# Information Leakage Prior to SEC Form Filings—Evidence from TAQ Millisecond Data

Keywords: Information Leakage, Private Information, Big Data, Informed Trading, Millisecond Trades

Steven Wei Ho, Department of Economics, Columbia University
Mingrui Zhang, Foster School of Business, University of Washington

## Introduction

Is there information leakage prior to corporate SEC filings? Existing studies have investigated the issue of informed trading and information leakage prior to specific events, such as mergers and share repurchases, and these studies generally indicate presence of abnormal or informed trading activities prior to announcements. However, these studies are all utilize daily data and focus on a horizon on the order of days or months prior to the announcement, we on the other hand, utilize high frequency data and investigate information leakage on the minutes range. Also, we aim to take advantage of big data and conduct our analysis on a comprehensive sample of all filings from EDGAR, including Form 10-Q, 10-K, 8-K, 4, Schedule 13D, and Schedule 13G among others, without restricting our attention on one particular event. We investigate information leakage as proxied by equity price movements by looking at the all-inclusive sample of SEC form filings in the years 1994-2017. Within the event sample that are made during normal trading hours, we find strong evidence that in the 30-minute intervals around the Edgar acceptance timestamp, the form filings in which stock prices increased the most before the timestamp are also those in which stock prices increase the most after the timestamp.

### Announcement List & Daily TAQ

**> From 1994-2017, 11,560,676 event observations and 30,5226 unique company identifiers.**

**> From September 2003 - present intraday transactions data (trades and quotes) for all securities listed on the New York Stock Exchange (NYSE), American Stock Exchange (AMEX), as well as Nasdaq National Market System (NMS) and SmallCap issues.**

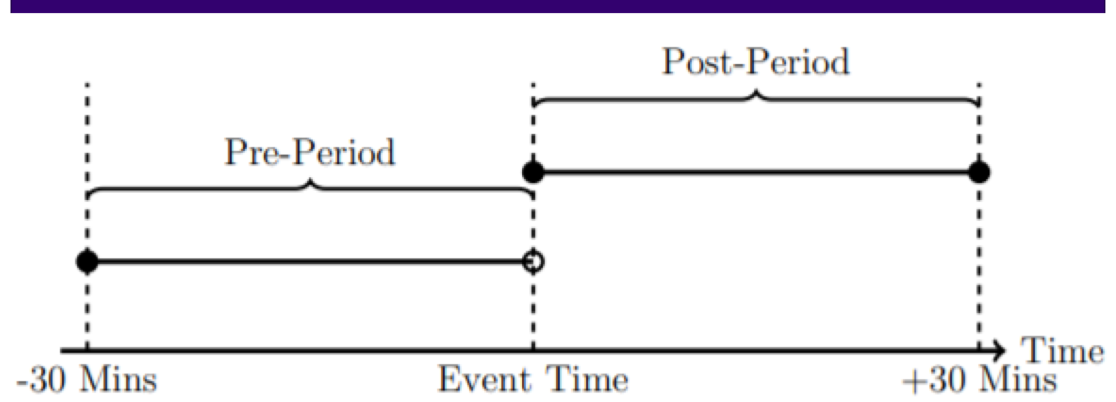


Figure 1: Definition of Pre- and Post- event period in 30 minutes case

## Pre/Post Release Drift

In summary, to confirm the existence of information leakage, we construct 5 portfolios based on pre-release price run-up, $Q_{30} = \log\left(\frac{P_0}{P_{-30}}\right)$, and compare the post- (trades occurred 30 minutes after the announcement) and the pre-release (trades occurred 30 minutes before the announcement) drift for all matched SEC form submissions and the results show that the mean returns started to move much earlier than the announcement as shown in Figure 2 which plots the return for every 15-second interval from 30 minutes prior (t=0) to event time to 30 minutes after event time (t=240). Incidentally, the event time (i.e., the EDGAR acceptance time stamp, t=120) has the most extreme returns for that 15-second interval, as expected. Figure 3 plots the cumulative return, and the fact that the returns of the portfolio's cumulative return run-up prior to event time agree in the same direction (and magnitude) with their cumulative return after event time is evidence in support of information leakage.

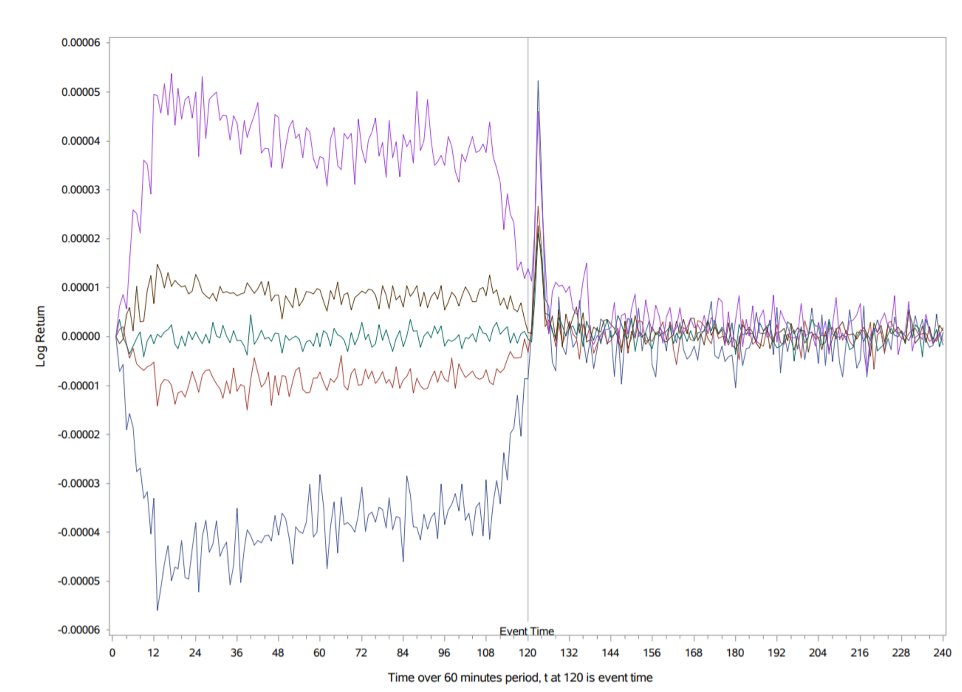


Figure 2: Average excess return for 15 seconds interval from 30 minutes before the acceptance to 30 minutes after the acceptance SEC form filings, in the years 1994 to 2017.

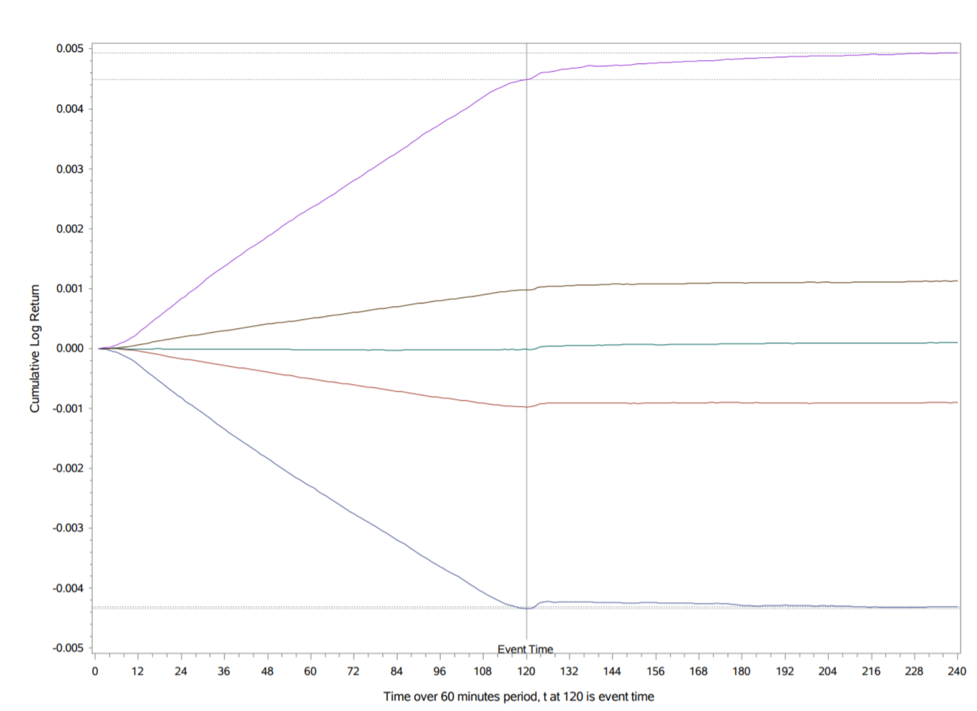


Figure 3: Cumulative average excess return for 15 seconds interval from 30 minutes before the acceptance to 30 minutes after the acceptance of SEC form filings, for all SEC form-types, in the years 1994 to 2017.

## Single/Double Sort

The Single-Sort table, Table 1, reports average returns and differences in returns, together with t-statistics and standard errors, and we find that generally speaking, as we go from portfolio 1 to portfolio 5, the higher the price runup prior to the event, the higher the cumulative return post the event, consistent with the information leakage conjecture.

**all form-types**

| | Panel A: $J$=30-mins | | | Panel B: $J$=90-mins | | | Panel C: $J$=180-mins | | |
|---|---|---|---|---|---|---|---|---|---|
| *PORT* | Mean | $t$-Stat | StdErr | Mean | $t$-Stat | StdErr | Mean | $t$-Stat | StdErr |
| 1 | -0.000051** | -2.426 | 2.1e-05 | -0.000018 | -0.433 | 4.1e-05 | -0.000218 | -1.319 | 0.000165 |
| 2 | 5E-6 | 0.437 | 1.1e-05 | -0.000064*** | -2.612 | 2.5e-05 | 0.000232** | 2.462 | 9.4e-05 |
| 3 | 0.000053*** | 4.95 | 1.1e-05 | 0.000113*** | 5.15 | 2.2e-05 | 0.000274*** | 3.041 | 9e-05 |
| 4 | 0.000143*** | 12.889 | 1.1e-05 | 0.000223*** | 9.396 | 2.4e-05 | 0.000164* | 1.814 | 9e-05 |
| 5 | 0.000504*** | 23.823 | 2.1e-05 | 0.000616*** | 16.113 | 3.8e-05 | 0.000771*** | 4.91 | 0.000157 |
| 5−1 | 0.000555*** | 18.681 | 3e-05 | 0.000632*** | 11.196 | 5.6e-05 | 0.000983*** | 4.28 | 0.00023 |
| | Total Observations N=832256 | | | Total Observations N=492775 | | | Total Observations N=64188 | | |

Table 1: Single Sort Results . The sample includes all form-types, for the time period September, 2003 to December, 2018. Each firm-announcement level observation are included in this sample if J-minutes prior or after the event-time still falls within the normal trading hours, (i.e., 9:30am to 4pm), where J=30, 90, or 180 minutes. The event-time here is the EDGAR "acceptance timestamp". All observations in the sample are sorted into five portfolios based on the sorting variable $Q_J$, with portfolio 1 representing the lowest quintile and portfolio 5 representing the highest quintile. We present equal-weighed returns for each portfolio. The long-short returns of going long for the highest quintile and going short for the lowest quintile ("5-1") are also shown. The table reports average returns and differences in returns, together with t-statistics and standard errors. To indicate significance: *** p < 0.01, ** p < 0.05, *p < 0.1.

In the Double Sort, or Portfolio Difference results, Table 2, we clearly see that there is universally highly statistically significant positive return differences between portfolio 5 and portfolio 1 (and between higher numbered portfolio and lower numbered portfolio), indicating that the higher the price runup prior to the form release, the higher the post release cumulative return, consistent with the conjecture that there is significant information leakage prior to releases.

Generally, the 30-minutes results are stronger than the 90- and 180-minutes results. In addition, we also present results in which we restrict the sample period to dates after March, 2015, when SEC updated its PDS release system, in the appendix of the paper online, and we note that there is no substantial qualitative difference for the results due to the change in sample period.

**all form-types, $J$=30 minutes**

| | Mean | | | | $t$-Stat | | | | StdErr | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $PORT_j$ \ $PORT_i$ | 5 | 4 | 3 | 2 | 5 | 4 | 3 | 2 | 5 | 4 | 3 | 2 |
| 4 | 0.000362*** | - | - | - | 15.155 | - | - | - | 2.4e-05 | - | - | - |
| 3 | 0.000452*** | 0.00009*** | - | - | 19.074 | 5.874 | - | - | 2.4e-05 | 1.5e-05 | - | - |
| 2 | 0.0005*** | 0.000138*** | 0.000048*** | - | 20.85 | 8.809 | 3.117 | - | 2.4e-05 | 1.6e-05 | 1.5e-05 | - |
| 1 | 0.000555*** | 0.000194*** | 0.000104*** | 2.349 | 18.681 | 8.164 | 4.397 | 2.349 | 3e-05 | 2.4e-05 | 2.4e-05 | 2.4e-05 |
| | Total Observations N=833184 | | | | | | | | | | | |

Table 2: Portfolio Return Differences. The sample includes all form-types. Each firm-announcement level observation are included in this sample if J-minutes prior or after the event-time still falls within the normal trading hours, (i.e., 9:30am to 4pm), where J=30 minutes. The event-time here is the EDGAR "acceptance timestamp". All observations in the sample are sorted into five portfolios based on the sorting variable $Q_J$, with portfolio 1 representing the lowest quintile and portfolio 5 representing the highest quintile. We present equal-weighed return difference between portfolio i and portfolio j, where i can be equal to 5, 4, 3, or 2, and j can be equal to 4, 3, 2, or 1. In other words, the long-short returns of going long for portfolio i and going short for portfolio j is shown. The table reports average returns and differences in returns, together with t-statistics and standard errors. To indicate significance: *** p < 0.01, ** p < 0.05, *p < 0.1.

## Pre-release period Buyer/Seller-initiated Trade Volume Categorized According to Lee and Ready (1991)

We conduct this exercise by regressing the cumulative return of 30-minute interval prior to the event to an "aggressive-buy-indicator", $X_{LR}$, which is calculated from the difference between the number of shares in trades that are classified as buyer initiated according to Lee and Ready (1991), $V_B$, and the number of shares in trades that are classified as seller initiated according to Lee and Ready (1991), $V_S$, and divided by the total volume traded in the pre-event horizon, $V$, formally, $X_{LR} = \frac{V_B - V_S}{V} = \frac{V_B - V_S}{V_B + V_S}$. While the intercept is statistically significant only for all forms combined and for form 10-K, the coefficient on $X_{LR}$ is highly statistically significant regardless of the form type, consistent with our central proposition that the there is information leakage prior to SEC form announcements, and the price movement prior to release is due to conscientious trading activity and not due to random market microstructure noise.

| All Forms | | Coef. | Std Err. | $t$-Stat | $P>\|t\|$ |
|---|---|---|---|---|---|
| | Intercept | 0.0*** | 0.0 | 2.651 | 0.008 |
| | $\beta$ | 0.008*** | 0.0 | 20.032 | 0.0 |
| 8K | | Coef. | Std Err. | $t$-Stat | $P>\|t\|$ |
| | Intercept | 0.0 | 0.0 | 0.619 | 0.536 |
| | $\beta$ | 0.011*** | 0.001 | 12.381 | 0.0 |
| 10Q | | Coef. | Std Err. | $t$-Stat | $P>\|t\|$ |
| | Intercept | 0.0 | 0.0 | 0.564 | 0.573 |
| | $\beta$ | 0.01*** | 0.001 | 10.296 | 0.0 |
| 10K | | Coef. | Std Err. | $t$-Stat | $P>\|t\|$ |
| | Intercept | 0.001* | 0.0 | 1.838 | 0.066 |
| | $\beta$ | 0.008*** | 0.002 | 4.5 | 0.0 |
| 4 | | Coef. | Std Err. | $t$-Stat | $P>\|t\|$ |
| | Intercept | 0.0 | 0.0 | 0.988 | 0.323 |
| | $\beta$ | 0.006*** | 0.001 | 7.528 | 0.0 |
| SC 13D | | Coef. | Std Err. | $t$-Stat | $P>\|t\|$ |
| | Intercept | 0.001 | 0.001 | 0.744 | 0.457 |
| | $\beta$ | 0.01*** | 0.004 | 2.7 | 0.007 |
| SC 13G | | Coef. | Std Err. | $t$-Stat | $P>\|t\|$ |
| | Intercept | 0.0 | 0.0 | 1.588 | 0.112 |
| | $\beta$ | 0.008*** | 0.001 | 10.706 | 0.0 |
| SC 13F | | Coef. | Std Err. | $t$-Stat | $P>\|t\|$ |
| | Intercept | 0.0 | 0.0 | 0.451 | 0.652 |
| | $\beta$ | 0.008*** | 0.001 | 8.968 | 0.0 |

Table 3: Panel Regression results of 30-minutes pre-event returns on XLR, September, 2009 to January, 2019. Each firm-announcement level observation are included in this sample if 30-minutes prior or after the event-time (i.e., EDGAR "acceptance timestamp" ) still falls within the normal trading hours, (i.e., 9:30am to 4pm). The 30-minutes pre-event returns is calculated as $Q_{30} = \log\left(\frac{P_0}{P_{-30}}\right)$ where $P_0$ is the price at the event time and $P_{-30}$ is the price 30 minutes prior to the event time. We perform panel regression in the form of: $Q_{30} = Intercept + \beta \times X_{LR}$, where $X_{LR}$ is the difference between buy volume and sell volume divided by the total volume between the event time and 30 minutes prior the event time. We run the regression on the sample with all forms combined, as well as on each of the form 4, 8K, 10K, 10Q, SC 13D, SC 13G, and SC 13F. The table reports intercept, coefficient β and standard errors, together with t-statistics and standard errors. To indicate significance: *** p < 0.01, ** p < 0.05, * p < 0.1.

## Conclusion

Using the TAQ millisecond data and a comprehensive sample of the acceptance timestamp of SEC form filings, we find strong evidence of information leakage in the 30-minute intervals around Edgar acceptance timestamp of corporate SEC filings, in that if the stocks are ranked into 5 portfolios bins based on the price run-up prior to filing release, for all form types, the events with the highest run-up would also have the highest price increase post filing release. Also, depending on the type of the SEC filing, for filings that could contain both positive and negative information, for example, form 8K 10K and 10Q (as opposed to SC 13D or 13G which generally can only be good new for stock price), the events with the most run-down prior to the release would also have the most price decrease post filing release. Our finding is not explained by momentum. Incidentally, panel regression results in Table 3 provide evidence in support of the presence of informed trading.