

## Case Study: Using Email Receipt Data to trade US Equities ([Eagle-Alpha](#))

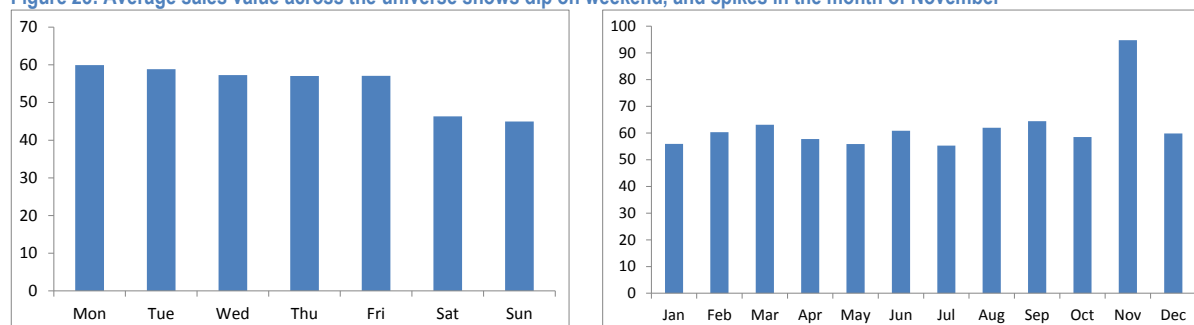
Aggregated consumer transaction data is a prominent category within business-process data. In the example below, we analyze a trading strategy based on email receipt data provided by Eagle Alpha. [Eagle Alpha](#) provides a variety of services across the spectrum of alternative data, including data sourcing, bespoke projects, teach-ins and analytical services. The firm provides ~500 data points spanning 20 categories, including online search trends, expert views and trade data.

The data was collected by an Eagle Alpha partner firm, which tracks transaction data representing 80% of all online purchases. The firm tracks such data for >5000 retailers and provides item and SKU-level transaction data filtered into 53 product categories. The email receipt data time series we analyzed below starts in 2013. This data included a constant group of users (that are still active and were active prior to Dec 1, 2013). The data included aggregated figures for total dollar spend, total number of orders, and the total number of unique buyers.

We have analyzed a dataset of email receipts for 97 companies. 36 of these were private companies, and 61 public, 31 of which were S&P 500 constituents. Taking liquidity into consideration, we decided to test trading signals for the S&P 500 companies only.

Given the disparity of actual spend for different companies, it is necessary to normalize dollar spend into relative score (e.g. percentage change, or normalize data for an average spend). There is also meaningful seasonality in the data. For instance spend data is slightly lower on weekends and is higher during month of November.

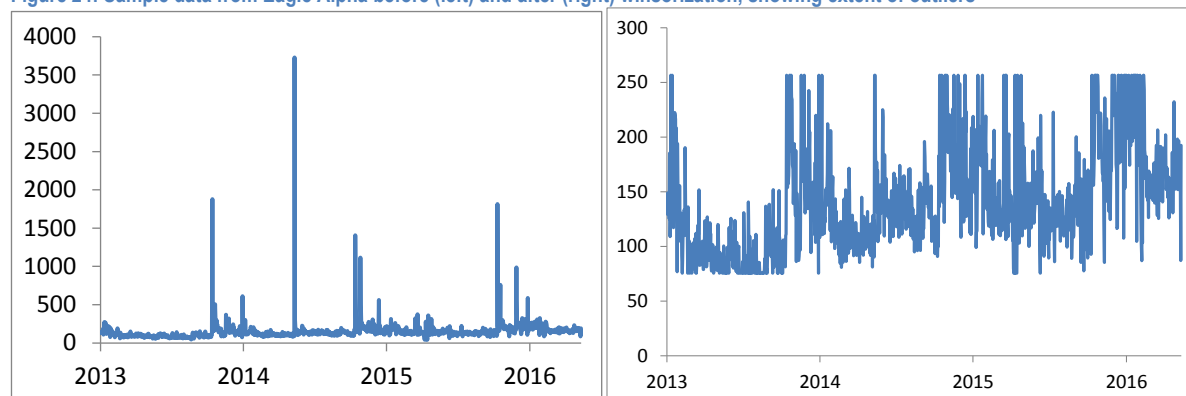
Figure 23: Average sales value across the universe shows dip on weekend, and spikes in the month of November



Source: J.P.Morgan Macro QDS, Eagle Alpha

The actual daily sales time series were volatile and we noticed data outliers. For this reason, we winsorized the time series. The Figure below shows a time series for one particularly noisy company without and with winsorization.

Figure 24: Sample data from Eagle Alpha before (left) and after (right) winsorization, showing extent of outliers



Source: J.P.Morgan Macro QDS, Eagle Alpha

We analyzed three time series: the dollar spend, number of orders and number of buyers. While number of orders and number of buyers are highly correlated (~99%), dollar spend is not highly correlated with number of buyers/orders (~25%).

We aggregated the daily spend/order/buyer data into a weekly score and calculated week-over-week percentage change for each. After winsorizing to 5<sup>th</sup>-95<sup>th</sup> percentile, we tested both the level and z-score as signals. Based on a cross-sectional comparison, we went long the top 5 stocks and short the bottom 5 stocks. The portfolio was rebalanced weekly.

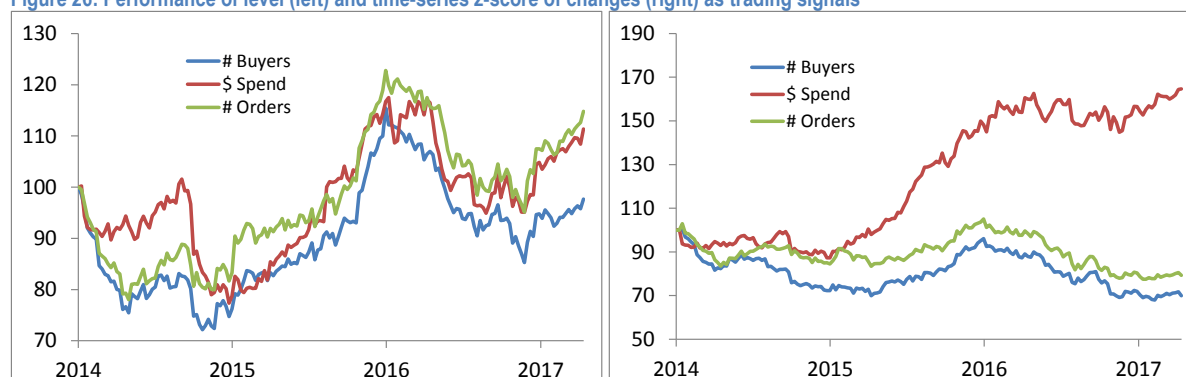
Figure 25: Sharpe ratios of various strategies using Dollar Spend, Buyer Count and Order Count

Dollar Spend Data	Top 6/ Bottom 6	Buyer Count Data	Top 6/ Bottom 6	Order Count Data	Top 6/ Bottom 6
Level	0.29	Level	0.02	Level	0.36
Z-score 4 week	1.13	Z-score 4 week	-0.71	Z-score 4 week	-0.49
Z-score 5 week	0.72	Z-score 5 week	-0.49	Z-score 5 week	-0.14
Z-score 6 week	0.67	Z-score 6 week	0.04	Z-score 6 week	0.11

Source: J.P.Morgan Macro QDS, Eagle Alpha

We also plot cumulative returns using the level (i.e. percentage of aggregated figure) and the 4-week z-score for all 3 data sets.

Figure 26: Performance of level (left) and time-series z-score of changes (right) as trading signals



Source: J.P.Morgan Macro QDS, Eagle Alpha