Online Appendix: Extraction and Cleaning of IPO Data from SDC Platinum

Michelle Lowry, Roni Michaely, Ekaterina Volkova

August 9, 2017

Table of Contents

Introduction 2
Extracting Data from SDC Platinum 2
SDC Data Selection 2
SDC Report 5
Cleaning SDC Request Data 6
Loading of the SDC Request 6
Preview of loaded SDC request. 7
Excluding Extra Observations 8
Exclude Non-IPO Offerings 8
Keep only Offerings of Common Shares 8
Exclude REITs 9
Exclude ADRs 9
Exclude Closed-End Funds 9
Exclude Units 9
Exclude Penny Stocks 10
Matching SDC and CRSP database 10
Exclude Observations w/o CRSP Record 11
Variable Construction 12
Initial Returns 12
Proceeds 12
Nominal Proceeds 12
Real Proceeds in 2016 dollars. 12
Initial Price Range 13
Syndicate Members 15
Registration Period 15
Matching SDC and SEC EDGAR 15
SEC EDGAR IPO Filings 15
IPO CIK Codes 17
Correction to SDC Data 17

Introduction

This is an Online Appendix to Lowry, Michaely, and Volkova, 2017, "Initial public offerings: A synthesis of the literature and directions for future research", in Foundations and Trends in Finance (working paper version available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2912354).

During the work on the review chapter, we found variety of features in the SDC Platinum data that influence the completeness of an output IPO sample. In this Appendix, we provide an extensive explanation on the construction of the data sample together with `R` codes. Full codes used to construct the review chapter sample, figures and tables are available at https://github.com/volkovacodes/IPO-Review-Chapter.

We are very thankful to Jay Ritter for emphasizing the sensitive points in the data construction.

Extracting Data from SDC Platinum

SDC Data Selection

We pull the sample of IPO firms from the SDC Platinum database. The request from SDC satisfies the following criteria:

1. Select "Common Stock" in "Global New Issues Database" (Figure A.1)
2. Select dates between 01/01/1973 and 12/31/2016 (Figure A.2)
3. Exclude companies whose 'Primary Exchange Nation' is any country other than United States (Figure A.3 and Figure A.4)[1]
4. These filters result in over 40,000 observations (Figure A.5)[2]
5. Extract the following variables in the SDC request: Filing date, Issue date, Issuer, State, Nation, Offer price, Security, Description, REIT, Unit, Depositary, Deal number, Closed-End Fund, CUSIP, CUSIP9, Proceeds, Original Low/High Filing Price, Amendment History of Low/High Filing Price, VC, Gross spread, Manager codes, Tech industry, IPO Flag, Original IPO Flag. (Figure A.6.)

[1] SDC records the primary exchange listing of a company in one of three ways: a specific country, 'unknown', or blank. If one simply selects IPOs whose primary exchange listing is the United States, one misses all IPOs for which this variable is blank (many of which went public on a US exchange). To avoid this problem, after choosing the variable 'primary exchange listing', select all countries and then de-select the United States. Importantly, one must deselect United States within both of the regions within which it appears: the 'Americas' region and the 'North America' region. See Figure A.4 for an illustration. After this step, we obtained between 42,000 and 44,000 observations, depending on the SDC version / interface

[2] We do not limit out initial SDC request to "Common Shares" only because some security offerings are recorded as "Ordinary Shares" or "Class A Shares" and thus would be inappropriately filtered out from the request. Table A.2 of this Appendix shows the most common types of securities.

Figure A.1

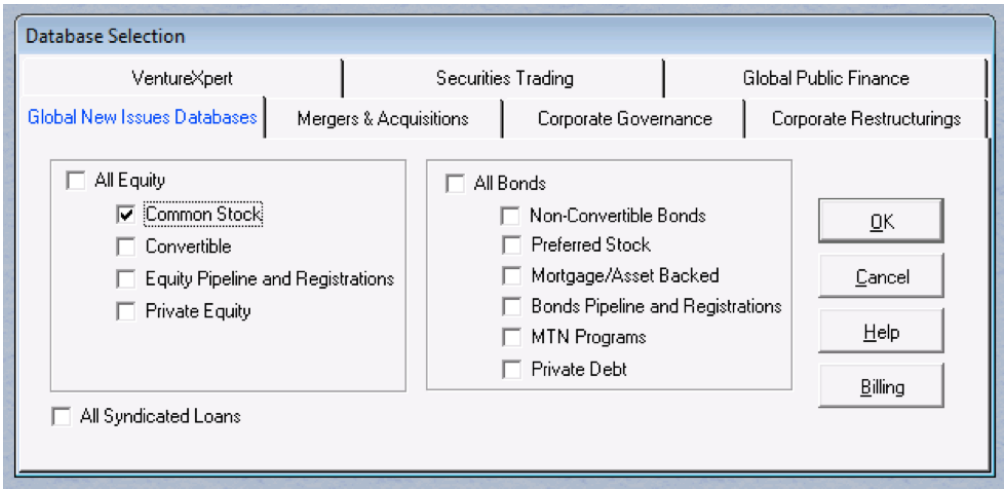

Figure A.2

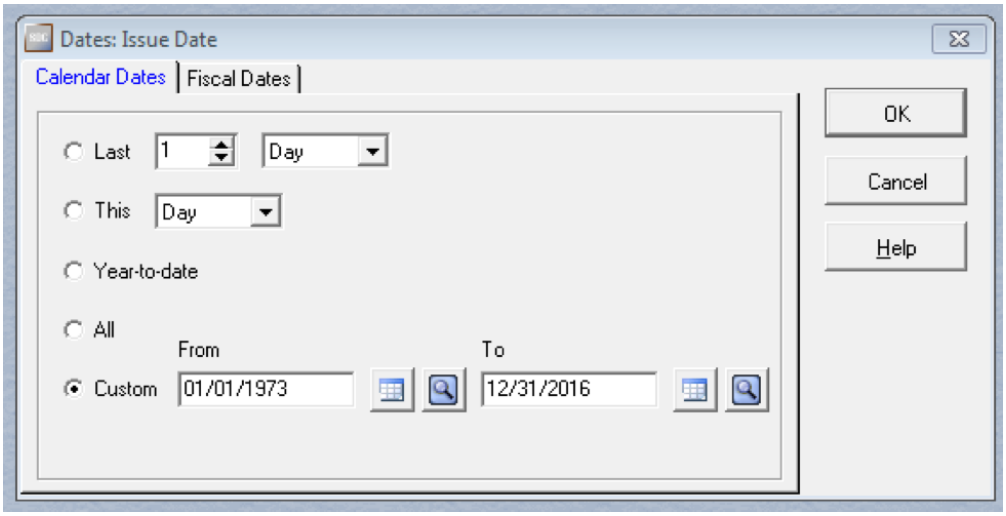

Figure A.3

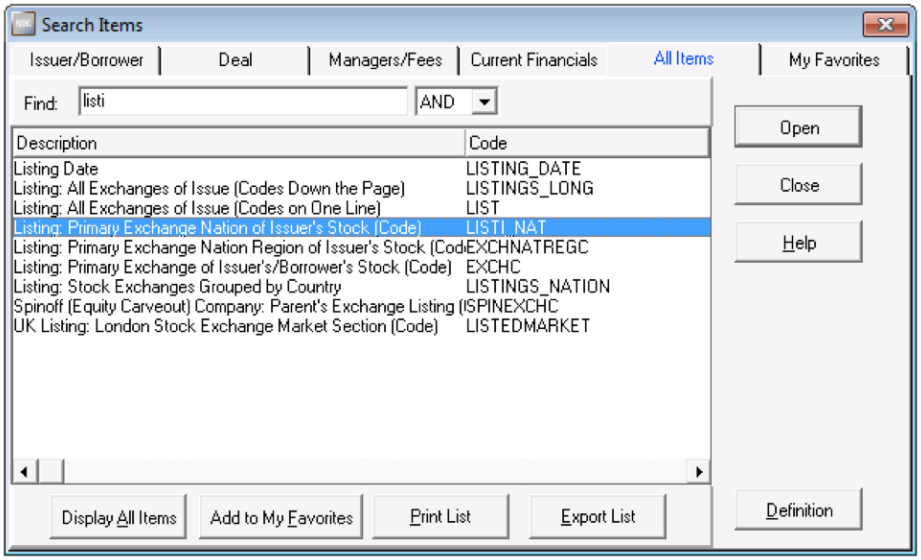

Figure A.4

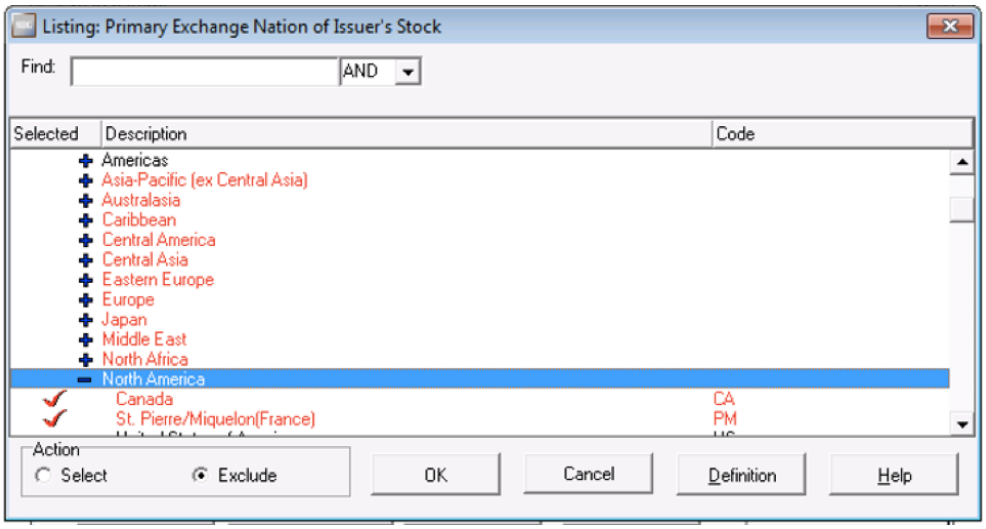

Figure A.5

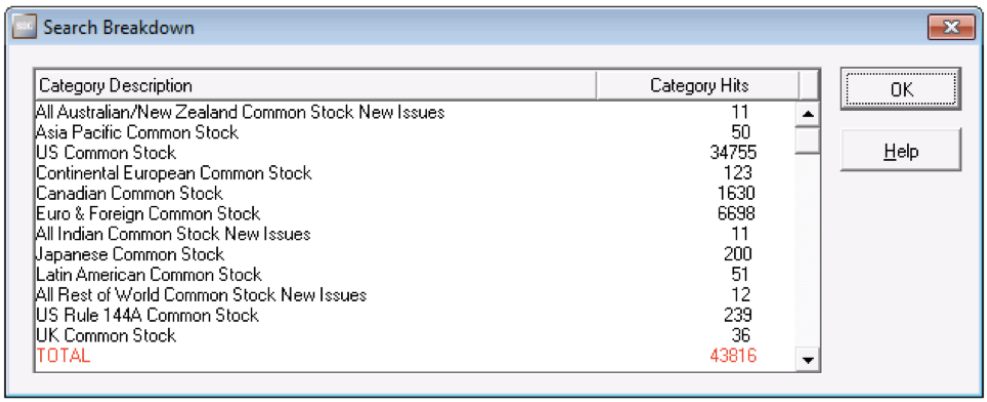

Figure A.6

SDC Report

```
Deal Type: NI
Author: ev9
Created: 7/15/2017 2:57:29 PM
Comments:



Custom Report Contents:
FILED
D
I
ST
NAT
IPO
ORIG_IPO
P
SECUR
DESCR
REIT_TYPE
UIT
DEPOSITARY
DEAL_NO
CLOSED_END_FUND_TRUST
CU
CUSIP9
PROCDS
VE
GPCT
ALLMGRROLECODE
HITECHP
LFILE
HFILE
AH_LFILE
AH_HFILE
```

Cleaning SDC Request Data

This section explains how to clean the request extracted from SDC. We illustrate every step with R code below the text. Researchers using R can execute the provided code file by saving the SDC request data under the name "IPO data all.csv" in the R working directory.

Loading of the SDC Request

Using R, we read "IPO data all.csv" into the `ipo` variable. (For researchers using other programs, these data would likely be read as one dataset with multiple variables.) Because the variable names of the initial SDC request might include multiple spaces and string breaks, we rename all the variables for convenience:

```
require(data.table)
ipo <- fread("IPO data all.csv")
colnames(ipo) <- c("Filing_date", "Issue_date", "Issuer", "State", "Nation",
                   "IPO_Flag", "Original_IPO_Flag", "Offer_Price",
                   "Type", "Description", "REIT", "Unit", "ADR",
                   "Deal_number", "CEF",  "CUSIP", "CUSIP9",
                   "Proceeds", "VC", "Gross_spread",  "Mgr_codes",
                   "Tech_ind", "Low_Price", "High_Price",
                   "Low_Price_History", "High_Price_History")
```

Preview of loaded SDC request.

Table A.1. Preview of Loaded SDC Request

This table shows the first six lines of the SDC request for the observations with Issue_date after 2001-01-01. Variables `Filing_date`, `Low_Price`, `High_Price`, `Low_Price_History`, `High_Price_History` are missing for early sample years.

Filing_date	Issue_date	Issuer	State	Nation	IPO_Flag	Original_IPO_Flag	Offer_Price	Type	Description	REIT	Unit
6/21/2000	1/2/2001	Horizon Industries Ltd	Foreign	Canada	Yes	Yes	0.2	Common Shares	2,000,000.0 Common Shares		
9/19/2000	1/3/2001	Arsenal Capital Inc	Foreign	Canada	Yes	Yes	0.2	Common Shares	1,500,000.0 Common Shares		
7/28/2000	1/10/2001	UTI Energy Corp	Texas	United States	No		28.5	Common Shares	1,300,000.0 Common Shares		No
6/21/2000	1/10/2001	Westlinks Resources Ltd	Foreign	Canada	Yes		4.5	Common Shares	1,000,000.0 Common Shares		No
10/6/2000	1/11/2001	Alpha Technologies Group Inc	California	United States	No		7.25	Common Shares	270,946.0 Common Shares		No
7/10/2000	1/11/2001	Crown Castle International	Texas	United States	No		26.25	Common Shares	10,200,000.0 Common Shares		No

ADR	Deal_number	CEF	CUSIP	CUSIP9	Proceeds	VC	Gross_spread	Mgr_codes	Tech_ind	Low_Price	High_Price	Low_Price_History	High_Price_History
	1079234091	No	440432	440432102	0.3	No	10	BM	0				
	1643779091	No	04287Q	04287Q108	0.2	No	0.02	BM	0				
No	1173174002	No	903387	903387108	37.1	No	3.509	BM	0	22	22		
No	1082483002	No	960574	960574101	4.5	No	10	BM JL	0			4.5 4.5 4.5 4.5	5.5 5.5 5.5 5.5
No	1285528002	No	020781	020781100	2	No	na	BM	319	7.25	7.25		
No	1082049002	No	228227	228227104	267.8	No	2.857	BM SD GL	419	37.938	37.938		

Exclude Extra Observation

In this section, we provide a detailed explanation of how to construct a sample of companies that went public between 1973 and 2016 on the NYSE, Nasdaq and AMEX stock exchanges. IPOs with an offer price below $5, REITs, ADRs, units, and companies without CRSP prices within one week after the IPO are excluded.

Exclude Non-IPO Offerings

In the first step we drop observations where either `IPO_Flag` or `Original_IPO_Flag` variable equals "No". The variable `IPO_Flag` has two values: "Yes" and "No", and the variable `Original_IPO_Flag` is either blank or equals "No". Dropping observation with either of these two flags lowers the sample from 43,817 to 16,454 observations.

```
ipo <- ipo[IPO_Flag != "No" & Original_IPO_Flag != "No"]
```

Keep only Offerings of Common Shares

We want to limit the sample to offerings of common shares. To determine the type of the offering security we use the `Type` variable. Unfortunately, SDC's coding of this variable is inconsistent. Below, we tabulate the top ten values of the `Type` variable:

```
types <- data.table(table(ipo$Type))
ind <- order(types$N, decreasing = T)[1:10]
```

Table A.2 Top 10 Most Common Securities Types

Type of Security	N
Common Shares	11743
Ord/Common Shs.	1319
Class A Shares	683
Units	402
Shs Benficl Int	206
Ordinary Shares	185
Ltd Prtnr Int	74
MLP-Common Shs	68
Class B Shares	43
Beneficial Ints	19

Ideally, we would want to keep only common shares. However, as shown in Table A.2, SDC uses a variety of names to label common shares. For instance, among the top ten types of securities listed above, six different types represent common shares: "Common Shares", "Ord/Common Shares", "Class A Shares", "Ordinary Shares", "Class B Shares", "Class A Ord Shs". The full sample contains significantly more different labels for common shares.

Such variation in SDC codes of security types makes it easier to specify cases on the securities that we would want to drop rather than listing all the securities we would want to keep. We exclude offerings of the following types: "Units", "Ltd Prtnr Int", "MLP-Common Shs", "Shs Benficl Int", "Ltd Liab Int", "Stock Unit", "Trust Units", "Beneficial Ints". After this step the sample decreases from 16454 to 15107.

```
ex_types <- c("Units", "Ltd Prtnr Int", "MLP-Common Shs", "Shs Benficl Int",
              "Ltd Liab Int", "Stock Unit", "Trust Units", "Beneficial Ints")
ipo <- ipo[!Type %in% ex_types,]
```

Exclude REITs

In this step, we remove real estate investment trusts. If the offering is not a REIT than the value of `REIT` variable is blank. In other cases, the value of the variable is a two-digit code. The table below represent the frequency of each value of `REIT` variable (i.e., in 14,695 cases the variable is blank, in 193 cases it equals “EQ”, etc.).

```
table(ipo$REIT)

##
##          EQ    HY    MO    UN
## 14695   193     5    46   168
```

To exclude REITs we drop observations where the value of `REIT` variable is not missing. This decreases the sample from 15,107 observations to 14,695 observations.

```
ipo <- ipo[REIT == ""]
```

Exclude ADRs

We use the `ADR` variable to remove ADRs from the sample. This variable takes three values: blank, "No" and "Yes".

```
table(ipo$ADR)

##
##          No   Yes
##   467 13636   592
```

We keep the observations where `ADR` equals "No". This lowers the number of observations from 14,695 observations to 13,636 observations.

```
ipo <- ipo[ADR == "No"]
```

Exclude Closed-End Funds

We use the `CEF` variable to remove all Closed-End Funds from the sample. As in the pervious case, this variable takes three values: blank, "No" and "Yes".

```
table(ipo$CEF)

##
##          No   Yes
##     2 12627  1007
```

We keep the observations where `CEF` equals "No". This lowers the number of observations from 13,636 observations to 12,627 observations.

```
ipo <- ipo[CEF == "No"]
```

Exclude Units

In the next step, we remove Units from the sample. The variable `Unit` takes three values: blank, "No" and "Yes".

```
table(ipo$Unit)

##
##          No   Yes
##  1398 11187    42
```

Unfortunately, if we keep only observations where the variable `Unit` equals "No" we will drop some offerings of common shares (such as the Visa Inc IPO in 2009, which should be in the sample). Thus, we instead drop only observations where `Unit` variable equals to "Yes". This lowers the number of observations from 12,627 to 12,585. However, we note that the unit variable is not well-coded by SDC, and the only way to ensure a sample contains no units is through reading of prospectuses.

```
ipo <- ipo[Unit == "No" | Unit == ""]
```

Exclude Penny Stocks

Additionally, we remove IPOs with an offer price below $5.00. SDC uses commas in the offer prices as a thousand separator. We remove commas from the `Offer_Price` variable, convert it to numerical value, and drop observations with the offer price below $5.00. This limits our sample from 12,585 to 11,103 observations.

```
ipo[, Offer_Price := gsub(",","", Offer_Price)]
ipo[, Offer_Price := as.numeric(as.character(Offer_Price))]
ipo <- ipo[Offer_Price >= 5.0 & !is.na(Offer_Price)]
```

Matching SDC and CRSP database

To match the `ipo` dataset with the CRSP database we use the CRSP.DSF table (CRSP daily stock file). For matching and sample construction, we need the following CRSP variables: `date`, `PERMNO`, `NCUSIP`, `PRC`, `RET`, `SHROUT`, `EXCHCD`, `SHRCD`. In this script, we use "crsp.rds" file pre-loaded from WRDS website.

```
crsp <- readRDS("crsp.rds")
crsp <- as.data.table(crsp)
head(crsp)

##          date PERMNO   NCUSIP    PRC       RET SHROUT EXCHCD SHRCD
## 1: 1986-01-08  10000 68391610 -2.500 -0.024390   3680      3    10
## 2: 1986-01-09  10000 68391610 -2.500  0.000000   3680      3    10
## 3: 1986-01-10  10000 68391610 -2.500  0.000000   3680      3    10
## 4: 1986-01-13  10000 68391610 -2.625  0.050000   3680      3    10
## 5: 1986-01-14  10000 68391610 -2.750  0.047619   3680      3    10
## 6: 1986-01-15  10000 68391610 -2.875  0.045455   3680      3    10
```

In the CRSP sample, we keep only common shares (first digit of `SHRCD` equals 1) that were traded on NYSE, NASDAQ or AMEX. Additionally, we exclude rows where price information is missing (`PRC` variable equals to "NA").

```
crsp <- crsp[SHRCD %in% 10:19 & EXCHCD %in% 1:3]
crsp <- crsp[, PRC := as.numeric(as.character(PRC))]
crsp <- crsp[!is.na(PRC)]
```

We convert the `date` variable to the `R` date format and make sure that the `crsp` data.table is sorted by PERMNO and date.

```
require(lubridate)
crsp[, date := ymd(date)]
setkey(crsp, PERMNO, date)
```

We also create `NCUSIP6` variable, equal to the first six characters of the `NCUSIP` variable.

```
require(stringr)
crsp[, `:=` (NCUSIP6 = substr(NCUSIP, 1, 6))]
```

In the next step we prepare the `ipo` data.table for merging with `crsp` and for subsequent calculations. We convert `Filing_date` and `Issue_date` variables to the date format and create `Year` variable, equal to the year when a company completed its IPO.

```
ipo[, `:=`(Filing_date = mdy(Filing_date), Issue_date = mdy(Issue_date))]
ipo[, Year := year(Issue_date)]
```

We create the eight-digit CUSIP variable (`CUSIP8`) from `CUSIP9` and `CUSIP` variables extracted from SDC. If information in `CUSIP9` is not missing than `CUSIP8` equals to the first eight digits of it. In other cases, we create `CUSIP8` by adding "10" at the end of `CUSIP`.

```
ipo[, CUSIP8 := paste0(CUSIP, 10)]
ipo[str_length(CUSIP9) > 1, CUSIP8 := substr(CUSIP9, 1,8)]
print(ipo[1:5]$CUSIP8)

## [1] "65531210" "87233710" "00103210" "19579610" "89328710"
```

We match the ipo data to the crsp data in two ways, by both `CUSIP8` and by six-digit cusip. Specifically, we first match the `CUSIP8` variable in the `ipo` data to the `NCUSIP` variable from `crsp`. We use `crsp` data to find the Permno and the date when the security first appears on CRSP. These variables are called `Permno_ncusip` and `First_CRSP_date_ncusip`.

```
ipo$Permno_ncusip <- crsp$PERMNO[match(ipo$CUSIP8, crsp$NCUSIP)]
match <- match(ipo$Permno_ncusip, crsp$PERMNO)
ipo$First_CRSP_date_ncusip <- crsp$date[match]
```

Second, and similarly, we match the `ipo` and `crsp` data by six-digit CUSIPs. In this step, we again record the CRSP Permno and first CRSP date, which we name `Permno_ncusip6` and `First_CRSP_date_ncusip6`.

```
ipo$Permno_ncusip6 <- crsp$PERMNO[match(substr(ipo$CUSIP8,1,6), crsp$NCUSIP6)]
match <- match(ipo$Permno_ncusip6, crsp$PERMNO)
ipo$First_CRSP_date_ncusip6 <- crsp$date[match]
```

Now in our `ipo` data we have two Permno variables: `Permno_ncusip` and `Permno_ncusip6`. We want to find which variable we should use for the future calculations.

First, we look at the Permno obtained from the eight-character CUSIP match (i.e. `Permno_ncusip`). We assign the value of this variable to the `Permno` variable if a company with this Permno appears on CRSP no earlier than one day before and no later than seven days after the issue date.

```
ipo[, Permno := -999] # -999 used as a "dummy" value for non-match results
ipo[, dif := as.numeric(as.character(First_CRSP_date_ncusip - Issue_date))]
ipo[(dif >= -1 & dif <= 7) & (Permno == - 999) & !is.na(dif), Permno := Permno_ncusip]
```

If we did not assign any value to `Permno` variable (i.e. it still equals to -999) we go to the `Permno_ncusip6` variable. Here again, we replace value `Permno` with `Permno_ncusip6` is a company with such Permno appeared on CRSP no earlier than one day before and no later than seven days after the issue date.

```
ipo[, dif := as.numeric(as.character(First_CRSP_date_ncusip6 - Issue_date))]
ipo[ (dif >= -1 & dif <= 7) & (Permno == - 999) & !is.na(dif), Permno := Permno_ncusip6]
```

Exclude Observations w/o CRSP Record

After the above step, we drop companies for which we were not able to find a proper match. Additionally, we drop observations with duplicated values of `Permno` (usually that happens when a company offers different classes of securities at the same time). These criteria lower the number of observations from 11,103 to 8,995 observations.

```
ipo <- ipo[Permno != -999]
ipo <- ipo[!duplicated(ipo$Permno)]
```

Variable Construction

Initial Returns

For the calculation of initial returns, we use `Permno` variable (as calculated in prior step) to match the `ipo` and `crsp` datasets. We record the first closing price on CRSP into `Close_Price` (which we store in the `ipo` dataset). Sometimes CRSP data includes negative prices. This occurs in cases when the price record was not available and CRSP replaced price with an average of bid and ask quotes. To adjust for this caveat in the data we take the absolute value of the closing price. Initial return (`IR` variable) is calculated as the return between the first closing price and the offer price.

```
m <- match(ipo$Permno, crsp$PERMNO)
ipo$Close_Price <- crsp$PRC[m]
ipo[, Close_Price := abs(Close_Price)]
ipo[, IR := Close_Price/Offer_Price - 1]
```

Proceeds

Nominal Proceeds

Nominal proceeds come directly from the SDC data. You just need to remove the comma thousand separators and convert the data to a numeric format.

```
ipo[, Proceeds := as.numeric(as.character(gsub(",","", Proceeds)))]
```

Real Proceeds in 2016 dollars.

Estimation of nominal proceeds requires data on GDP price deflator. GDP price deflator information is available on https://fred.stlouisfed.org/series/GDPDEF. We download these data into "GDPDEF.csv" file in the working directory. This file includes quarterly values of GDP price deflator. For our calculation, we use the value of `GDPDEF` for the first quarter of each year. The variable `factor_2016` shows how much more valuable one dollar was in a given year, relative to the value in 2016.

```
# factor 2016 calculation
gdp <- fread("GDPDEF.csv")
gdp[, year := year(DATE)]
gdp <- gdp[!duplicated(year)] # use first quater only
gdp[, base := GDPDEF[year == 2016]]
gdp[, factor_2016 := base/GDPDEF]
head(gdp)

##          DATE GDPDEF year   base factor_2016
## 1: 1947-01-01 12.566 1947 110.63    8.803915
## 2: 1948-01-01 13.379 1948 110.63    8.268929
## 3: 1949-01-01 13.717 1949 110.63    8.065175
## 4: 1950-01-01 13.490 1950 110.63    8.200890
## 5: 1951-01-01 14.596 1951 110.63    7.579474
## 6: 1952-01-01 14.863 1952 110.63    7.443316
```

Matching `factor_2016` to the `ipo` database

```
match <- match(ipo$Year, gdp$year)
ipo$factor_2016 <- gdp$factor_2016[match]
ipo$Proceeds_2016 <- ipo$Proceeds*ipo$factor_2016
```

Initial Price Range

SDC Platinum provides several variables for the initial price range. Some variables provide the initial low/high price, some provide the price update history, and some provide the last low/high filing price. More information on all SDC variables is provided on http://mergers.thomsonib.com/td/DealSearch/help/nidef.htm.

Here is the brief overview of available price range variables on SDC:

1. **LFILE** variable has information on the **first** value of the low price of the filing price range. SDC keeps track of this variable starting from 01-01-1983.
2. **LFILE2** variable has information on the **last** value of the low price of the filing price range. SDC keeps track of this variable starting from 01-01-1983.
3. **C_LFILE** variable also has information on the **last** value of the low price of the filing price range. SDC keeps track of this variable starting from 01-01-1983. In our sample this variable seems to be identical to **LFILE2**.
4. **AH_LFILE** variable keeps the history of price update starting with the first **amendment** of the prospectus. So if the company files IPO prospectuses in the following sequence: S-1, S-1/A, S-1/A, S-1/A, the variable would have information from the last three filings (S-1/As) but not from the first one. SDC keeps track of this variable starting from 01-01-1996.
5. **ABOVEBELOW** variable has information on whether the issue is priced below/within/above the **last** price range. SDC keeps track of this variable starting from 01-01-1983.

Similarly, SDC has variables **HFILE**, **HFILE2**, **C_HFILE**, **AH_HFILE** that have similar information on the high end of the price range.

Importantly, the variables **LFILE2**, **C_LFILE** and **ABOVEBELOW** cannot be used to construct initial price range as they have information regarding the last (i.e., amended) price range.

For the construction of the lower bound of the initial price range, we use two variables: **LFILE** and **AH_LFILE**. Most information is taken from **LFILE** variable and if it is missing, we use **AH_FILE** information.

In sum, the roadmap for the construction of the initial price range is the following:

1. Starting with 01-01-1983 we use the **LFILE** SDC variable for the lower end of the price range (we call this SDC variable `Low_Price` in the loaded `ipo` dataset). We repeat this step for **HFILE**
2. If the value of the previously defined variable is missing and the offer date is after 01-01-1996 we use the first value of the **AH_LFILE** variable (we call this SDC variable `Low_Price_History` in the loaded `ipo` dataset). We repeat this step for **AH_HFILE**
3. Sometimes SDC reverses the low and high ends of the price range. To adjust for such mistakes we create variables `min_range` which equals the minimum of the low and high ends, and `max_range` which equals the maximum.
4. If neither the `min_range` nor `max_range` variables have missing values we use them to determine whether the issue was price below/within or above the initial price range.

The code below incorporates this described roadmap. SDC separates values in the price update history variable with the end of the list character (i.e. `\n` code) and function `first_price()` extracts the first non-missing value from `Low_Price_History` and `High_Price_History` variables.

```r
# Estimation of Initial Price Range

# function that helps to format to numeric
n <- function(x) return(as.numeric(as.character(x)))
ipo[, Issue_priced_range := "NA"]

# Step 1.
# Use LFILE/HFILE for price range first
ipo[Year >= 1983 & Low_Price!= "",constructed_low  := n(gsub(",", "", Low_Price))]
ipo[Year >= 1983 & High_Price!= "",constructed_high := n(gsub(",", "", High_Price))]


# info from AH_LFILE/AH_HFILE (ammendment price history list) variables
first_price <- function(a){
  prices <- strsplit(as.character(a), split = "\n")
  prices <- lapply(prices, function(x) return(n(x)))
  prices <- lapply(prices, function(x) return(x[!is.na(x)]))
  first_element_or_na <- function(x){
    if(length(x) >= 1) return(x[1])
    if(length(x) == 0) return(NA)
  }
  prices <- lapply(prices, first_element_or_na)
  return(unlist(prices))
}

# Step 2.
# After 1996 fill the missing information from AH_LFILE/AH_HFILE variables
a <- ipo[Year >= 1996 & is.na(constructed_low)]$Low_Price_History
ipo[Year >= 1996 & is.na(constructed_low)]$constructed_low <- first_price(a)

b <- ipo[Year >= 1996 & is.na(constructed_high)]$High_Price_History
ipo[Year >= 1996 & is.na(constructed_high)]$constructed_high <- first_price(b)


# Step 3.
# Adjusted to SDC mixing high/low price
min_range <- 0.5*(ipo$constructed_low + ipo$constructed_high) -
         0.5*abs(ipo$constructed_low - ipo$constructed_high)
max_range <- 0.5*(ipo$constructed_low + ipo$constructed_high) +
         0.5*abs(ipo$constructed_low - ipo$constructed_high)

# Step 4.
# Create Issue_priced_range variables
not_missing <- (!is.na(ipo$constructed_low)) & (!is.na(ipo$constructed_high))
ipo[not_missing & Offer_Price < min_range,
               Issue_priced_range := "Below range"]
ipo[not_missing & Offer_Price >= min_range & Offer_Price <= max_range,
               Issue_priced_range := "Within range"]
ipo[not_missing & Offer_Price > max_range,
               Issue_priced_range := "Above range"]
```

Syndicate Members

We use the `Mgr_codes` variable to determine the number of lead managers, co-managers, and other syndicate members. SDC has six codes for manager roles: "BM" (book manager), "JB" (joint book), "JL" (joint lead), "CM" (co-manager), "SD" (syndicate member) and "GL"(general lead). Managers that are coded as "GL" are usually also coded as either "BM" or "JB". Therefore, to avoid double counting we ignore "GL" code.

Codes "BM", "JB" and "JL" all correspond to the lead underwriters. Code "CM" indicates a co-manager role and "SD" indicates other syndicate members. We count the number of lead managers, co-managers and other syndicate members by the number of occurrences of the listed codes in the `Mgr_codes` variable.

```
ipo[, n_lead := str_count(Mgr_codes,"BM") +
                str_count(Mgr_codes,"JB") +
                str_count(Mgr_codes,"JL"), by = Deal_number]
ipo[, n_co := str_count(Mgr_codes,"CM"), by = Deal_number]
ipo[, n_syn := str_count(Mgr_codes,"SD"), by = Deal_number]
```

Registration Period

Length of the registration period is defined as a difference (in days) between the issue date and the filing date. We convert both variables to `R` date format, as previously in this document. In cases where the variables are not formatted, the function `ymd()` from the `lubridate` package can be used.

```
ipo[, rp := as.numeric(as.character(Issue_date - Filing_date))]
```

Matching SDC and SEC EDGAR

SDC Platinum originally extracts information from SEC filings. For companies with IPOs after 1997 these filings are publically available on the SEC EDGAR website. SDC provides information that can be used to match IPO data with the SEC EDGAR database. Such a link between the SDC and SEC databases is highly useful in empirical research. First, it allows researchers to understand the origins of the data and to double check the information. Second, one can use SEC to extract additional information from different types of filings (such as 10-K, 8-K, S-4, UPLOAD, etc).

SEC EDGAR IPO Filings

SDC provides the link to SEC EDGAR filings in "SEC File Form (Number)" variable. You can extract this information for all observations in an additional SDC request. In this file, we provide an example of the "Square Inc" IPO to illustrate the process.

According to SDC data, Square Inc filed for an IPO on October 14, 2015 and went public on November 18, 2015. The value of "SEC File Form (Number)" (coded by SDC as `FILEN`) is 207411. If you put the prefix "https://www.sec.gov/cgi-bin/browse-edgar?action=getcompany&filenum=333-" to this number it leads to the sequence of Square Inc IPO filings on the SEC EDGAR website. All IPOs use the same prefix.

Here is a current version of https://www.sec.gov/cgi-bin/browse-edgar?action=getcompany&filenum=333-207411 page preview. The first prospectus (form S-1) in the sequence was submitted on October 14, 2015, which corresponds to the SDC filing date. According to this page, the company went public on November 18, 2015 (date of EFFECT form). This information is also consistent with SDC data.

Square, Inc. CIK#: 0001512673 (see all company filings)

SIC: 7372 - SERVICES-PREPACKAGED SOFTWARE
State location: CA | State of Inc.: **DE** | Fiscal Year End: 1231
(Assistant Director Office: 3)
Get **insider transactions** for this **issuer**.
File Number: **333-207411**

Business Address
1455 MARKET STREET
SUITE 600
SAN FRANCISCO CA 94103
415-375-3176

Mailing Address
1455 MARKET STREET
SUITE 600
SAN FRANCISCO CA 94103

Filter Results: Filing Type: Prior to: (YYYYMMDD) Ownership? include exclude only Limit Results Per Page 40 Entries Search Show All

Items 1 - 11 RSS Feed

Filings	Format	Description	Filed/Effective	File/Film Number
CT ORDER	Documents	Confidential treatment order Acc-no: 9999999997-17-002097 (33 Act) Size: 10 KB	2017-03-09	333-207411 17678255
424B4	Documents	Prospectus [Rule 424(b)(4)] Acc-no: 0001193125-15-382249 (33 Act) Size: 12 MB	2015-11-19	333-207411 151244368
CT ORDER	Documents	Confidential treatment order Acc-no: 9999999997-15-014749 (33 Act) Size: 10 KB	2015-11-19	333-207411 151242745
EFFECT	Documents	Notice of Effectiveness Acc-no: 9999999995-15-003082 (33 Act) Size: 1 KB	2015-11-18 16:00:00	333-207411 151242288
S-1/A	Documents	**[Amend]** General form for registration of securities under the Securities Act of 1933 Acc-no: 0001193125-15-378578 (33 Act) Size: 8 MB	2015-11-16	333-207411 151236394
FWP	Documents	Filing under Securities Act Rules 163/433 of free writing prospectuses Acc-no: 0001193125-15-371917 (34 Act) Size: 7 KB	2015-11-09	333-207411 151215585
FWP	Documents	Filing under Securities Act Rules 163/433 of free writing prospectuses Acc-no: 0001193125-15-371912 (34 Act) Size: 9 KB	2015-11-09	333-207411 151215555
S-1/A	Documents	**[Amend]** General form for registration of securities under the Securities Act of 1933 Acc-no: 0001193125-15-370613 (33 Act) Size: 273 KB	2015-11-09	333-207411 151213773
S-1/A	Documents	**[Amend] [Cover]**General form for registration of securities under the Securities Act of 1933 Acc-no: 0001193125-15-369092 (33 Act) Size: 10 MB	2015-11-06	333-207411 151202377
S-1/A	Documents	**[Amend]** General form for registration of securities under the Securities Act of 1933 Acc-no: 0001193125-15-352937 (33 Act) Size: 8 MB	2015-10-26	333-207411 151174531
S-1	Documents	General form for registration of securities under the Securities Act of 1933 Acc-no: 0001193125-15-343733 (33 Act) Size: 18 MB	2015-10-14	333-207411 151158297

IPO CIK Codes

SEC EDGAR uses CIK codes as identifiers of companies and investors. Researchers can use the previous webpage to automatically extract the CIK code of the IPO firm.

For this purpose, define the `sec_filings` variable as the content (i.e., html code) of the page with URL address of (prefix + "SEC File Form (Number)".

```
filen <- "207411"
prefix <- "https://www.sec.gov/cgi-bin/browse-edgar?action=getcompany&filenum=333-"

link <- paste0(prefix, filen)
sec_filings <- readLines(url(link))
```

Then search for "Central Index Key" phrase inside this variable:

```
ind <- grep("Central Index Key", sec_filings)
line <- sec_filings[ind]
print(line)

## [1] "         <span class=\"companyName\">Square, Inc. <acronym title=\"Central Index Ke
y\">CIK</acronym>#: <a href=\"/cgi-bin/browse-edgar?action=getcompany&amp;CIK=0001512673&am
p;owner=include&amp;count=40\">0001512673 (see all company filings)</a></span>"
```

The variable `line` contains the contents of the line of the `sec_filigs` html with the information about the CIK code. To extract the CIK code just remove everything before ";CIK=" and after "&owner=". In the code below, `.*` is used to match everything before and after the cut offs.

```
CIK <- gsub(".*;CIK=", "", line)
CIK <- gsub("&amp;owner=.*", "", CIK)
print(CIK)

## [1] "0001512673"
```

The output of the code above for Square Inc's CIK code is "0001512673". The CIK code can be used to access Square filings on the SEC EDGAR website. SEC master files provide the links to all their filings: https://www.sec.gov/Archives/edgar/full-index/.

Corrections to SDC Data

We have used the procedures described above to establish a link between SDC Platinum and SEC EDGAR for all companies that went public between 1996 and 2016. In total, we have 3,324 observations in this time period. For each of these IPOs, we verified whether the SEC File Number, Issue Date and Filing Date were correctly reported by SDC.

Manually going though these 3,324 observations, we identify 31 cases where the SEC File Number was either incorrect or missing in SDC. Table A.3 lists these cases together with the hand-corrected information.

The date of IPO was almost always correct at SDC. At least, there were no observations for which the Issue date reported by SDC was more than two days apart from SEC data.

There were 41 cases where the SDC filing date was either incorrect or missing. We have listed these observations together with correct values in Table A.4

Table A.3 Correction of SEC File Number variable in SDC Platinum for IPOs between 1996 and 2016.

We hand checked "SEC File Number" variable is SDC Platinum database for 3,324 IPOs between 1996 and 2016. There were 31 cases in which information reported by SDC was either missing or incorrect. The first three columns of this table list Issue_date, Issuer and SDC Deal Number for these observations. The fourth column provides information on SEC File Number as reported by SDC for the cases in which this field is either missing or incorrect. The fifth column provides hand-corrected values for SDC File Number. Adding the prefix https://www.sec.gov/cgi-bin/browse-edgar?action=getcompany&filenum=333- in front of the last column leads to the firm's IPO directory on EDGAR.

Issue_date (Reported by SDC)	**Issuer** (Reported by SDC)	**Deal_number** (Reported by SDC)	**SEC File Number** (Reported by SDC)	**CORRECTED: SEC File Number** (Hand-corrected)
6/10/96	McLeod Inc	558747009		3112
10/31/96	Triangle Pharmaceuticals Inc	601647002	1793	11793
7/8/97	Telegroup Inc	1096046009		25065
10/6/97	SCM Microsystems Inc	701039009		29073
10/9/97	Stoneridge Inc	1095835009		33285
6/8/99	The Pantry Inc	1094419009		74221
6/30/99	Clarent Corp	1094301009		76051
7/27/99	Focal Communications Corp	1094177009		77995
8/10/99	Red Hat Inc	1094135009		80051
9/21/99	Broadbase Software Inc	1094069009		82251
10/5/99	Silicon Image Inc	929829009		83665
11/23/99	OpenTV Corp	1093776009		89609
2/3/00	Alamosa PCS Holdings Inc	1093596009		89995
3/2/00	UTStarcom Inc	1093434009		93069
3/14/00	Entropin Inc	956451002	93565	11308
6/29/00	Mobility Electronics Inc	1092914009		30264
7/27/00	Lexent Inc	1030170009		30660
8/3/00	Bruker Daltonics Inc	1032981009		34820
9/28/00	Docent Inc	1092587009		34546
9/28/00	Simple Technology Inc	1092584009		32478
10/16/00	EndWave Corp	1092531009		41302
10/17/00	Monsanto Co	1056838009		36956
4/30/01	Reliant Resources Inc	1179524009		48038
7/12/04	Domino's Pizza Inc	1514070002	0	114442-01
3/18/08	Visa Inc	1923350009		147296
3/18/10	Golden Minerals Co	2118265009		162486
3/19/13	Tetraphase Pharmaceuticals Inc	2495337009		186574
9/26/13	Pattern Energy Group Inc	2557099009		190538
6/18/14	Cellular Biomedicine Group	2648690009		210337
8/10/16	Medpace Inc	2977673002	212810	212236
9/22/16	Apptio Inc	2397201002	0	213334

Table A.4 Correction of the Filings Date in SDC Platinum for IPOs between 1996 and 2016.

We hand checked "Filing Date" variable in SDC Platinum database for 3,324 IPOs between 1996 and 2016 using SEC EDGAR. There were 44 cases in which the Filing Date from SDC was more than two days apart from the first filing date on SEC EDGAR. The first three columns of the table provide Issue_date, Issuer and SDC deal number of these observations. The fourth column provides the filing date as reported by SDC Platinum. The last column provides the corrected filing date based on SEC EDGAR information.

Issue_date (Reported by SDC)	**Issuer** (Reported by SDC)	**Deal_number** (Reported by SDC)	**Filing_Date** (Reported by SDC)	**CORRECTED: Filing_Date** (Hand-corrected)
3/10/97	Pluma Inc	648003002	11/25/87	12/24/96
8/27/97	Authentic Specialty Foods Inc	691349002	5/27/97	6/25/97
6/26/98	Cumulus Media Inc	769812002	7/28/92	3/30/98
1/28/99	Tut Systems Inc	847312002		7/31/98
4/5/99	PLX Technology Inc	869198002	3/1/99	2/4/99
5/5/99	NorthPoint Communications Grp	878288002	4/19/99	2/26/99
6/8/99	The Pantry Inc	1094419009	6/8/99	3/11/99
6/9/99	Skechers USA Inc	891399002	6/8/95	7/29/98
7/27/99	Focal Communications Corp	1094177009	7/27/99	5/7/99
1/28/00	interWAVE Communications	967001002	12/29/99	12/17/99
2/15/00	VarsityBooks.com Inc	976518002	11/8/99	10/14/99
3/10/00	OTG Software Inc	984681002	2/15/00	12/23/99
8/8/00	ChipPAC Inc	1034511002	7/14/00	6/16/00
3/6/01	Seattle Genetics Inc	1162559002	2/23/01	11/20/00
10/11/01	TheraSense Inc	1228444002	8/13/01	7/3/01
8/14/02	Crescent Finl Corp,Cary,NC	1321633002		5/3/02
8/6/03	CapitalSource Inc	1414465002	7/15/03	6/12/03
10/7/03	Acusphere Inc	1410226002	8/6/03	7/1/03
3/24/04	Ultra Clean Holdings Inc	1467355002	2/17/04	1/14/04
10/4/04	Texas Roadhouse Inc	1525081002	10/4/04	5/7/04
10/6/04	IntraLase Corp	1531022002	10/5/04	5/28/04
10/20/04	Tower Group Inc	1525245002	9/15/04	5/7/04
10/27/04	Calamos Asset Management Inc	1591271002	10/12/04	8/2/04
12/8/04	Foundation Coal Holdings Inc	1595750002	11/15/04	8/20/04
12/8/04	Symmetry Medical Inc	1531027002	11/17/04	5/28/04
12/13/04	Bluelinx Holdings Inc	1597969002	12/10/04	9/2/04
12/14/04	Cascade Microtech Inc	1503175002	4/13/04	3/3/04
12/14/04	Great Wolf Resorts Inc	1593683002	12/7/04	8/12/04
12/15/04	Advance America Cash Advance	1594026002	12/13/04	8/13/04
12/16/04	Arbinet-thexchange Inc	1586148002	8/25/04	7/9/04
12/16/04	Warren Resources Inc	1596347002	12/2/04	8/25/04
1/25/05	Majesco Holdings Inc	1631181002	11/12/04	10/29/04
2/9/05	Nasdaq Stock Market Inc	1624412002	1/25/05	12/14/04
4/12/05	PRB Gas Transportation Inc	1635318002	2/28/05	11/1/04
5/16/05	Xerium Technologies Inc	1582496002	6/10/04	4/22/04
8/23/05	Ready Mix Inc	1639349002	8/19/05	2/11/05
12/14/05	Somaxon Pharmaceuticals Inc	1699924002	11/9/05	10/7/05
5/14/07	Continental Resources Inc	1741066002	5/1/07	3/7/06
12/13/07	MEMSIC Inc	1910258002	12/3/07	9/28/07
12/19/07	NetSuite Inc	1884200002	12/5/07	7/2/07
1/22/08	Meridian Interstate Bancorp	2032177002		9/28/07
1/24/12	Guidewire Software Inc	2335107002	8/5/11	9/2/11
9/17/13	Benefitfocus Inc	2506403002	9/4/13	8/14/13
5/13/14	ServisFirst Bancshares Inc	2618976002	3/17/14	1/17/14