Martyn Jepson and Mehul Salhotra

MSCF Investments 2019

Songs and Stocks

Using the Billboard Number 1 song to predict S&P returns.

# Introduction (Why are we doing this?)

## Background

Knowledge of investor sentiment can be used to build a profitable trading strategy due to its mean reverting nature. If sentiment is high (positive) and stock prices increase, then they are likely to decrease as sentiment reverts to normal.

## Hypothesis

Our hypothesis is that the Billboard number 1 song can be used to measure investor sentiment; if the number 1 song is positive then sentiment is positive and vice versa. From this sentiment the future S&P 500 moves can be predicted.

To summarize, we hypothesize that Billboard number 1 song sentiment is negatively correlated with future stock returns.

## Value

If future stock return is correlated (albeit negatively) to Billboard number 1 song sentiment, then a profitable trading strategy can be created where the investor shorts the index when the Billboard number 1 is positive and long the index when Billboard number 1 is negative.

# Data Collection

## Number 1s

Our plan was to utilize an existing data set of weekly number 1 songs for the 10-year date range from 2005 to 2015. Initially we attempted to scrape directly from the Billboard site but server API limits made this unfeasible for such a large amount of data.

The dataset used is from <https://data.world/kcmillersean/billboard-hot-100-1958-2017> created by Sean Miller.

## Song Sentiment

Our plan was to utilize existing Natural Language Processing analysis of song sentiment which we can match to the songs in our scope. We used the dataset from <https://github.com/kevinschaich/billboard> created by [Kevin Schaich](https://kevinschaich.io/).

## S&P Returns

[MS]

# Data Clean-up and Analysis

## Song Data

The song data was cleaned and analyzed in a 3-step process:

1. Number 1 data was parsed from the previously mentioned data set filtering for the number 1 at each week and then selected for the data required. (See Top\_song\_pull.pdf)
2. Sentiment data was taken for each song available; the “negativity” of the song was subtracted from the “positivity” to give a net sentiment value for each song. This was then adjusted by subtracting the mean, so mean sentiment was 0 over the period. (See Song\_lookup\_notebook.pdf)
3. A unique key of title and artist was created after cleaning up differences between the data sets. Each week was then assigned the net sentiment value of the number one song. (See Sentiment\_merge.pdf)

## S&P Data

MS

# Data Features

## Predictor (Billboard Sentiment Normalized)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | Descriptor | Value | | count | 521 | | mean | 0.000 | | std | 0.126 | | min | (0.468) | | 25% | (0.061) | | 50% | 0.001 | | 75% | 0.067 | | max | 0.532 | | C:\Users\mjjep\AppData\Local\Packages\Microsoft.Office.Desktop_8wekyb3d8bbwe\AC\INetCache\Content.MSO\499CCB30.tmp C:\Users\mjjep\AppData\Local\Packages\Microsoft.Office.Desktop_8wekyb3d8bbwe\AC\INetCache\Content.MSO\3C5ED5A4.tmp |

## Response 1 (Log Adjusted 1 Week Return)

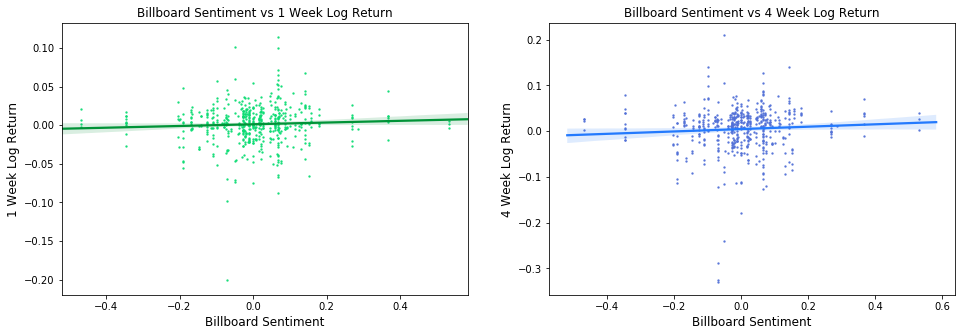
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | Descriptor | Value | | count | 521 | | mean | 0.001224 | | std | 0.023725 | | min | -0.200838 | | 25% | -0.008155 | | 50% | 0.002378 | | 75% | 0.013751 | | max | 0.113559 | | C:\Users\mjjep\AppData\Local\Packages\Microsoft.Office.Desktop_8wekyb3d8bbwe\AC\INetCache\Content.MSO\B420C8BE.tmp  C:\Users\mjjep\AppData\Local\Packages\Microsoft.Office.Desktop_8wekyb3d8bbwe\AC\INetCache\Content.MSO\AD72337C.tmp |

## Response 2 (Log Adjusted 4 Week Return)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | Descriptor | Value | | count | 521 | | mean | 0.005003 | | std | 0.045145 | | min | -0.33073 | | 25% | -0.013521 | | 50% | 0.011538 | | 75% | 0.029614 | | max | 0.209323 | | C:\Users\mjjep\AppData\Local\Packages\Microsoft.Office.Desktop_8wekyb3d8bbwe\AC\INetCache\Content.MSO\F21DF2EA.tmp  C:\Users\mjjep\AppData\Local\Packages\Microsoft.Office.Desktop_8wekyb3d8bbwe\AC\INetCache\Content.MSO\C25ADA88.tmp |

# Results

## Scatter Plot of Predictor and Responses



## OLS Regression

### Billboard Sentiment vs Log Adjusted 1 Week Returns

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Dep. Variable:** | 1\_Week\_Log\_Return | **R-squared:** | 0.003 |  |  |  |
| **Model:** | OLS | **Adj. R-squared:** | 0.001 |  |  |  |
| **Method:** | Least Squares | **F-statistic:** | 1.449 |  |  |  |
| **No. Observations:** | 501 | **Prob (F-statistic):** | 0.229 |  |  |  |
| **Df Residuals:** | 499 | **Log-Likelihood:** | 1117.6 |  |  |  |
| **Df Model:** | 1 | **AIC:** | -2231 |  |  |  |
| **Covariance Type:** | nonrobust | **BIC:** | -2223 |  |  |  |
|  | **coef** | **std err** | **t** | **P>|t|** | **[0.025** | **0.975]** |
| **const** | 0.0012 | 0.001 | 0.998 | 0.319 | -0.001 | 0.003 |
| **Billboard\_Sentiment** | 0.0111 | 0.009 | 1.204 | 0.229 | -0.007 | 0.029 |

|  |  |
| --- | --- |
| Plot of Residuals vs Fitted Values  C:\Users\mjjep\AppData\Local\Packages\Microsoft.Office.Desktop_8wekyb3d8bbwe\AC\INetCache\Content.MSO\F8F2D4E0.tmp | QQ Plot of Residuals vs Normal Distribution  C:\Users\mjjep\AppData\Local\Packages\Microsoft.Office.Desktop_8wekyb3d8bbwe\AC\INetCache\Content.MSO\4DE86DEE.tmp |

### Billboard Sentiment vs Log Adjusted 4 Week Returns

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Dep. Variable:** | 4\_Week\_Log\_Return | **R-squared:** | 0.004 |  |  |  |
| **Model:** | OLS | **Adj. R-squared:** | 0.002 |  |  |  |
| **Method:** | Least Squares | **F-statistic:** | 2.156 |  |  |  |
| **No. Observations:** | 501 | **Prob (F-statistic):** | 0.143 |  |  |  |
| **Df Residuals:** | 499 | **Log-Likelihood:** | 792.11 |  |  |  |
| **Df Model:** | 1 | **AIC:** | -1580 |  |  |  |
| **Covariance Type:** | nonrobust | **BIC:** | -1572 |  |  |  |
|  | **coef** | **std err** | **t** | **P>|t|** | **[0.025** | **0.975]** |
| **const** | 0.0045 | 0.002 | 2.009 | 0.045 | 9.85E-05 | 0.009 |
| **Billboard\_Sentiment** | 0.026 | 0.018 | 1.468 | 0.143 | -0.009 | 0.061 |

|  |  |
| --- | --- |
| Plot of Residuals vs Fitted Values  C:\Users\mjjep\AppData\Local\Packages\Microsoft.Office.Desktop_8wekyb3d8bbwe\AC\INetCache\Content.MSO\641B802C.tmp | QQ Plot of Residuals vs Normal Distribution  C:\Users\mjjep\AppData\Local\Packages\Microsoft.Office.Desktop_8wekyb3d8bbwe\AC\INetCache\Content.MSO\F5F0D31A.tmp |