA` +
              `Free Cash Flow Yield` +
              ROIC +
              `Inflationary Environment`
            , data = drop_all_na_df)
```

We see that in Model H, all variables besides LEV became statistically significant when explaining the one-quarter lagged total return of the respective stock while provided a significant boost over R-squared from all previous model. This model's R-squared jumped from Model C's 0.001 level to 0.008. Therefore, we can conclude that the FCFY, ROIC, and INF actually contain much more information on implied one-quarter ahead of future total quarterly returns than in the current sense.

##Model I

$Y2 = \beta 0 + \beta 1FCFY + \beta 2INF + \beta 3FCFY * INF + \epsilon$

$$INF = \begin{cases} 0, QPCE < 0.0049 \\ 1, Otherwise \end{cases}$$

```
lagged_short_model3 <- lm(`Lagged Total Return (%)` ~
              `Free Cash Flow Yield` +
              `Inflationary Environment` +
              `Free Cash Flow Yield`*`Inflationary Environment`
            , data = drop_all_na_df)
```

Due to Model C's significance in explaining the current total quarterly return, we want to explore its utility in explaining the one-quarter ahead future total quarterly return as well. However, the interaction term between FCFY and INF actually became insignificance. Moreover, the model's adjusted R-squared did not provide a significant boost from Model H. Therefore, we decided to select Model H as the primary model in explaining the future total returns.

```
get_robust_se <- function(model) {
    sqrt(diag(sandwich::vcovHC(model)))
}

labels <- c("Net Debt", "FCFY", "ROIC", "INF", "FCFY:INF")

star = stargazer(lagged_short_model, lagged_short_model2, lagged_short_model3,
    covariate.labels = labels, title = "Results with future Total Return - short models",
    align = TRUE, type = "latex", header = FALSE, se = list(get_robust_se(lagged_short_model),
        get_robust_se(lagged_short_model2), get_robust_se(lagged_short_model3)),
    single.row = TRUE)
```

```
cat(paste0(star), sep = "\n")
```

Table 1: Results with future Total Return - short models

	Dependent variable:		
	'Lagged Total Return (%)'		
	(1)	(2)	(3)
Net Debt	0.0005 (0.0005)	0.0004 (0.0005)	
FCFY	0.135 *** (0.030)	0.127 *** (0.030)	0.177 *** (0.044)
ROIC	-0.030 (0.019)	-0.028 (0.019)	
INF		-0.077 *** (0.005)	-0.073 *** (0.006)
FCFY:INF			-0.096 (0.059)
Constant	0.098 *** (0.004)	0.140 *** (0.005)	0.135 *** (0.004)
Observations	29,710	29,710	29,710
R^2	0.001	0.008	0.008
Adjusted R^2	0.001	0.008	0.008
Residual Std. Error	0.452 (df = 29706)	0.450 (df = 29705)	0.450 (df = 29706)
F Statistic	7.324*** (df = 3; 29706)	59.916*** (df = 4; 29705)	79.582*** (df = 3; 29706)
Note:			*p<0.1; **p<0.05; ***p<0.01

##Model J

$Y2 = \beta0 + \beta1LEV + \beta2FCFY + \beta3ROIC + \beta4INF + \beta5EN + \beta6MAT + \beta7IND + \beta8CD + \beta9CS + \beta10HC + \beta11FIN + \beta12IT + \beta13COM + \beta14UTIL + \beta15RE + \epsilon$

```
sector_names <- data.frame(Dummy_Variable=c('EN', 'MAT', 'IND', 'CD', 'CS', 'HC', 'FIN', 'IT', 'COM', 'UTI
sector_names
```

```
##    Dummy_Variable            GICS_Sector
## 1              EN                 Energy
## 2             MAT              Materials
## 3             IND            Industrials
## 4              CD Consumer Discretionary
## 5              CS       Consumer Staples
## 6              HC            Health Care
## 7             FIN             Financials
## 8              IT Information Technology
## 9             COM  Communication Services
## 10           UTIL              Utilities
## 11             RE            Real Estate
```

The sector dummy variables (EN, MAT, IND, CD, CS, HC, FIN, IT, COM, UTIL, RE) take the form of 0 if the constituent company of interest is not in the specified sector, and 1 otherwise.

$$EN, MAT, etc. = \begin{cases} 0, Company\ Not\ in\ GICS\ Sector \\ 1, Otherwise \end{cases}$$

```
lagged_model <- lm(`Lagged Total Return (%)` ~
              `Net Debt to EBITDA` +
              `Free Cash Flow Yield` +
              ROIC +
              `Inflationary Environment` +
              Materials +
              Financials +
              `Communication Services` +
              `Information Technology` +
              `Industrials` +
```

```
               `Energy` +
               `Consumer Staples` +
               `Health Care` +
               `Real Estate` +
               `Consumer Discretionary`
             , data = drop_all_na_df)
```

We see that in Model G, FCFY and INF are still statistically significant when explaining the one-quarter lagged total return of the respective stock where as the interaction term between FCY and INF did not offer any statistical significance. Its R-squared value remains low and combined with its difficulty to interpret makes it not an ideal candidate over Model I

##Model K

$Y2 = \beta 0 + \beta 1LEV + \beta 2FCFY + \beta 3ROIC + \beta 4INF + \beta 5EN + \beta 6MAT + \beta 7IND + \beta 8CD + \beta 9CS + \beta 10HC + \beta 11FIN + \beta 12IT + \beta 13COM + \beta 14UTIL + \beta 15RE + \beta 16(INF * LEV) + \beta 17(INF * FCFY) + \beta 18(INF * ROIC) + \epsilon$

Again, the sector dummy variables (EN, MAT, IND, CD, CS, HC, FIN, IT, COM, UTIL, RE) take the form of 0 if the constituent company of interest is not in the specified sector, and 1 otherwise. The sector codes can be referenced in the table above.

$$EN, MAT, etc. = \begin{cases} 0, Company\ Not\ in\ GICS\ Sector \\ 1, Otherwise \end{cases}$$

```
lagged_model2 <- lm(`Lagged Total Return (%)` ~
               `Net Debt to EBITDA` +
               `Free Cash Flow Yield` +
               ROIC +
               `Inflationary Environment` +
               `Inflationary Environment`*`Net Debt to EBITDA` +
               `Inflationary Environment`*`Free Cash Flow Yield` +
               `Inflationary Environment`*ROIC +
               Materials +
               Financials +
               `Communication Services` +
               `Information Technology` +
               `Industrials` +
               `Energy` +
               `Consumer Staples` +
               `Health Care` +
               `Real Estate` +
               `Consumer Discretionary`
             , data = drop_all_na_df)
```

##Model L

$Y2 = \beta 0 + \beta 1LEV + \beta 2FCFY + \beta 3ROIC + \beta 4INF + \beta 5EN + \beta 6MAT + \beta 7IND + \beta 8CD + \beta 9CS + \beta 10HC + \beta 11FIN + \beta 12IT + \beta 13CS + \beta 14UTIL + \beta 15RE + \beta 16(INF * LEV) + \beta 17(INF * FCFY) + \beta 18(INF * ROIC) + \beta 19(LEV * EN) + \beta 20(LEV * MAT) + \beta 21(LEV * IND) + \beta 22(LEV * CD) + \beta 23(LEV * CS) + \beta 24(LEV * HC) + \beta 25(LEV * FIN) + \beta 26(LEV * IT) + \beta 27(LEV * COM) + \beta 28(LEV * UTIL) + \beta 29(LEV * RE) + \beta 30(FCFY * EN) + \beta 31(FCFY * MAT) + \beta 32(FCFY * IND) + \beta 33(FCFY * CD) + \beta 34(FCFY * CS) + \beta 35(FCFY * HC) + \beta 36(FCFY * FIN) + \beta 37(FCFY * IT) + \beta 38(FCFY * COM) + \beta 39(FCFY * UTIL) + \beta 40(FCFY * RE) + \beta 41(ROIC * EN) + \beta 42(ROIC * MAT) + \beta 43(ROIC * IND) + \beta 44(ROIC * CD) + \beta 45(ROIC * CS) + \beta 46(ROIC * HC) + \beta 47(ROIC * FIN) + \beta 48(ROIC * IT) + \beta 49(ROIC * COM) + \beta 50(ROIC * UTIL) + \beta 51(ROIC * RE) + \epsilon$

The sector dummy variables (EN, MAT, IND, CD, CS, HC, FIN, IT, COM, UTIL, RE) take the form of 0 if the constituent company of interest is not in the specified sector, and 1 otherwise.

$$EN, MAT, etc. = \begin{cases} 0, Company\ Not\ in\ GICS\ Sector \\ 1, Otherwise \end{cases}$$

```r
lagged_model3 <- lm(`Lagged Total Return (%)` ~
              `Net Debt to EBITDA` +
              `Free Cash Flow Yield` +
              ROIC +
              `Inflationary Environment` +
              `Inflationary Environment`*`Net Debt to EBITDA` +
              `Inflationary Environment`*`Free Cash Flow Yield` +
              `Inflationary Environment`*ROIC +
              Materials +
              Materials*`Net Debt to EBITDA` +
              Materials*`Free Cash Flow Yield` +
              Materials*ROIC +
              Financials +
              Financials*`Net Debt to EBITDA` +
              Financials*`Free Cash Flow Yield` +
              Financials*ROIC +
              `Communication Services` +
              `Communication Services`*`Net Debt to EBITDA` +
              `Communication Services`*`Free Cash Flow Yield` +
              `Communication Services`*ROIC +
              `Information Technology` +
              `Information Technology`*`Net Debt to EBITDA` +
              `Information Technology`*`Free Cash Flow Yield` +
              `Information Technology`*ROIC +
              `Industrials` +
              `Industrials`*`Net Debt to EBITDA` +
              `Industrials`*`Free Cash Flow Yield` +
              `Industrials`*ROIC +
              `Energy` +
              `Energy`*`Net Debt to EBITDA` +
              `Energy`*`Free Cash Flow Yield` +
              `Energy`*ROIC +
              `Consumer Staples` +
              `Consumer Staples`*`Net Debt to EBITDA` +
              `Consumer Staples`*`Free Cash Flow Yield` +
              `Consumer Staples`*ROIC +
              `Health Care` +
              `Health Care`*`Net Debt to EBITDA` +
              `Health Care`*`Free Cash Flow Yield` +
              `Health Care`*ROIC +
              `Real Estate` +
              `Real Estate`*`Net Debt to EBITDA` +
              `Real Estate`*`Free Cash Flow Yield` +
              `Real Estate`*ROIC +
              `Consumer Discretionary` +
              `Consumer Discretionary`*`Net Debt to EBITDA` +
              `Consumer Discretionary`*`Free Cash Flow Yield` +
              `Consumer Discretionary`*ROIC
            , data = drop_all_na_df)
```

We show the F-tests of each of the models mentioned above below.

```r
#  f-test
anova(lagged_short_model, lagged_model, test = "F")
```

```
## Analysis of Variance Table
##
## Model 1: `Lagged Total Return (%)` ~ `Net Debt to EBITDA` + `Free Cash Flow Yield` +
##     ROIC
## Model 2: `Lagged Total Return (%)` ~ `Net Debt to EBITDA` + `Free Cash Flow Yield` +
##     ROIC + `Inflationary Environment` + Materials + Financials +
##     `Communication Services` + `Information Technology` + Industrials +
##     Energy + `Consumer Staples` + `Health Care` + `Real Estate` +
##     `Consumer Discretionary`
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1  29706 6063.8
## 2  29695 6012.2 11    51.662 23.197 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(lagged_model, lagged_model2, test = "F") # I reject lagged model 2 here
```

```
## Analysis of Variance Table
##
## Model 1: `Lagged Total Return (%)` ~ `Net Debt to EBITDA` + `Free Cash Flow Yield` +
##     ROIC + `Inflationary Environment` + Materials + Financials +
##     `Communication Services` + `Information Technology` + Industrials +
##     Energy + `Consumer Staples` + `Health Care` + `Real Estate` +
##     `Consumer Discretionary`
## Model 2: `Lagged Total Return (%)` ~ `Net Debt to EBITDA` + `Free Cash Flow Yield` +
##     ROIC + `Inflationary Environment` + `Inflationary Environment` *
##     `Net Debt to EBITDA` + `Inflationary Environment` * `Free Cash Flow Yield` +
##     `Inflationary Environment` * ROIC + Materials + Financials +
##     `Communication Services` + `Information Technology` + Industrials +
##     Energy + `Consumer Staples` + `Health Care` + `Real Estate` +
##     `Consumer Discretionary`
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1  29695 6012.2
## 2  29692 6010.8  3    1.3342 2.1969 0.08618 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(lagged_model2, lagged_model3, test = "F")
```

```
## Analysis of Variance Table
##
## Model 1: `Lagged Total Return (%)` ~ `Net Debt to EBITDA` + `Free Cash Flow Yield` +
##     ROIC + `Inflationary Environment` + `Inflationary Environment` *
##     `Net Debt to EBITDA` + `Inflationary Environment` * `Free Cash Flow Yield` +
##     `Inflationary Environment` * ROIC + Materials + Financials +
##     `Communication Services` + `Information Technology` + Industrials +
##     Energy + `Consumer Staples` + `Health Care` + `Real Estate` +
##     `Consumer Discretionary`
## Model 2: `Lagged Total Return (%)` ~ `Net Debt to EBITDA` + `Free Cash Flow Yield` +
##     ROIC + `Inflationary Environment` + `Inflationary Environment` *
##     `Net Debt to EBITDA` + `Inflationary Environment` * `Free Cash Flow Yield` +
##     `Inflationary Environment` * ROIC + Materials + Materials *
##     `Net Debt to EBITDA` + Materials * `Free Cash Flow Yield` +
##     Materials * ROIC + Financials + Financials * `Net Debt to EBITDA` +
##     Financials * `Free Cash Flow Yield` + Financials * ROIC +
```

```
##     `Communication Services` + `Communication Services` * `Net Debt to EBITDA` +
##     `Communication Services` * `Free Cash Flow Yield` + `Communication Services` *
##     ROIC + `Information Technology` + `Information Technology` *
##     `Net Debt to EBITDA` + `Information Technology` * `Free Cash Flow Yield` +
##     `Information Technology` * ROIC + Industrials + Industrials *
##     `Net Debt to EBITDA` + Industrials * `Free Cash Flow Yield` +
##     Industrials * ROIC + Energy + Energy * `Net Debt to EBITDA` +
##     Energy * `Free Cash Flow Yield` + Energy * ROIC + `Consumer Staples` +
##     `Consumer Staples` * `Net Debt to EBITDA` + `Consumer Staples` *
##     `Free Cash Flow Yield` + `Consumer Staples` * ROIC + `Health Care` +
##     `Health Care` * `Net Debt to EBITDA` + `Health Care` * `Free Cash Flow Yield` +
##     `Health Care` * ROIC + `Real Estate` + `Real Estate` * `Net Debt to EBITDA` +
##     `Real Estate` * `Free Cash Flow Yield` + `Real Estate` *
##     ROIC + `Consumer Discretionary` + `Consumer Discretionary` *
##     `Net Debt to EBITDA` + `Consumer Discretionary` * `Free Cash Flow Yield` +
##     `Consumer Discretionary` * ROIC
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1  29692 6010.8
## 2  29662 6002.7 30    8.1638 1.3447 0.09857 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
get_robust_se <- function(model) {
    sqrt(diag(sandwich::vcovHC(model)))
}

long_labels <- c("Net Debt", "FCFY", "ROIC", "INF", "Net Debt:INF",
    "FCFY:INF", "ROIC:INF", "Net Debt:Materials", "FCFY:Materials",
    "ROIC:Materials", "Net Debt:Financials", "FCFY:Financials",
    "ROIC:Financials", "Net Debt:Comm", "FCFY:Comm", "ROIC:Comm",
    "Net Debt:Info Tech", "FCFY:Info Tech", "ROIC:Info Tech",
    "Net Debt:Industrials", "FCFY:Industrials", "ROIC:Industrials",
    "Net Debt:Energy", "FCFY:Energy", "ROIC:Energy", "Net Debt:Con. S.",
    "FCFY:Con. S.", "ROIC:Con. S.", "Net Debt:Health Care", "FCFY:Health Care",
    "ROIC:Health Care", "Net Debt:Real Estate", "FCFY:Real Estate",
    "ROIC:Real Estate", "Net Debt:Con. D.", "FCFY:Con. D.", "ROIC:Con. D.",
    "Materials", "Financials", "Comm", "Info Tech", "Industrials",
    "Energy", "Con. S.", "Health Care", "Real Estate", "Con. D.")

star = stargazer(lagged_model, lagged_model2, lagged_model3,
    covariate.labels = long_labels, title = "Results with future Total Return - long models",
    align = TRUE, type = "latex", header = FALSE, se = list(get_robust_se(lagged_model),
        get_robust_se(lagged_model2), get_robust_se(lagged_model3)),
    single.row = TRUE)
```

```
cat(paste0(star), sep = "\n")
```

Table 2: Results with future Total Return - long models

	Dependent variable:		
	'Lagged Total Return (%)'		
	(1)	(2)	(3)
Net Debt	0.0002 (0.0005)	0.001 * (0.001)	0.0002 (0.003)
FCFY	0.143 *** (0.031)	0.201 *** (0.045)	0.017 (0.126)
ROIC	-0.030 (0.019)	-0.005 (0.027)	-0.083 (0.215)
INF	-0.077 *** (0.005)	-0.064 *** (0.008)	-0.065 *** (0.008)
Net Debt:INF	-0.039 ** (0.016)	-0.040 *** (0.016)	-0.056 * (0.031)
FCFY:INF	0.001 (0.016)	-0.001 (0.016)	-0.007 (0.026)
ROIC:INF	-0.044 ** (0.017)	-0.045 *** (0.017)	-0.073 ** (0.031)
Net Debt:Materials	-0.011 (0.013)	-0.012 (0.013)	-0.014 (0.025)
FCFY:Materials	-0.012 (0.013)	-0.013 (0.013)	-0.015 (0.026)
ROIC:Materials	-0.061 *** (0.017)	-0.062 *** (0.017)	-0.090 *** (0.030)
Net Debt:Financials	-0.014 (0.015)	-0.015 (0.015)	-0.012 (0.036)
FCFY:Financials	-0.014 (0.013)	-0.015 (0.013)	-0.040 (0.027)
ROIC:Financials	0.012 (0.016)	0.012 (0.016)	-0.014 (0.052)
Net Debt:Comm	-0.028 ** (0.013)	-0.029 ** (0.013)	-0.038 (0.025)
FCFY:Comm		-0.002 * (0.001)	-0.001 (0.001)
ROIC:Comm		-0.100 * (0.060)	-0.094 (0.060)
Net Debt:Info Tech		-0.045 (0.038)	-0.044 (0.036)
FCFY:Info Tech			-0.003 (0.004)
ROIC:Info Tech			0.568 ** (0.247)
Net Debt:Industrials			0.033 (0.246)
FCFY:Industrials			-0.002 (0.004)
ROIC:Industrials			0.193 (0.128)
Net Debt:Energy			0.095 (0.216)
FCFY:Energy			0.002 (0.003)
ROIC:Energy			0.540 ** (0.227)
Net Debt:Con. S.			0.046 (0.252)
FCFY:Con. S.			0.0003 (0.003)
ROIC:Con. S.			0.239 (0.175)
Net Debt:Health Care			0.037 (0.216)
FCFY:Health Care			0.001 (0.003)
ROIC:Health Care			0.058 (0.145)
Net Debt:Real Estate			0.097 (0.226)
FCFY:Real Estate			0.007 (0.005)
ROIC:Real Estate			0.136 (0.161)
Net Debt:Con. D.			0.197 (0.260)
FCFY:Con. D.			-0.008 (0.007)
ROIC:Con. D.			0.287 (0.198)
Materials			0.080 (0.245)
Financials			-0.001 (0.003)
Comm			0.655 *** (0.197)
Info Tech			0.025 (0.227)
Industrials			0.004 (0.006)
Energy			0.257 (0.174)
Con. S.			0.063 (0.451)
Health Care			0.001 (0.003)
Real Estate			0.033 (0.150)
Con. D.			0.143 (0.219)
Constant	0.157 *** (0.012)	0.150 *** (0.012)	0.157 *** (0.022)
Observations	29,710	29,710	29,710
R^2	0.009	0.009	0.011

Adjusted R^2	0.009	0.009	0.009
Residual Std. Error	0.450 (df = 29695)	0.450 (df = 29692)	0.450 (df = 29662)
F Statistic	19.809*** (df = 14; 29695)	16.703*** (df = 17; 29692)	6.902*** (df = 47; 29662)
Note:			*p<0.1; **p<0.05; ***p<0.01

#4. Results

#5. Limitations ##5a.Statistical Limitations

We concluded that the model H which has the equation $Y2 = \beta 0 + \beta 1 FCFY + \beta 2 INF + \beta 3 FCFY * INF + \epsilon$ gave us the most information with the ease of interpretation.

1. **IID Data:**

There are factors such as time series data and sector clustering that makes the data violate the independent assumptions. However, in an economic model these assumptions are generally hard to hold true and less prevalent. As for identically distributed, we argue that they are subjected to the same economic atmosphere and therefore, their distribution should be identical.

2. **No Perfect Colinearity:**

```
library(car)
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```
## The following object is masked from 'package:purrr':
##
##     some
```

```
vif(lagged_short_model2)
```

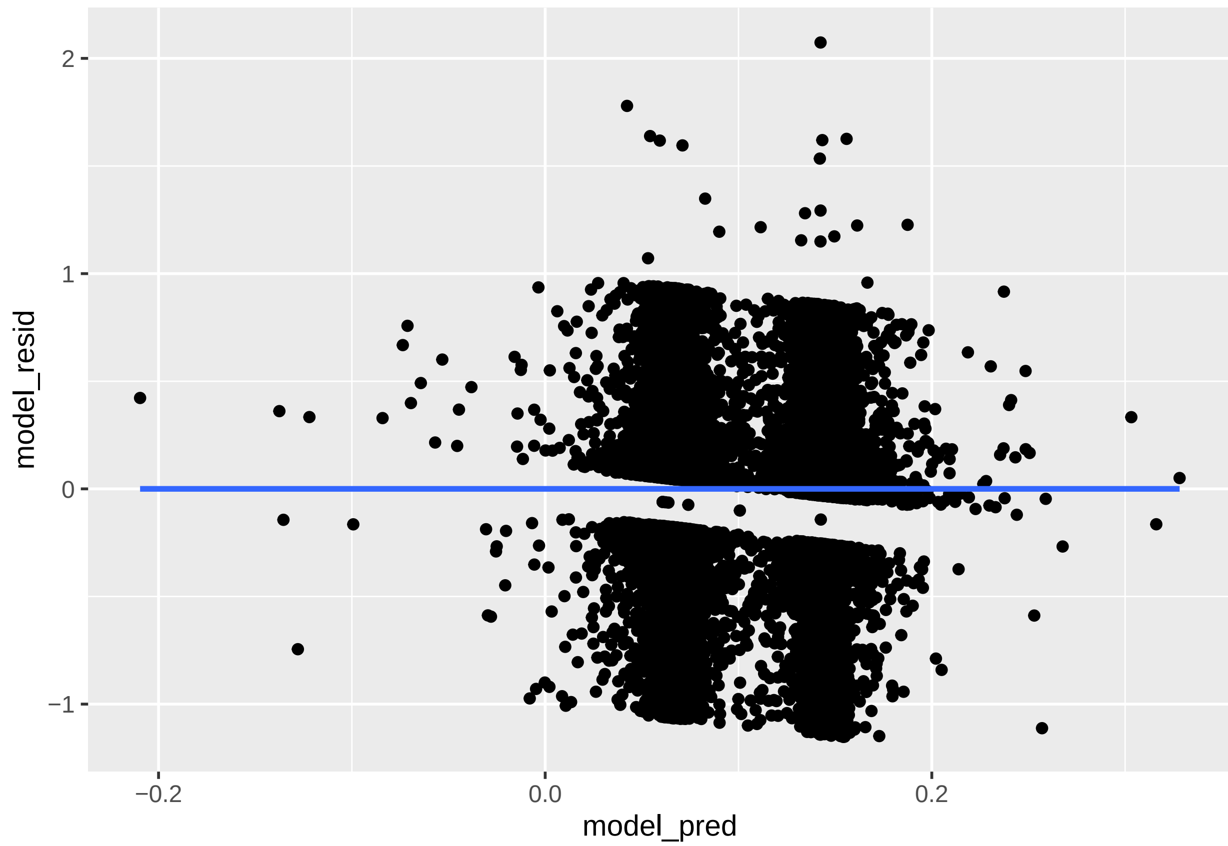
```
##       `Net Debt to EBITDA`     `Free Cash Flow Yield`
##                   1.027951                   1.015652
##                       ROIC `Inflationary Environment`
##                   1.024822                   1.000618
```

The VIF result shows no param has an above 10 score meaning there should not be any colinearity issue.

3. **Linear Conditional Expectation:**

```
drop_all_na_df %>%
  mutate(
    model_pred = predict(lagged_short_model2),
    model_resid = resid(lagged_short_model2)
  ) %>%
  ggplot(aes(model_pred, model_resid)) +
  geom_point() +
  geom_smooth()
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

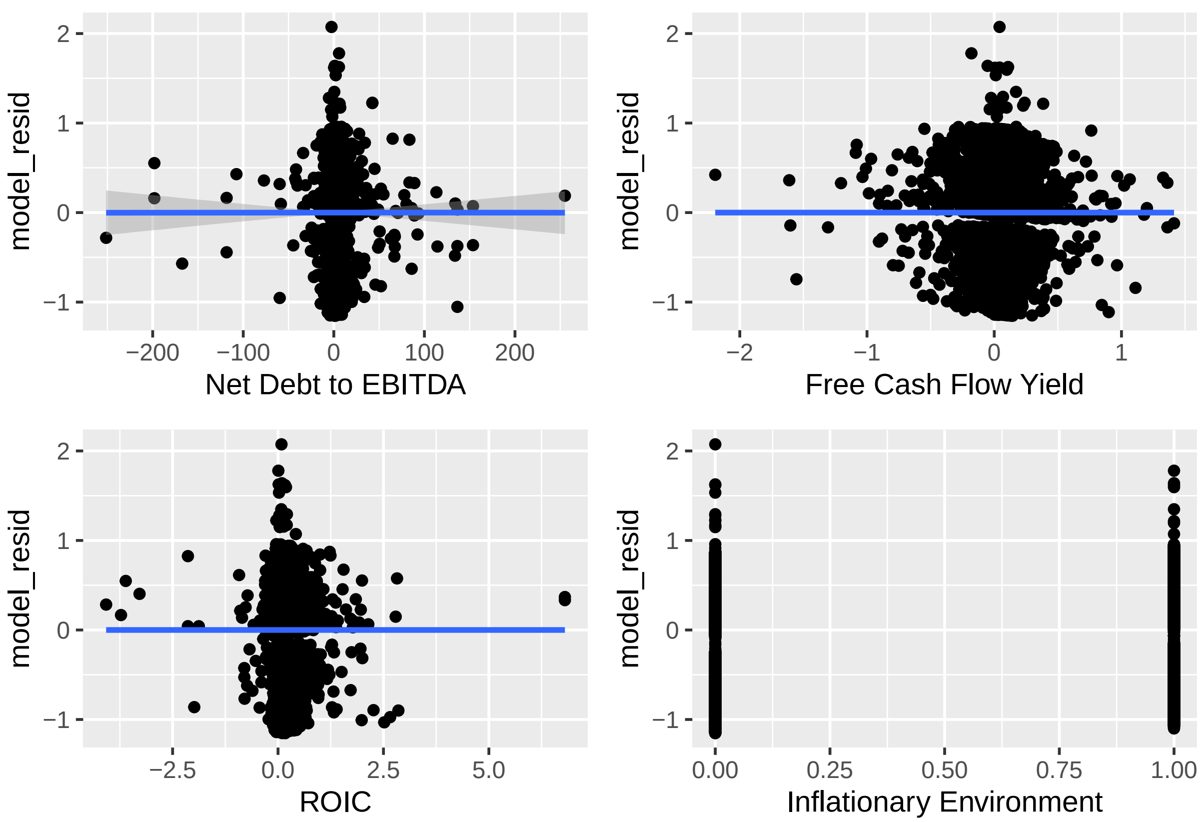



```
LEV_resid <- drop_all_na_df %>%
  mutate(
    model_resid = resid(lagged_short_model2)
  ) %>%
  ggplot(aes(`Net Debt to EBITDA`, model_resid)) +
  geom_point() +
  stat_smooth()

FCFY_resid <- drop_all_na_df %>%
  mutate(
    model_resid = resid(lagged_short_model2)
  ) %>%
  ggplot(aes(`Free Cash Flow Yield`, model_resid)) +
  geom_point() +
  stat_smooth()

ROIC_resid <- drop_all_na_df %>%
  mutate(
    model_resid = resid(lagged_short_model2)
  ) %>%
  ggplot(aes(`ROIC`, model_resid)) +
  geom_point() +
  stat_smooth()

INF_resid <- drop_all_na_df %>%
  mutate(
    model_resid = resid(lagged_short_model2)
  ) %>%
  ggplot(aes(`Inflationary Environment`, model_resid)) +
```

```
  geom_point() +
  stat_smooth()

(LEV_resid + FCFY_resid + ROIC_resid + INF_resid +
  plot_layout(ncol = 2))
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

```
## Warning: Computation failed in `stat_smooth()`:
## x has insufficient unique values to support 10 knots: reduce k.
```

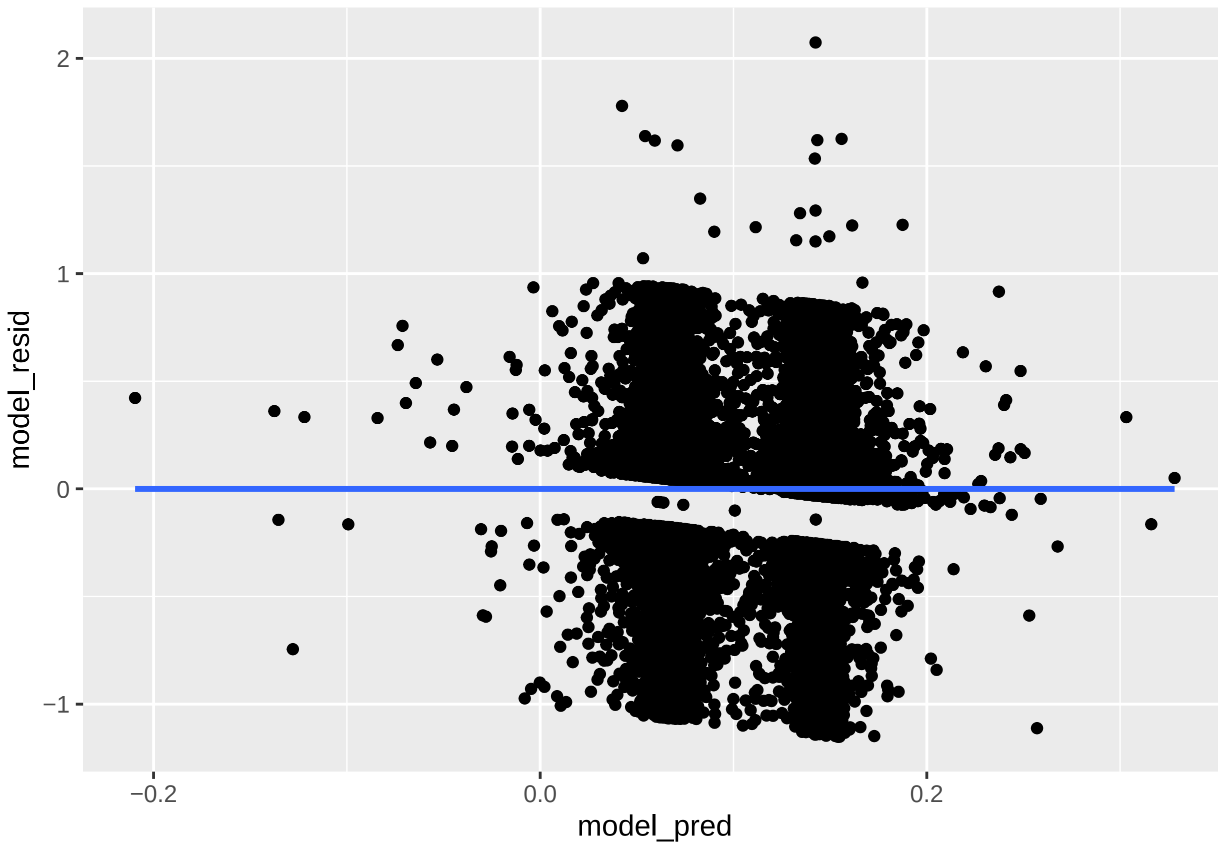


The running the model residual against the model predicted values, it seems to suggest there is a straight line hence the linear conditional expectation is not violated

4. **Homoskedastic Errors:**

```
drop_all_na_df %>%
  mutate(
    model_pred = predict(lagged_short_model2),
    model_resid = resid(lagged_short_model2)
  ) %>%
  ggplot(aes(model_pred, model_resid)) +
  geom_point() +
  geom_smooth()
```

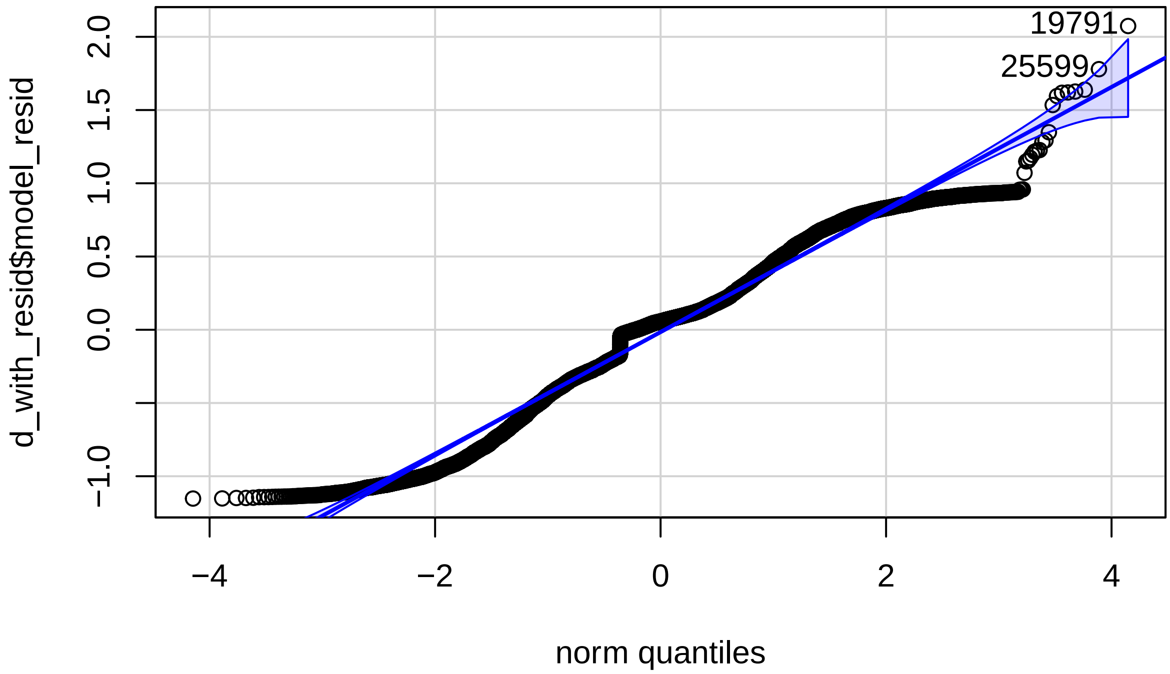
```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



The gray area (the confidence interval) of the residual does have even thickness across the graphs suggesting homoskedastic error assumption is not violated.

5. **Normally Distributed Errors:**

```
d_with_resid <- drop_all_na_df %>%
  mutate(
    model_resid = resid(lagged_short_model2)
  )

qqPlot(d_with_resid$model_resid)
```




```
## [1] 19791 25599
```

Plotting residuals on the QQ-plot, we can see that many plots fall outside of the confidence interval of the QQ line. Therefore, we can conclude that the assumption of normally distributed error is violated.

##5b. Structural Limitations

There are several important variables that are likely important which we were not able to measure in our analysis. There is a multitude of financial data that is publicly available; however, as a group, we decided that the most important omitted variables in our models included operating margin growth, change in percentage of buybacks, and growth in dividends. The latter two are mathematically linked to total returns while the former is thought to be a future predictor of returns assuming the underlying company is managed by a capable C-Suite.

We consider each of the three main omitted variables and the potential bias by leveraging the below equation.

$Estimate = TrueParameter + OmittedBias$ $\alpha 1 = \beta 1 + \beta 2 * \delta 1$

In each case, we see the three omitted variables as having overestimated the true parameter. This is because in each case the omitted variables can be observed as having a positive relationship with quarterly total returns, and thus we are moving away from zero.

Note that with more time, we would have also liked to make adjustments to the figures based on the Management's Discussion and Analysis in the Financial Statements as well as one-off or non-recurring items. Finally, we would have liked to also have considered the potential for reverse causality within our model.