```
# nrow(final_link) # 182,615

size_btm_leverage_IR_instown_analyst = size_btm_leverage_IR_instown %>% left_join(final_link, by = c("LPERMNO", "datadate")) %>%
                                       mutate(NUMEST = ifelse(is.na(NUMEST), 0, NUMEST)) %>%
                                       mutate(NUMEST = log(1 + NUMEST)) %>%
                                       rename("analysts" = "NUMEST")

# nrow(size_btm_leverage_IR_instown_analyst) # 220,407
```

7. M&A: indicator for recent material M&A

```
sdc = read_excel(paste0(path, "1_data_processing/data/control_variables/raw/SDC_MA.xlsx"))

# read in fundq
crsp_compustat_fundq = read.csv(paste0(path, "1_data_processing/data/sample_selection/wrds/raw/crsp_compustat_fundq_20250617.csv"))

# select datadate from 2010 to latest
crsp_compustat_fundq = crsp_compustat_fundq %>% mutate(datadate = as.Date(datadate)) %>%
                       filter(datadate >= "2008-01-01", datadate <= "2024-12-31")

# filter LINKPRIM = P or C
crsp_compustat_fundq = crsp_compustat_fundq %>% filter(LINKPRIM %in% c("P", "C"))

# filter datafqtr != ""
crsp_compustat_fundq = crsp_compustat_fundq %>% filter(datafqtr != "")

# filter rdq is not NA
crsp_compustat_fundq = crsp_compustat_fundq %>% filter(rdq != "")

# add calender year of fiscal year end
crsp_compustat_fundq = crsp_compustat_fundq %>% mutate(year = year(datadate)) %>%
                       mutate(GVKEY = str_pad(GVKEY, 6, pad = "0")) %>%
                       mutate(rdq = as.Date(rdq))

# merge in shrcd
crsp_compustat_fundq = crsp_compustat_fundq %>% inner_join(mse %>% dplyr::select(permno, namedt, nameendt, shrcd), by = c("LPERMNO" = "permno")) %>%
                                                filter(datadate >= namedt, datadate <= nameendt)
```

```
## Warning in inner_join(., mse %>% dplyr::select(permno, namedt, nameendt, : Detected an unexpected many-to-many relationship between `x` and `y`.
## ℹ Row 1 of `x` matches multiple rows in `y`.
## ℹ Row 30298 of `y` matches multiple rows in `x`.
## ℹ If a many-to-many relationship is expected, set `relationship =
##   "many-to-many"` to silence this warning.
```

```
crsp_compustat_fundq = crsp_compustat_fundq %>% mutate(id = paste0(GVKEY, "_", datadate))

crsp_compustat_fundq = crsp_compustat_fundq %>%
                       group_by(id) %>% arrange(desc(datacqtr), .by_group = TRUE) %>% slice(1) %>% ungroup() %>%
                       dplyr::select(GVKEY, datadate, id, fqtr)

crsp_compustat_fundq = crsp_compustat_fundq %>%
                       group_by(GVKEY) %>% arrange(datadate, .by_group = TRUE) %>% mutate(datadatelag = lag(datadate)) %>% ungroup() %>%
                       mutate(datadatelag = ifelse(is.na(datadatelag), add_with_rollback(datadate, months(-3)), datadatelag)) %>%
                       mutate(datadatelag = as.Date(datadatelag, origin = "1970-01-01")) %>% dplyr::select(-fqtr)

# crsp_compustat_fundq %>% mutate(dif = datadate - datadatelag) %>% filter(dif > 92) # 3,915

crsp_compustat_fundq = crsp_compustat_fundq %>% mutate(dif = datadate - datadatelag) %>%
                       mutate(datadatelag = ifelse(dif > 92, datadate - 92, datadatelag)) %>%
                       mutate(datadatelag = as.Date(datadatelag, origin = "1970-01-01")) %>% dplyr::select(-dif)

crsp_combined = crsp_compustat_fundq %>% filter(datadate >= "2010-01-01")

# nrow(crsp_combined) # 288,273
```

```
sdc_mse_acquiror = mse %>% mutate(cusip6 = str_sub(ncusip, 1, 6)) %>% inner_join(sdc, by = c("cusip6" = "Acquiror 6-digit CUSIP")) %>%
                           filter(`Date Announced` >= namedt, `Date Announced` <= nameendt) %>%
                           group_by(`SDC Deal No`) %>% mutate(n = n()) %>% filter(n == 1 | (n > 1 & ticker == `Acquiror Primary Ticker Symbol`)) %>%
                           mutate(n = n()) %>% filter(n == 1) %>% ungroup() %>% dplyr::select(permno, `SDC Deal No`, `Date Announced`)
```

```
## Warning in inner_join(., sdc, by = c(cusip6 = "Acquiror 6-digit CUSIP")): Detected an unexpected many-to-many relationship between `x` and `y`.
## ℹ Row 2 of `x` matches multiple rows in `y`.
## ℹ Row 43891 of `y` matches multiple rows in `x`.
## ℹ If a many-to-many relationship is expected, set `relationship =
##   "many-to-many"` to silence this warning.
```

```
sdc_mse_target = mse %>% mutate(cusip6 = str_sub(ncusip, 1, 6)) %>% inner_join(sdc, by = c("cusip6" = "Target 6-digit CUSIP")) %>%
                         filter(`Date Announced` >= namedt, `Date Announced` <= nameendt) %>%
                         group_by(`SDC Deal No`) %>% mutate(n = n()) %>% filter(n == 1 | (n > 1 & ticker == `Target Primary Ticker Symbol`)) %>%
                         mutate(n = n()) %>% filter(n == 1) %>% ungroup() %>% dplyr::select(permno, `SDC Deal No`, `Date Announced`)
```

```
## Warning in inner_join(., sdc, by = c(cusip6 = "Target 6-digit CUSIP")): Detected an unexpected many-to-many relationship between `x` and `y`.
## ℹ Row 3 of `x` matches multiple rows in `y`.
## ℹ Row 48911 of `y` matches multiple rows in `x`.
## ℹ If a many-to-many relationship is expected, set `relationship =
##   "many-to-many"` to silence this warning.
```

```
acquiror = size_btm_leverage_IR_instown_analyst %>%
           inner_join(crsp_combined %>% dplyr::select(GVKEY, datadate, datadatelag), by = c("GVKEY", "datadate")) %>%
           left_join(sdc_mse_acquiror, by = c("LPERMNO" = "permno")) %>%
           filter(`Date Announced` > datadatelag, `Date Announced` <= datadate) %>%
           dplyr::select(GVKEY, datadate) %>% distinct()
```

```
## Warning in left_join(., sdc_mse_acquiror, by = c(LPERMNO = "permno")): Detected an unexpected many-to-many relationship between `x` and `y`.
## ℹ Row 1 of `x` matches multiple rows in `y`.
## ℹ Row 23561 of `y` matches multiple rows in `x`.
## ℹ If a many-to-many relationship is expected, set `relationship =
##   "many-to-many"` to silence this warning.
```

```
target = size_btm_leverage_IR_instown_analyst %>%
         inner_join(crsp_combined %>% dplyr::select(GVKEY, datadate, datadatelag), by = c("GVKEY", "datadate")) %>%
         left_join(sdc_mse_target, by = c("LPERMNO" = "permno")) %>%
         filter(`Date Announced` > datadatelag, `Date Announced` <= datadate) %>%
         dplyr::select(GVKEY, datadate) %>% distinct()
```

```
## Warning in left_join(., sdc_mse_target, by = c(LPERMNO = "permno")): Detected an unexpected many-to-many relationship between `x` and `y`.
## ℹ Row 64 of `x` matches multiple rows in `y`.
## ℹ Row 4457 of `y` matches multiple rows in `x`.
## ℹ If a many-to-many relationship is expected, set `relationship =
##   "many-to-many"` to silence this warning.
```

```
size_btm_leverage_IR_instown_analyst_ma = size_btm_leverage_IR_instown_analyst %>%
                                          left_join(acquiror %>% mutate(acquiror = 1), by = c("GVKEY", "datadate")) %>%
                                          mutate(acquiror = ifelse(is.na(acquiror), 0, 1)) %>%
                                          left_join(target %>% mutate(target = 1), by = c("GVKEY", "datadate")) %>%
                                          mutate(target = ifelse(is.na(target), 0, 1)) %>%
                                          mutate(MA = ifelse(acquiror + target == 0, 0, 1)) %>%
                                          dplyr::select(GVKEY:analysts, MA)

# nrow(size_btm_leverage_IR_instown_analyst_ma) # 220,407
```

8. segments: number of operating segments

```
comp_segments = readRDS(paste0(path, "1_data_processing/data/control_variables/wrds/RDS/compustat_segments_20250707.RDS"))
comp_segments_opseg = comp_segments %>% mutate(id = paste0(gvkey, "_", datadate)) %>%
                      dplyr::select(id, OPSEG) %>%
                      mutate(OPSEG = ifelse(is.na(OPSEG), 0, OPSEG))


size_btm_leverage_IR_instown_analyst_ma_segments = size_btm_leverage_IR_instown_analyst_ma %>% mutate(id = paste0(GVKEY, "_", datadate)) %>%
                                                   left_join(comp_segments_opseg, by = c("id")) %>%
                                                   group_by(GVKEY, fyear) %>% arrange(OPSEG, .by_group = TRUE) %>%
                                                   fill(OPSEG) %>% ungroup() %>% mutate(OPSEG = ifelse(is.na(OPSEG), 0, OPSEG)) %>%
                                                   dplyr::select(id, GVKEY:MA, OPSEG) %>%
                                                   rename("segments" = "OPSEG")

# nrow(size_btm_leverage_IR_instown_analyst_ma_segments) # 220,407
```

9. abnormal returns

```
ret_fundq = readRDS(paste0(path, "1_data_processing/data/control_variables/wrds/RDS/abnormal_returns_fundq_by_decile_20250707.RDS"))

size_btm_leverage_IR_instown_analyst_ma_segments_abr = size_btm_leverage_IR_instown_analyst_ma_segments %>%
                                                       left_join(ret_fundq %>% rename("abnormal_return_by_decile" = "abnormal_return"), by = c("id"))

# nrow(size_btm_leverage_IR_instown_analyst_ma_segments_abr) # 220,407
```

10. audit fees

```
audit_fee_combined = readRDS(paste0(path, "1_data_processing/data/control_variables/wrds/RDS/audit_fees_clean_20250707.RDS"))

size_btm_leverage_IR_instown_analyst_ma_segments_abr_audit = size_btm_leverage_IR_instown_analyst_ma_segments_abr %>% left_join(audit_fee_combined, by = c("id"))

# nrow(size_btm_leverage_IR_instown_analyst_ma_segments_abr_audit) # 220,407
```

save

```
saveRDS(size_btm_leverage_IR_instown_analyst_ma_segments_abr_audit, paste0(path, "1_data_processing/data/control_variables/control_variables_pr_cc_20250707.RDS"))
```

```
.end_time <- Sys.time()
cat("===== END TIME =====\n")
```

```
## ===== END TIME =====
```

```
print(.end_time)
```

```
## [1] "2026-02-02 21:35:25 PST"
```

```
cat("Elapsed:", difftime(.end_time, .start_time, units = "secs"), "seconds\n")
```

```
## Elapsed: 686.624 seconds
```