

Analysis of Stock Price Variance Close To Annual And Quarterly Report Filing Dates

Nelson Corrocher

Harrisburg University

Author Note

Nelson Corrocher, M.B.A.

Correspondence concerning this article should be addressed to Nelson Corrocher.

E-mail: ncorrocher@my.harrisburgu.edu

Abstract

This paper provides an analysis of the differences between stock price variation that occurs regularly and the variation that happens close to 10-Q and 10-K filing dates. Using publicly available information, variation for regular days and for days close to the filing release dates were calculated and then compared using a same-variance hypothesis test along with the graphical plotting. The results showed that, in general, the variance of a stock price close to its filing date is higher. The implications of this study can lead to a systematic straddle creation strategy that has potential to provide simplicity of execution and higher Sharpe Ratio than passive index strategies.

Keywords: straddle options financial reports price variance

Word count: 107

Analysis of Stock Price Variance Close To Annual And Quarterly Report Filing Dates

During a trading simulation in a graduate investment class at Boston University in 2014, the winning strategy used an option straddle close to the release date of the 10-K of a certain public company. This outcome started a discussion of using a systematic creation of straddles close to filing dates of companies. In the context of financial investment, a straddle is a neutral option strategy involving the simultaneous buying of a put and a call of the same underlying stock, expecting a great price variation before the option expiration date. Creating an option straddle close to the filing date of the financial statement from publicly traded companies has been used by fund managers to try to improve their funds returns. However, since they use additional and often undisclosed criteria to decide whose company's stocks they straddle, there is neither evidence that the variance is indeed greater close to financial report filing dates, nor that the strategy, by itself, is profitable. The goal of this research is to compare the differences in stock price variation that happens on a daily basis to those that happens close to 10-Q and 10-K filing dates using publicly available information. For the validation of the assumption that price variation is indeed higher close to filing dates, five years of historical stock prices will be collected for the top S&P index composing companies, also called Blue Chips (which have greater options liquidity - used in straddles) along with inflation rates and SEC filing dates. The stock prices will be adjusted to control for inflation and their daily price variation will be computed. The filings dates from each company will be used to mark the days that we expect to see greater variation. In this reserach we are using day minus 2 to day plus 2 on the filling dates to compose our test sample. The remaining days will be used as the control sample. Next, both samples will be first compared using a same-variance hypothesis test along with the graphical plotting to show the differences visually, if any. Based on the strategy used by the fund managers, it is expected that the test sample would show greater variance and thus, enabling further research into applying this concept in formulating a strategy and running simulations. Formulating the strategy and running the simulations are not in this paper's scope and will

be done at a later time.

Methods

Participants

No direct participants were involved in this research as the main inputs for the analysis were financial market data that are publicly available.

Measures and Procedures

This entire paper has been constructed almost entirely by using R scripting (exception for Inflation Daily Rates, explained below). For replicability purposes, the methodology is going to be explained through the commented script, enabling the reader to change it for his own needs. The code below is used to prepare the environment for the script. **Note:** the SETWD function in the block above sets the working directory and should be set to the folder where the inputs and outputs would be located.

```
list.of.packages <- c("ggplot2", "xlsx", "data.table")
for(i in list.of.packages) {
  if(!is.installed(i)) {
    install.packages(i)
  }
}

library(data.table)
library(xlsx)
library(ggplot2)
```

In the section below we select the stocks being used in the Analysis using their tickers. The stocks used are the top comprising blue chips that are part of the S&P index. As long as yahoo finance website its data, any ticker is valid to be added below. To add a new one,

put the ticker and the company name used on the SEC filings (case-sensitive) in the array below. To get the correct ticker and company filing, visit

<https://www.sec.gov/edgar/searchedgar/companysearch.html>. **Note:** In some cases, companies change their filing names (for example, AA - Alcoa). This requires a code modification, also explained further below.

```
stocks <- data.frame(  
  c(  
    "IBM",  
    "XOM",  
    "CVX",  
    "PG",  
    "MMM",  
    "JNJ",  
    "MCD",  
    "WMT",  
    "UTX",  
    "KO",  
    "BA",  
    "CAT",  
    "JPM",  
    "VZ",  
    "T",  
    "DD",  
    "MRK",  
    "DIS",  
    "HD",  
    "MSFT",
```

```
    "AXP",  
    "BAC",  
    "PFE",  
    "GE",  
    "INTC",  
    "C",  
    "GM"  
  )  
)  
stocks <- data.frame(  
  stocks,  
  c(  
    "INTERNATIONAL BUSINESS MACHINES CORP",  
    "EXXON MOBIL CORP",  
    "CHEVRON CORP",  
    "PROCTER & GAMBLE",  
    "3M CO",  
    "JOHNSON & JOHNSON",  
    "MCDONALDS CORP",  
    "WAL MART STORES INC",  
    "UNITED TECHNOLOGIES CORP",  
    "COCA COLA CO",  
    "BOEING CO",  
    "CATERPILLAR INC",  
    "JPMORGAN CHASE & CO",  
    "VERIZON COMMUNICATIONS INC",  
    "AT&T INC.",  
  )  
)
```

```

"DUPONT E I DE NEMOURS & CO",
"Merck & Co. Inc.",
"WALT DISNEY CO/",
"HOME DEPOT INC",
"MICROSOFT CORP",
"AMERICAN EXPRESS CO",
"BANK OF AMERICA CORP /DE/",
"PFIZER INC",
"GENERAL ELECTRIC CO",
"INTEL CORP",
"CITIGROUP INC",
"General Motors Co"
)
)
colnames(stocks) <- c("ticker", "company")

```

The code below imports data from yahoo finance website and download historical prices for previously selected stocks, consolidating all information in a single table. **Note:** At the time of writing this paper, this code was functional. However, should yahoo decide to change the website interface, the code may need tweaks to work correctly. The dates have been hardcoded in the section below (&a=8&b=1&c=2011 means August 1st, 2011). They can be changed to use other periods. In this research, the field adjusted closing (Adj Closing) price was the one selected for the price change calculations.

```

firstRun = TRUE
for (t in stocks$ticker) {
  URL <-
  paste(

```

```

    "http://chart.finance.yahoo.com/table.csv?s=",
    t,
    "&a=8&b=1&c=2011&d=7&e=31&f=2016&g=d&ignore=.csv",
    # Hard-coded dates
    sep = ""
  )
tmp <-
  fread(URL,
        drop = c("Open", "High", "Low", "Close", "Volume"))
tmp <- data.frame(tmp, t, 0)
colnames(tmp) <- c("date", "price", "ticker", "flag")
if (firstRun == TRUE) {
  historicalPrices <- data.frame(tmp)
  firstRun <- FALSE
} else
  historicalPrices <- rbind(historicalPrices, tmp)
}
rm(tmp)
rm(firstRun)
rm(t)
rm(URL)

```

There wasn't a good free formatted online source for daily US inflation rates so it was manually estimated using excel and distributing the monthly rate daily, including weekends and holidays. These daily rates are used to adjust stock prices for inflation. The method used is to change the daily price by daily rate (For example, if the stock price was \$100 and inflation rate was 1%, the adjusted price would be \$101). In practice, the inflation rate didn't change the values significantly for the periods considered.


```
fn <- "dailyInflationUS.xlsx"

if (!file.exists(fn)) {
  download.file(URL, fn, method="auto", cacheOK = FALSE)
}

inflationRates <-
  read.xlsx(
    fn,
    sheetIndex = 1,
    colIndex = c(1, 4),
    stringsAsFactors = FALSE
  )
```

Note:The table used can be downloaded from
<https://drive.google.com/open?id=0B2BIWokA6cNdM0RmQ1k1eFlaQWs>

The code below consolidates the historical prices and US inflation rates in the same file, excluding all the dates that are outside the interest range.

```
tmp <-
  merge(historicalPrices,
        inflationRates,
        by = "date",
        all = TRUE)
tmp$date <- as.Date(tmp$date, "%Y-%m-%d")
tmp <-
  subset(tmp,
        date >= "2011/09/01" &
```

```

        date <= "2016/07/31" &
        !is.na(ticker)
    )
historicalPricesInflation <- tmp
rm(tmp)
rm(inflationRates)
rm(historicalPrices)

```

Next, the price changes are calculated by the formula $(d + 1 - d) / (d)$, where d is the price on a specific day. In sequence, the price changes are then normalized to the normal distribution.

```

tmp <- setorder(historicalPricesInflation,
               "ticker",
               "date"
            )
tmp <- data.frame(tmp, 0)
colnames(tmp)[6] <- "change"

for (i in 1:nrow(tmp)) {
  if (i == 1 || tmp[i,"ticker"] != tmp[i - 1,"ticker"] ) {
    tmp[i, "change"] <- NA
  } else {
    tmp[i, "change"] <-
      (tmp[i, "price"] -
       tmp[i - 1, "price"] *
       (1 + tmp[i, "rate"])) /
      tmp[i-1,"price"]*(1 + tmp[i, "rate"])
  }
}

```

```
}  
}  
adjustedPriceChanges <- subset(tmp,!is.na(change))  
rm(i)  
rm(tmp)  
  
for (t in stocks$ticker) {  
  tmp <- subset(adjustedPriceChanges, ticker == t)  
  tmp$change <- scale(tmp$change)  
  if (!exists("scaledPriceChanges"))  
    scaledPriceChanges <- tmp  
  else  
    scaledPriceChanges <- rbind(scaledPriceChanges, tmp)  
}  
rm(t)
```

The following step captures filing dates from SEC.org website in the period considered (the period can be changed in the two “*for*” lines, where **i** is year and **j** is quarter). As now the code doesn’t check for the filing existence so in the case it doesn’t, the script will return an error. Next, consolidate all lines together, filtering date by filing type, 10-Q and 10-K, and by companies, the ones declared at the beggining. There is a known issue: since the names companies use to file their financial statements can have small changes, the operator %like% is used so that the script can look inside the fields for contained text instead of equal text. The problem with this approach is that sometimes it brings up more than one company, which needs separated treatment.

```
firstRun <- TRUE
for (i in 2011:2016) {
  for (j in 1:4) {
    if (!(i == 2016 & j == 4)) {
      URL <-
        paste("ftp://ftp.sec.gov/edgar/full-index/",
              i,
              "/QTR",
              j,
              "/master.zip",
              sep = "")

      FNC <- paste("master_QTR", j, "_", i, ".zip", sep = "")
      FNU <- paste("master_QTR", j, "_", i, ".idx", sep = "")
      if (!file.exists(FNC))
        download.file(URL, FNC, method = "libcurl")
      unzip(FNC, "master.idx")
      file.rename("master.idx", FNU)
      if (firstRun == TRUE) {
        filingDates <- fread(FNU)
        filingDates <-
          subset(filingDates, V3 == "10-K" | V3 == "10-Q")
        firstRun <- FALSE
      } else {
        tmp <- fread(FNU)
        tmp <- subset(tmp, V3 == "10-K" | V3 == "10-Q")
        filingDates <- rbind(filingDates, tmp)
      }
    }
  }
}
```

```

    }
  }
}

colnames(filingDates) <- c("CIK","company","filing","date","location")
rm(i)
rm(j)
rm(URL)
rm(firstRun)
rm(tmp)

for (n in 1:nrow(stocks)) {
  tmp <-
    data.frame(subset(filingDates, company %like% stocks$company[n]),
               stocks$ticker[n])
  if (!exists("relevantFilings"))
    relevantFilings <- tmp
  else
    relevantFilings <- rbind(relevantFilings, tmp)
}

```

This is the place to treat the exceptions above. In this example, four additional companies were brought in the sample which are not related to the analysis and must be eliminated. **Note:** If other stocks are added, this section may need to include other incorrectly included companies.

```

relevantFilings <-
  subset(
    relevantFilings,

```

```

company != "STRUCTURED PRODUCTS CORP CORTS TRUST FOR BOEING CO NOTES" &
company != "STRATS SM TRUST FOR JPMORGAN CHASE & CO SEC SERIES 2004-9" &
company != "PORTLAND GENERAL ELECTRIC CO /OR/" &
company != "CINTEL CORP"
)

```

The section below joins the data from the filings and from historical pricing.

```

relevantFilings <-
  data.frame(as.Date(relevantFilings$date, "%Y-%m-%d"),
    relevantFilings$stocks.ticker.n.,
    1)
colnames(relevantFilings) <- c("date", "ticker", "flag")
scaledPriceChanges <-
  data.frame(scaledPriceChanges$date,
    scaledPriceChanges$ticker,
    scaledPriceChanges$change)
colnames(scaledPriceChanges) <- c("date", "ticker", "change")

finalData <-
  merge(
    scaledPriceChanges,
    relevantFilings,
    by = c("date", "ticker"),
    all.x = TRUE
  )
for (i in 1:nrow(finalData)) {
  if (is.na(finalData$flag[i]))

```

```
finalData$flag[i] <- 0
}
```

This section marks the days that the dates affected by the “Release” of the filings. For this project, it was considered day - 2 to day + 2. These days are marked and will constitute the subset called “Release” in the final analysis. The remaining days will be treated as “Regular” days. After this last step, the data will be ready for analysis.

Results

Below the variance of both subsets is shown.

```
## [1] "Regular set variance: 0.99"
```

```
## [1] "Release set variance: 1.08"
```

On the same data, a hypothesis test for the two populations to have the same variance is ran:

```
##
```

```
## F test to compare two variances
```

```
##
```

```
## data: finalData.reg and finalData.fin
```

```
## F = 0.9, num df = 30000, denom df = 3000, p-value = 0.004
```

```
## alternative hypothesis: true ratio of variances is not equal to 1
```

```
## 95 percent confidence interval:
```

```
## 0.87 0.97
```

```
## sample estimates:
```

```
## ratio of variances
```

```
## 0.92
```

The graphs of the distribution densities are shown in Figure 1 through 4 for visual comparison of the two samples.

Discussion

The objective of this study was to show the existence or not of any relation between price change variance. To this end, the variance of both subsets were shown and a basic variance test was used to compare the two populations. Results above shows that not only the variance of the two populations were different, with the “Release” set being higher than the “Regular” but the p-value from the hypothesis test allows us to reject the main hypothesis that the two populations have the same variance. Figures 1 to 4 show that while the distribution around the center is similar for both populations, the distribution gets very different in the tails especially below -5 and above 5. These results suggest a potential strategy for option straddle for two main reasons. First, there is indeed a greater variation close to the release date, important for neutral straddles. Secondly, difference is notoriously on the graphs tails, as the cost to form a straddle with puts and calls with strike price far from the current are significantly cheaper, increasing the potential for returns. In summary, the results shown here provide the basis for the next steps: to check the price variance in time, to collect options data on analyzed stocks and to run simulations for the different possible straddle strategies. If the results are optimistic still optimistic, re-running the analysis on a different set of stocks for backtesting would confirm the feasibility of the strategy.

References

Session Info

```
sessionInfo()
```

```
## R version 3.3.2 (2016-10-31)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 7 x64 (build 7601) Service Pack 1
##
## locale:
## [1] LC_COLLATE=English_United States.1252
## [2] LC_CTYPE=English_United States.1252
## [3] LC_MONETARY=English_United States.1252
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.1252
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## other attached packages:
## [1] ggplot2_2.2.1      xlsx_0.5.7          xlsxjars_0.6.1      rJava_0.9-8
## [5] data.table_1.10.0  papaja_0.1.0.9477
##
## loaded via a namespace (and not attached):
## [1] Rcpp_0.12.9      knitr_1.15.1      magrittr_1.5       munsell_0.4.3
## [5] colorspace_1.3-2 highr_0.6          stringr_1.1.0      plyr_1.8.4
## [9] tools_3.3.2      grid_3.3.2        gtable_0.2.0       htmltools_0.3.5
## [13] assertthat_0.1   yaml_2.1.14       lazyeval_0.2.0     rprojroot_1.2
```

```
## [17] digest_0.6.12      tibble_1.2          bookdown_0.3        evaluate_0.10
## [21] rmarkdown_1.3       labeling_0.3         stringi_1.1.2        scales_0.4.1
## [25] backports_1.0.5
```

Aust, F., & Barth, M. (2016). *Papaja: Create apa manuscripts with rmarkdown*. Retrieved from <https://github.com/crsh/papaja>

Berk, P. M., Jonathan BDeMarzo. (n.d.). *Corporate finance*.

Bodie, Z., Kane, A., & Marcus, A. J. (n.d.). *Investments*.

Dowle, M., Srinivasan, A., Short, T., R Saporta, S. L. with contributions from, & Antonyan, E. (2015). *Data.table: Extension of data.frame*. Retrieved from <https://CRAN.R-project.org/package=data.table>

Dragulescu, A. A. (2014a). *Xlsx: Read, write, format excel 2007 and excel 97/2000/xp/2003 files*. Retrieved from <https://CRAN.R-project.org/package=xlsx>

Dragulescu, A. A. (2014b). *Xlsxjars: Package required poi jars for the xlsx package*. Retrieved from <https://CRAN.R-project.org/package=xlsxjars>

Investopedia. (2003). *Straddle definition*. Retrieved from <http://www.investopedia.com/terms/s/straddle.asp>

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>

Sec.gov | company search page_2016. (2016). *Sec.gov*. Retrieved from <https://www.sec.gov/edgar/searchedgar/companysearch.html>

Sec.org | information for ftp users - 2016. (2016). *Sec.gov*. Retrieved from <https://www.sec.gov/edgar/searchedgar/ftpusers.htm>

Urbanek, S. (2016). *RJava: Low-level r to java interface*. Retrieved from <https://CRAN.R-project.org/package=rJava>

Wickham, H. (2009). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York.

Retrieved from <http://ggplot2.org>

Yahoo finance - business finance, stock market, quotes, news_2016. (2016).

Finance.yahoo.com. Retrieved from <http://finance.yahoo.com/>

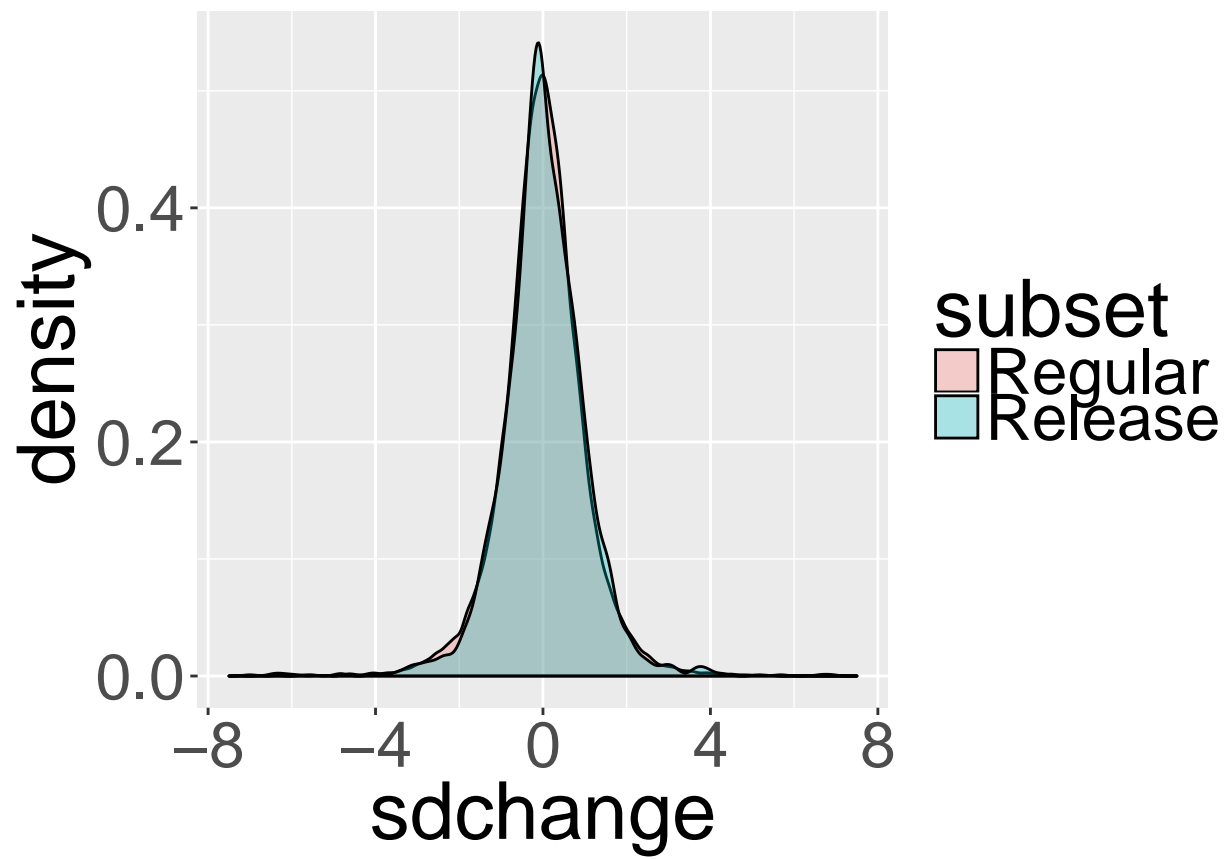


Figure 1. Superposed distribution density of the two datasets.

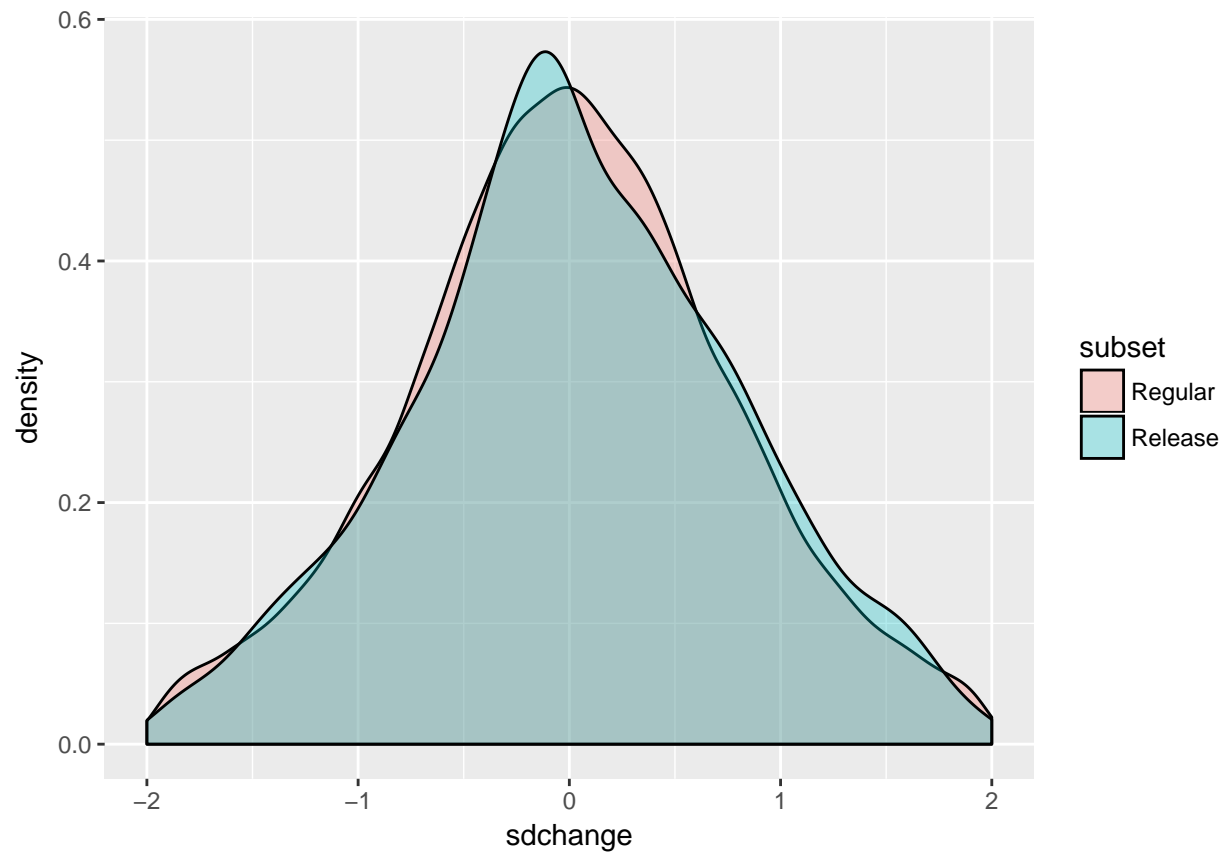


Figure 2. Distribution density zoomed around the center.

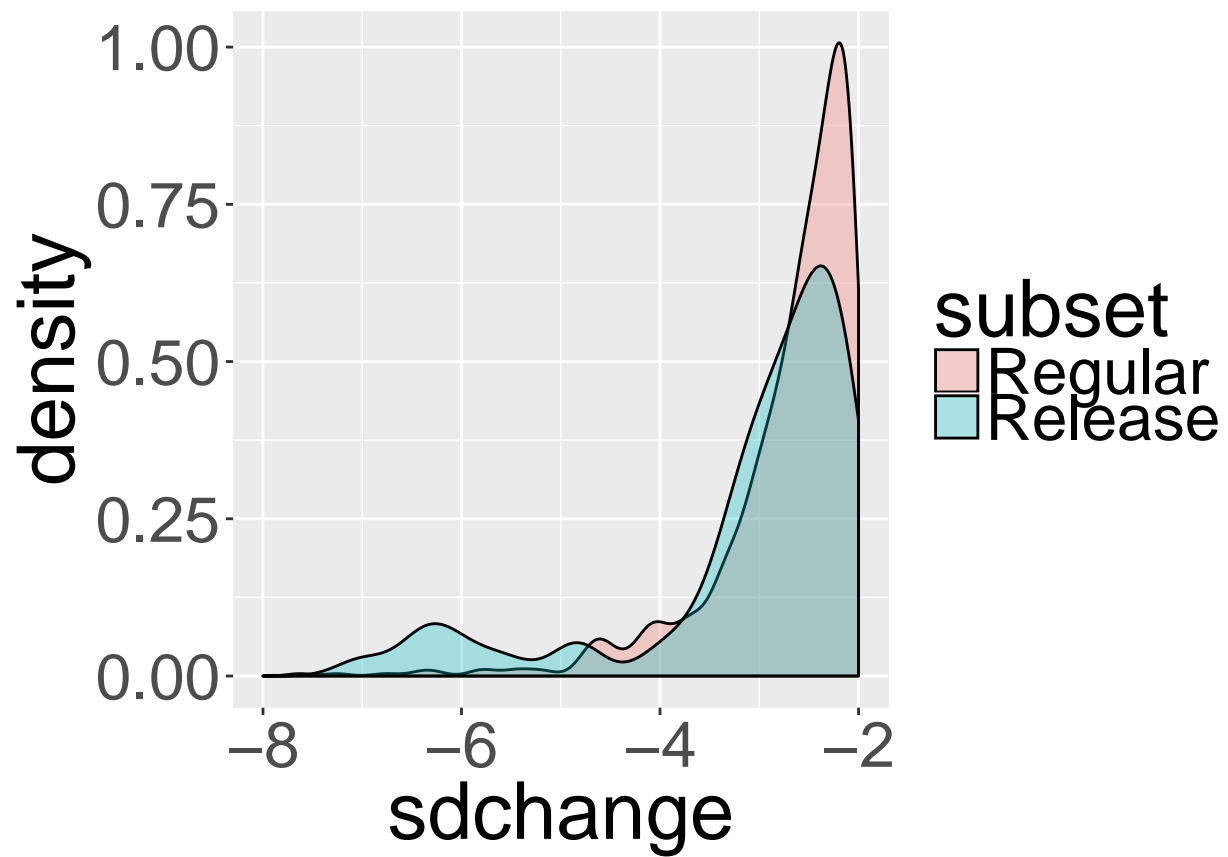


Figure 3. Distribution density zoomed on the left tail.

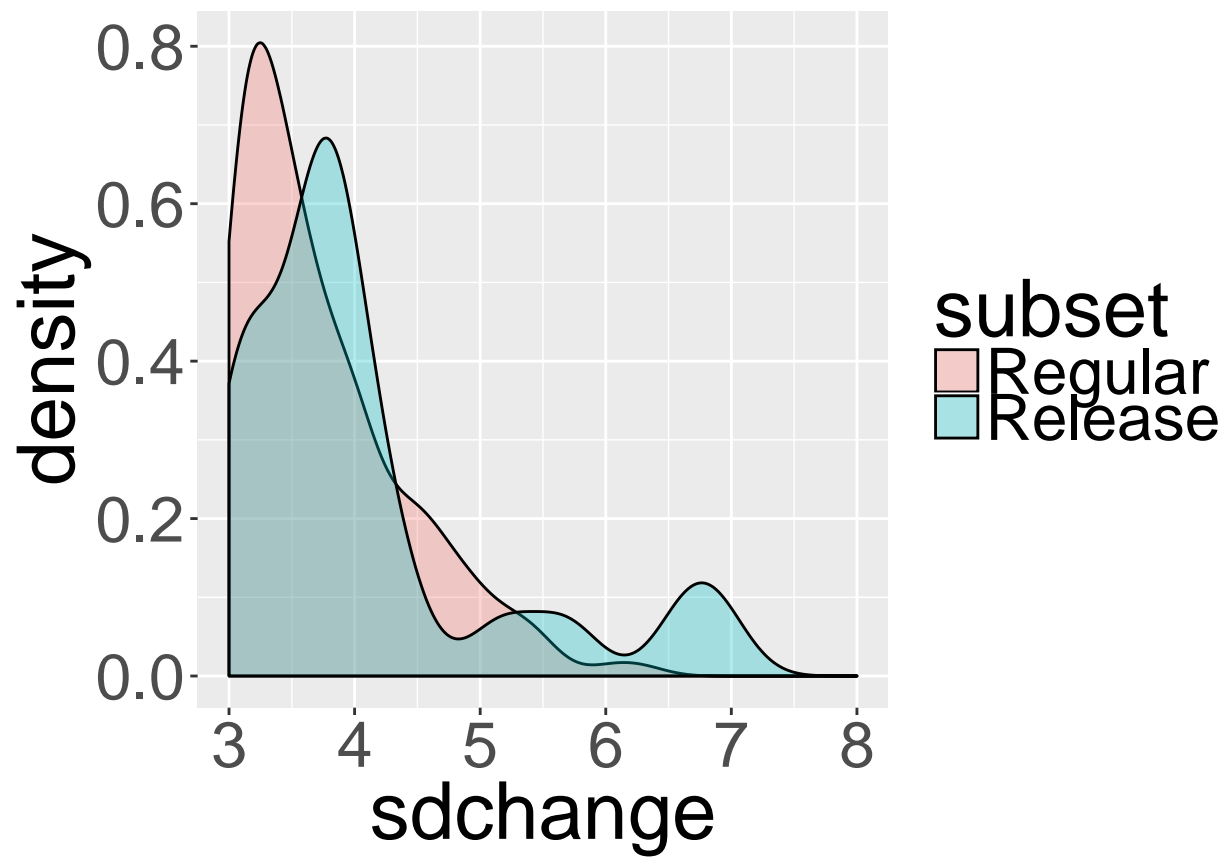


Figure 4. Distribution density zoomed on the right tail.