

# Online Appendix

for

PEAD.txt: Post-Earnings-Announcement Drift Using Text

Vitaly Meursault, Pierre Jinghong Liang, Bryan R. Routledge, Madeline Marco Scanlon

## Online Appendix A. Dataset Construction Details

We construct the corpus of earnings call transcripts using the Capital IQ Transcripts database, which is available through the Wharton Research Data Services (WRDS) platform. Various numerical variables are constructed based on the Center for Research in Security Prices (CRSP), Compustat, and Institutional Brokers' Estimate System (IBES) datasets available through the WRDS platform. All continuous variables are winsorized at the 1% and 99% levels.

To merge the Transcripts, CRSP, and IBES datasets, we use the identifier crosswalk provided with the Transcripts dataset, as well the CRSP to IBES linking table constructed using the script by Qingyi (Freda) Song Drechsler.<sup>20</sup> We link Compustat to other datasets using the Compustat / CRSP merged dataset available at WRDS.

To compute announcement day abnormal returns used to create the ML-based earnings call surprise measure, we use the WRDS Event Studies tool to compute one-day abnormal returns using the Fama–French plus momentum risk model with default estimation window, number of valid returns, and gap parameters. In the Fama-French three-factor model, proposed in Fama and French (1993), stock returns are adjusted for excess returns on the market portfolio of stocks, the size factor, and the book-to-market factor. These factors adjust for some major correlations between stock returns. The excess return on the market portfolio of stocks captures the common variation in stock returns. The size factor relates to profitability, as small firms tend to have lower earnings on assets than large firms. The book-to-market factor follows from the fact that firms with a low (high) stock price relative to book value tend to have persistently low (high) earnings. The Carhart momentum factor, proposed in Carhart (1997), extends the Fama–French three-factor model to account for the profitable trading strategy associated with stock price momentum. Momentum is the tendency of stock price to continue increasing following a

---

<sup>20</sup><https://www.fredasongdrechsler.com/full-python-code/iclink>. Last accessed: 08/12/2020.

rise and to continue decreasing following a decline. By using the Fama–French three-factor model with the Carhart momentum factor in the calculation of abnormal returns, we extract the “unexpected” portion of returns that we hypothesize to be associated with the information contained in the earnings call.

For the econometric analysis of the earnings call surprise measure, we compute long-run cumulative abnormal returns based on the Fama–French six size and book-to-market matched portfolios.<sup>21</sup>

We compute classic earnings surprises, following Livnat and Mendenhall (2006).<sup>22</sup> We compute three versions of standardized unexpected earnings (SUE). The first version (*SUE1*) is based on earnings reported in Compustat and measures the surprise as the difference between the current earnings and the earnings in the same quarter the previous year. The second version (*SUE2*) is computed in the same way as (*SUE1*) but uses earnings adjusted for special items. The main specification, *SUE3*, uses analyst forecasts to measure deviations in reported earnings from expectations. Prior to firm earnings announcements, financial analysts assigned to the firm issue predictions of future firm performance based on firm fundamentals, prior industry experience, and other information. These forecasts are then compared to reported earnings to calculate unexpected earnings (also called earnings surprises) and adjusted by share price. Earnings surprises are summary statistics for new information contained in earnings announcements. The prices tend to move in the direction of the surprise, and as a result, buying (selling) firms with positive (negative) *SUE* can generate excess returns.

Earnings calls are often conducted either before market open or after market close. This influences which daily return should be used as the earnings call day return. For calls that happened in the morning of day  $t$ , the return on day  $t$  (price change between the

---

<sup>21</sup>The cutoffs used to match stocks to their benchmark portfolios and the portfolio returns are from Kenneth R. French’s data library at [https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data\\_library.html](https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). Last accessed: 08/12/2020.

<sup>22</sup>We use a python script available at <https://www.fredasongdrechsler.com/full-python-code/pead>. Last accessed: 08/12/2020.

market close on day  $t - 1$  and the market close on day  $t$ ) is appropriate. But if the call happens on day  $t$  after the market close, the return on day  $t + 1$  (price change between the market close on day  $t$  and the market close on day  $t + 1$ ) is the one that covers the time of the call. The Capital IQ Transcripts database does not include the time of the call, but it can be estimated using the transcript creation time. We use the following heuristic. Given that the date of the earnings call is  $t$ ,

- If the first version of the transcript is marked as preliminary and was uploaded before 3:00 PM Eastern Time on the day  $t$ , we use the return between  $t - 1$  and  $t$  as the call day return.
- If the first version of the transcript is marked as final (edited) and was uploaded sometime during the day  $t$ , we use the return between  $t - 1$  and  $t$  as the call day return.
- Otherwise, we use the return between  $t$  and  $t + 1$  as the call day return.

Note that this approach picks the later return as earnings call day return in cases when we might be unsure about when the call happened, safeguarding against using information from the future.

## Online Appendix B. Regularized Logistic Regression

We use a regularized logistic text regression to predict short-term market reactions based on earnings transcript text. Then we use model predictions to construct a measure of earnings call text surprises. We consider an earnings call to have a positive SUE.txt if the model assigns high log-odds to one-day abnormal returns being positive and low log-odds to one-day abnormal returns being negative.

We use the elastic net model developed by Zou and Hastie (2005) and implemented in R package *glmnet* (Friedman, Hastie and Tibshirani, 2010).<sup>23</sup> It is a multinomial logistic regression model that combines  $L1$  and  $L2$  penalties of the lasso and ridge methods. The model, in log-odds form, is the following:

$$\begin{aligned} & \text{for } r \in \{H, F, L\}, \\ \text{log-odds}(r) &= \log \frac{\Pr(R_{t=0} = r | X = x)}{\Pr(R_{t=0} \neq r | X = x)} = \beta_{0r} + \beta_r^T x, \end{aligned} \tag{6}$$

where  $R_{t=0}$  is the earnings call day return split into categories  $r \in \{H, F, L\}$ , *high*, *flat*, and *low* (more on the category split below);  $x$  is a vector of predictor variables (word frequencies or standard numerical variables);  $\beta_{0r}$  is the intercept; and  $\beta_r$  is the vector of regression coefficients.

The objective function is the following:

$$\begin{aligned} L(\{\beta_{0r}, \beta_r\}_{r \in \{H, F, L\}}) &= \\ & - \left[ \frac{1}{N} \sum_{i=1}^N \left( \sum_{r \in \{H, F, L\}} y_{ir} (\beta_{0r} + x_i^T \beta_r) - \log \left( \sum_{r \in \{H, F, L\}} e^{\beta_{0r} + x_i^T \beta_r} \right) \right) \right] + \\ & \lambda \left[ (1 - \alpha) \|\beta\|_2^2 / 2 + \alpha \sum_{j=1}^p \|\beta_j\|_1 \right], \end{aligned} \tag{7}$$

---

<sup>23</sup>While the model we use is distinct from support vector regression used in Frankel et al. (2016), the general approach of using a regularized linear model is the same. Our approaches also share a feature of reestimating the model over time.

where  $y_{ir}$  is an indicator variable equal to one if the one-day return for an earnings call  $i$  falls into category  $r$ ,  $\alpha$  is the mixing hyperparameter that controls the relative strength of  $L1$  and  $L2$  regularization,  $\lambda$  is the hyperparameter that controls overall strength of regularization, and  $p$  is the total number of coefficients associated with all categories. To avoid excessive subscripts, we denote the combination of vectors  $\beta_H, \beta_F$ , and  $\beta_L$  as  $\beta$ , and individual coefficients as  $\beta_j$ . We set  $\alpha$  to 0.5 and choose  $\lambda$  using 10-fold cross-validation on the training sets.

Our target variable is one-day abnormal returns split into *high*, *flat*, and *low* categories in the following way:

1. *Flat*: 33% of observations closest to zero (first tercile of the absolute value of abnormal returns). The cutoffs for training and test sets are based on the terciles of the training set.
2. *High*: large positive abnormal returns (positive returns not falling into the *flat* category).
3. *Low*: large negative abnormal returns (negative returns not falling into the *flat* category).

The SUE.txt is an a priori measure in the sense that one can compute it without access to returns associated with a specific call. We achieve this by setting up our problem as a sliding window prediction task, as demonstrated in Table 2. At each iteration, we use eight consecutive quarters of observations as the training set, and one subsequent quarter as the test set. We slide the window one quarter at a time. We perform 40 iterations, with test quarters ranging from the first quarter of 2010 to the fourth quarter of 2019. We only use test set predictions to calculate SUE.txt, therefore ending up with 40 quarters of a priori measurement. Within each training set, we perform 10-fold cross-validation to find the best value of the regularization strength hyperparameter  $\lambda$ .

## Online Appendix C. Paragraph Groups Examples

1. **Financial accounting.** Discussions of financial results focused around metrics from various financial statements.

- (a) **Bottom line.** Net earnings / income, net losses or EPS. Keywords: *earnings, eps, net earnings, net income, net loss.*

Thank you, Jack. Good morning, everyone. For the third quarter, we generated net income of \$118.9 million or \$2.96 per fully diluted share compared with \$100.4 million or \$2.33 per diluted share in the third quarter last year. (The Hanover Insurance Group, Inc., Q3 2019 Earnings Call, Oct 31, 2019.)

- (b) **Metrics.** A wide range of financial metrics excluding the bottom line ones above. Keywords: *amortization, book-to-bill, capex, cash, cash flow, cashflow, charge, cogs, cost, current assets, current liabilities, ebit, ebitda, ebt, expenditure, expense, fee, goodwill, gross margin, income, inventory, liability, long - term debt, loss, margin, operational income, operational margin, payable, pp&e, profit, profit margin, profitability, receivable, retained earnings, revenue, sale, tax, treasury stock, working capital.*

Turning to some other key metrics. In addition to delivering robust constant currency revenue growth, which we were able to achieve even when facing a – even when faced with a difficult prior-year comparable, it was – I was also pleased to see the year-over-year and sequential expansion in both our adjusted gross and operating margin. And during the quarter, our adjusted gross margin reached 58.6%, which was an increase of 160 basis points as compared to the prior year and 90 basis points sequentially, while our adjusted operating margin totaled 27%, which was an increase of 100 basis points as compared to the prior year and 180 basis points sequentially. (Teleflex Incorporated, Q3 2019 Earnings Call, Oct 31, 2019.)

- (c) **Adjustments.** Non-GAAP adjusted metrics, specific write-offs etc. Keywords: *accrual, adjusted, adjusting for, adjustment, allowance, benefit, conversion, depreciation, excluding, gaap, gain from, gain, gain on, impairment, non - gaap, non -gaap, non gaap, non- gaap, non-gaap, write - down, write off, write offs, write- offs, writedown, writeoff.*

Our MMP segment delivered adjusted earnings of \$210 million compared to \$300 million in the same period last year. Weak refining results affected MMP negatively this quarter. In addition, we also had a negative timing effect related to the valuation of gas in storage. According to accounting principles, we had a write-down of the gas inventories due to the drop in gas prices in the quarter. However, the gas inventory is sold forward at higher prices and expect to deliver is during winter. Without these timing effects, MMP would have delivered adjusted earnings within the guided range. (Equinor ASA, Q2 2019 Earnings Call, Jul 25, 2019.)

- (d) **Lending.** Financial metrics specific for lending institutions. Keywords: *bond, borrowing, capital ratio, debt, deposit, financing, interest, interest income, loan, noninterest income, return on assets, return on equity, roa, roe.*

Both Medallion Bank and Medallion Capital continue to add shareholder value. The consumer and commercial business segments recorded \$27.5 million of net interest income for the third quarter and \$77 million for the first 9 months of the year. Our consumer segments continue to grow as demand for the bank's products remain strong. In the 2019 9 months, Medallion Bank originated \$351.8 million of recreational and home improvement loans compared to \$333.4 million in the same period last year, including \$125.3 million in the 2019 third quarter. As of September 30, our net consumer portfolio stood at \$919 million, a 24% increase from the 2018 third quarter and has grown 21% year-to-date. (Medallion Financial Corp., Q3 2019 Earnings Call, Oct 31, 2019.)

- (e) **Financing.** Equity and debt raising, dividends and stock repurchases. Keywords: *dividend, equity raise, loan financing, pay interest, payback, payout, raise equity, repurchase, share purchase, stock purchase.*

Got it. And just for – and you can do this in multiple phases over the next 12 months? Or is it a 1/1 period you will get the full equity raise or one date, rather? (Regency Centers Corporation, Q3 2019 Earnings Call, Oct 31, 2019.)

2. **Operations management and marketing.** Production, delivery, supply and demand, and supply chain.

- (a) **Operational and marketing metrics.** Metrics that are not reported in traditional financial statements, like orders, backlog, volume, subscriptions etc. Keywords: *asset, backlog, barrel, booking, commission, orders, shipment, subscription, traffic, unit, volume.*

Moving to Page 10 and ER. Q3 2019 revenue grew 7% to \$64 million from \$60 million in the prior year. The growth in ER revenue was driven by higher volume and pricing. Adjusted EBITDA in Q3 declined to a loss of \$1.1 million from a profit of \$0.6 million in the prior year. The decline was primarily the result of unfavorable mix, higher material cost and was partially offset by pricing and volume. Our backlog grew 11% to \$195 million compared to \$176 million in the prior year. (Spartan Motors, Inc., Q3 2019 Earnings Call, Oct 31, 2019.)

- (b) **Segments.** Specific lines of business. Keywords: *application, business, category, group, product, project, segment, service, subsidiary.*

Turning to our business highlights by segment. In Personal Lines, we grew our topline by approximately 6%, as a result of new business growth at mid-single-digit rate increases. Our deep agency partnerships and account focus remain important differentiators for us in the market, especially in the face of increasing competition. We continue to increase our penetration in targeted markets, executing on our strategy to be



our agents carrier of choice for preferred account business. Account business now represents 85% of both new business and our overall portfolio. (The Hanover Insurance Group, Inc., Q3 2019 Earnings Call, Oct 31, 2019.)

- (c) **Supply chain.** Supply and demand, and specifically supply chain. Keywords: *demand, supplier, supply, supply chain.*

As discussed in prior calls, we expect our largest OEM customer for ligands to diversify their supply chain next year, which will reduce overall ligand demand for Repligen in 2020. To limit the impact on our overall proteins business, we continue to invest in a Repligen-owned ligand portfolio, and we are very encouraged by reports that the Protein A resin utilizing our NGL-Impact A ligand is gaining traction in the marketplace, with a number of key accounts expected to scale up over the next few years. (Repligen Corporation, Q3 2019 Earnings Call, Oct 31, 2019.)

- (d) **Production.** Production cycle, including development, launch, production etc. Keywords: *channel, completion, construct, develop, downstream, exploration, franchise, install, inventory management, open, out - source, outsource, overhaul, produce, production, relocate, remodel, repair, upgrade, upstream.*

And then, James, if I could, you're guiding up next quarter led by Defense and you called out GaN and X-band. Are you looking at production now with some of these large systems like SPY-60 or Gator? Or is it more development work? And if it's the former, can you give some kind of color on the run? I know this stuff has been in development for many years now, but some of these are very large systems with big unit volumes in the long term, just trying to get a feel for how Defense will play out over the next – well, long term actually, the next 12 months or so? And then I have a follow up. (Qorvo, Inc., Q2 2020 Earnings Call, Oct 31, 2019.)

- (e) **Interruptions.** Disruptions of firm operations including interruptions and shutdowns. Keywords: *delay, interruption, shutdown.*

As you know, we did encounter some challenges to begin the new year in early January during a time of extreme winter weather conditions. The Oil Sands base plant incurred a power interruption, which resulted in a controlled shutdown of extraction and upgrading. We executed on a very disciplined recovery process and returned the assets to service, with no lasting impacts to overall operations. We are now back at full production rates, and we remain on track to meet our guidance commitments for the year. (Suncor Energy Inc., Q4 2017 Earnings Call, Feb 08, 2018).

- (f) **Marketing.** Customers, prices and promotions. Keywords: *ad, check, client, customer, launch, population, price, promotion.*

Okay. Great. And you mentioned that ESG dealers place some orders early in the quarter. I wonder why that timing was there. Was there a

price increase that they were getting in front of? Or why those orders came in, was there sell-through to end customers? Was there a customer attached to those orders? (Federal Signal Corporation, Q3 2019 Earnings Call, Oct 31, 2019.)

3. **Global economics.** Discussions of global, national and regional economy as well as factors such as weather.

(a) **Foreign exchange.** Foreign exchange and currency environment. Keywords: *currency, foreign exchange, forex, fx.*

As a result of prevailing foreign exchange conditions, we now expect that foreign exchange will result in a 225-basis point headwind to full year revenue as compared to our previous expectation of a 150-basis point headwind. As such, we are lowering our as-reported revenue growth guidance from a range of between 6% and 6.5% to a revised range of between 5.75% and 6%. (Teleflex Incorporated, Q3 2019 Earnings Call, Oct 31, 2019.)

(b) **Seasonality and weather.** Seasonal factors including weather. Keywords: *season, seasonality, seasonally, weather.*

Even with the late season weather slowdown across much of the West, revenue, profitability and earnings improved significantly in 2018, and we expect 2019 to be another year of significant top and bottom line expansion. (Granite Construction Incorporated, Q4 2018 Earnings Call, Feb 20, 2019.)

(c) **General global economics.** General economic environment or conditions in which the firm operates. Keywords: *conditions, country, economic, economy, environment, export, fluctuation, geographic, import, inflation, international, macro, macroeconomy, region, stability, trend, uncertainty, world.*

We continue to meet competitive challenges by substantially reducing component and manufacturing costs, while introducing new products that improve productivity and increase flexibility for our customers. As widely reported, the macro economy has softened and the geopolitical climate has become more unstable over the last few months. As a result, we again found ourselves in a more uncertain position with limited near-term visibility to improved – improving business conditions. Despite the volatile geopolitical and business environment, we continue to invest in new products and application to substantially enhance our competitive position. (IPG Photonics Corporation, Q2 2019 Earnings Call, Jul 30, 2019.)

4. **Strategy.** Discussions of other parties with which the companies engage in strategic interactions, including competitors, partners and governments. Also discussions of mergers and restructuring, and general corporate strategy.

(a) **Competition.** Firm's competitors and competitive advantage. Keywords: *compete, competition, competitive, industry, market.*

You'll have to excuse me, Anthony, I don't go into the details on the pricing. I'm sure our competitors listen to the call, same as anyone else.

So I don't want to share that for competitive reasons. But I can tell you, Anthony, that we feel incredibly confident in our pricing strategy and we believe that it will not be a barrier to adoption. (Teleflex Incorporated, Q3 2019 Earnings Call, Oct 31, 2019.)

- (b) **Expansion.** Discussions of M&A, integrations and investments. Keywords: *acquisition, integration, investment, m a, m&a, merger, reinvestment.*

Now let me provide some more background on our latest acquisition. In late August, we purchased the assets of Backflow Direct, a California-based company that designs and manufactures large-diameter stainless steel backflows, primarily used in commercial fire protection applications. Backflow Direct was established about 7 years ago, and the founder is well known in the backflow prevention industries and to Watts as well. Strategically, Backflow Direct provides some innovative products in fire protection applications that broaden our offerings to meet customers' requirements. We are also excited that the former owner, a well-respected professional in backflow prevention, is joining our team to aid in future backflow development opportunities. We are excited to add Backflow Direct to our portfolio. (Watts Water Technologies, Inc., Q3 2019 Earnings Call, Oct 31, 2019.)

- (c) **Contraction.** Divestitures and divestments. Keywords: *divestiture, divestment.*

Yes. So about half of the revenue change is due to the divestitures, and a good chunk is due to FX. When I look at the range, I also hedge the range based on the lumpiness that we tend to see in the recall business, and we continue to see. But predominantly, it's FX and divestiture, and divestiture is about half of that. When I look at – on an EBITDA basis, the divestitures is about 1/3 of that, again, reflecting that these are lower-margin businesses. (Stericycle, Inc., Q3 2019 Earnings Call, Oct 31, 2019.)

- (d) **Partners.** Partners and joint ventures. Keywords: *joint venture, jv, partner.*

The first and largest is our agreement to form a 90/10 joint venture with a local real estate operator, we're the 90, for an initial 40-plus individual street retail properties in Hoboken, New Jersey. Our share of the investment approximates \$185 million. The properties, mostly apartments over Street retail, our prime retail – prime real estate sites on either Washington Street or 14th Street, two of Hoboken's main commercial thoroughfares. We're very bullish on Hoboken and its access to the increasingly important west side of Manhattan, including the \$20-billion-plus Hudson Yards development. That access is easier than in any areas of Manhattan through the PATH, Ferry and the bus through the immediately adjacent Lincoln Tunnel. One or more transportation choices of which is walkable from the buildings we're buying. (Federal Realty Investment Trust, Q3 2019 Earnings Call, Oct 31, 2019.)

- (e) **Deals.** Contracts, purchases and payments etc. Keywords: *award, bid, close, contract, deal, payment, purchase, transaction, win.*

Importantly, if we conclude the transaction, which we believe is likely, Noble will receive a payout of its share of the drilling contract and be free to market the Bully II. Given the effect of the advanced stages of the negotiations, we recognized a gross impairment on the Bully II in the third quarter of \$596 million or \$331 million net of noncontrolling interest. We would expect to receive the lump-sum payment in the fourth quarter of this year. (Noble Corporation PLC, Q3 2019 Earnings Call, Oct 31, 2019.)

- (f) **Government.** Government actions, regulations and approvals. Keywords: *action, approval, compliance, government, law, legal resolution, legal settlement, license, patent, regulation, regulator, regulatory, requirement, rule, trial.*

Additionally, the controlled rollout in Japan continues to gain traction as we generated \$9.6 million of sales in the quarter due to new sites and patient utilization. We opened 17 sites in Q2 and now are in approximately 100 hospitals, with 135 hospitals currently approved by the government. It is worth noting that we do expect to open fewer sites in the second half of the year as the team closes out the post-approval study and prioritizes a broader CP launch, which will require additional training at existing sites. (Abiomed, Inc., Q2 2020 Earnings Call, Oct 31, 2019.)

- (g) **Restructuring.** Restructuring, reorganization and lay-offs. Keywords: *lay - off, lay - offs, layoff, reorganization, restructuring.*

Then in addition, we put in an additional restructuring program in 2019, and that program doesn't generate savings until 2021. And there's also cash outflows that we have to invest in the project in 2019. There's also CapEx associated with it. (Teleflex Incorporated, Q3 2019 Earnings Call, Oct 31, 2019.)

- (h) **General strategy.** Strategy, initiatives and growth. Keywords: *efficiency, growth, initiative, opportunity, strategic, strategically, strategy.*

You guys have been very consistent with your strategy of organic investment and driving growth and margin expansion. So I'm going to focus first on growth. Your comments on the macro were helpful. I was wondering though if you could talk about some of your strategies of penetrating some of your international markets like the Middle East. What does that look going forward? Are you seeing traction there? And then also, if you could talk about some of your connected solutions? And maybe an update there. (Watts Water Technologies, Inc., Q3 2019 Earnings Call, Oct 31, 2019.)

5. **Forward-looking.** Paragraphs including forward-looking phrases that are identified following Muslu et al. (2015).

As a result of these changes, our revenues from new products introduced in the past 5 years is now at 15% compared to approximately 11% in 2016.

We aim to increase this number to above 20% by 2024. (Luxfer Holdings PLC, Q3 2019 Earnings Call, Oct 31, 2019.)