Analysis of US Airline Stocks Performance Using Latent Dirichlet Allocation (LDA)

Graduate Researcher : Amina Issoufou Anaroua, Master of Science Aviation Finance

Faculty Mentor : Dr. Hari Adhikari

Abstract

The primary objective is to measure the impact of US airlines' stock performance following aviation related news announcements of differing sentiment and/or topics. The categories which account our data for aviation news are fuel prices, interest rate, inflation, airfares, foreign exchange rates, capital expenditures, and growth in output. The amount of such data and documents is expected to be enormous. So, we use a natural language processing, Latent Dirichlet Allocation (LDA) to investigate and search for patterns that can explain the movement of US airline stock. First, we will mine the aviation related data through text mining and topic modeling. Second, we will employ the LDA model approach to help us identify and capture the extent of certain topics mentioned in aviation news releases. Finally, we will use Event Study Methodology (ESM), which is a common method used to investigate stock price reactions to news announcements. We will apply this method to discover the significance of the relationship between the stock return and the associated event. Findings of this research will help stakeholders or investors understand the relationship between news of the aviation related events and the stock returns of the US airlines.

Introduction

Research is intended to use aviation digitized information to analyze the performance of US airline stocks. Determining the factors that influences the change in stock prices has always been a critical research subject. Recently, it is an accepted approach to include not only quantitative but also qualitative information to explain the change in stock prices due to financial events. When judging the fair price of a stock, news plays a crucial role for investors. The proposed work may potentially enhance investors 'understanding of the implications of aviation news announcements on the stock prices for US Airlines. News announcements are correlated to stock price performance. Previous studies utilized deep learning models such as ARIMA, RNN, and Facebook Prophet to improve the accuracy of stock price predictions by gathering a large amount of time series data. They focused on a daily stock price dataset consisting of closing stock prices from the Standard and Poor's 500 companies. They did not focus on a particular industry. Our studies use LDA to identify the magnitude, content, and severity of time trends in textual aviation news announcements.

Methodology

LDA was deemed appropriate for this study due to its ability to capture the extent to which a particular aviation topic was brought up in aviation news release. Our model approach is composed of first doing a mining process, then LDA, and finally an event study methodology for calculating abnormal returns. The mining process will be composed of text mining and topic modelling. Text mining is a widely used technique which assesses a sizable collection of documents to extract the information one seeks and/or learn new information. Topic modelling was deemed the most appropriate method in developing a model to accurately predict the future stock performance. Topic modelling method is effective for our study to automatically search, understand, organize, and summarize our large electronic archives. LDA is a probabilistic model which generates explanation(s) to sets of data as to why there are similarities, if there is any, by uncovering underlying implicit topics from the documents. A label is assigned to a topic found by the computer based on words. LDA estimates the number of topics in the whole announcement as well as the proportion of that topic. Event study methodology (ESM) is one of the most common used methods to investigate stock price reactions from news disclosures. ESM was designed to measure the effect of an event on a specific dependent variable which is commonly the stock price in finance.

Latent Dirichlet Allocation

1) For each document, a distribution over topics is selected randomly
- α , proportion parameter
- $\theta(d)$, topic distribution for document d
- $\theta(d,k)$, proportion of topic k in document d
- $\beta(k)$, distribution over the vocabulary

2) Next , the process continues for each word in the document :
- From the distribution $\theta(d)$, a topic is randomly selected

$z(d,n)$, topic assignment for the nth word in document d

$z(d)$, topic assignment for all the words in document d

- From the distribution, a word is randomly selected from the distribution over the fixed vocabulary corresponding to the topic assignment $z(d,n)$

$w(d,n)$, n-th word in document d , $w(d)$, words observed in document d

Based on this notation, the generative process for LDA is :

$$P(\beta_{1:K},\theta_{1:D},z_{1:D},w_{1:D}) = \prod_{i=1}^{K} P(\beta_i) \prod_{d=1}^{D} P(\theta_d)$$

$$\prod_{n=1}^{N} P(z_{d,n}|\theta_d) P(w_{d,n}|\beta_{1:K},z_{d,n}).$$

From the above equation, the probability of observing a word $w(d,n)$ depends on the topic assignment $z(d,n)$ and all the topics $\beta(1:K)$. Goal of LDA is to:
- Find the hidden structure topics
- Name topic distribution over each document
- Name topic assignment and each word over each document
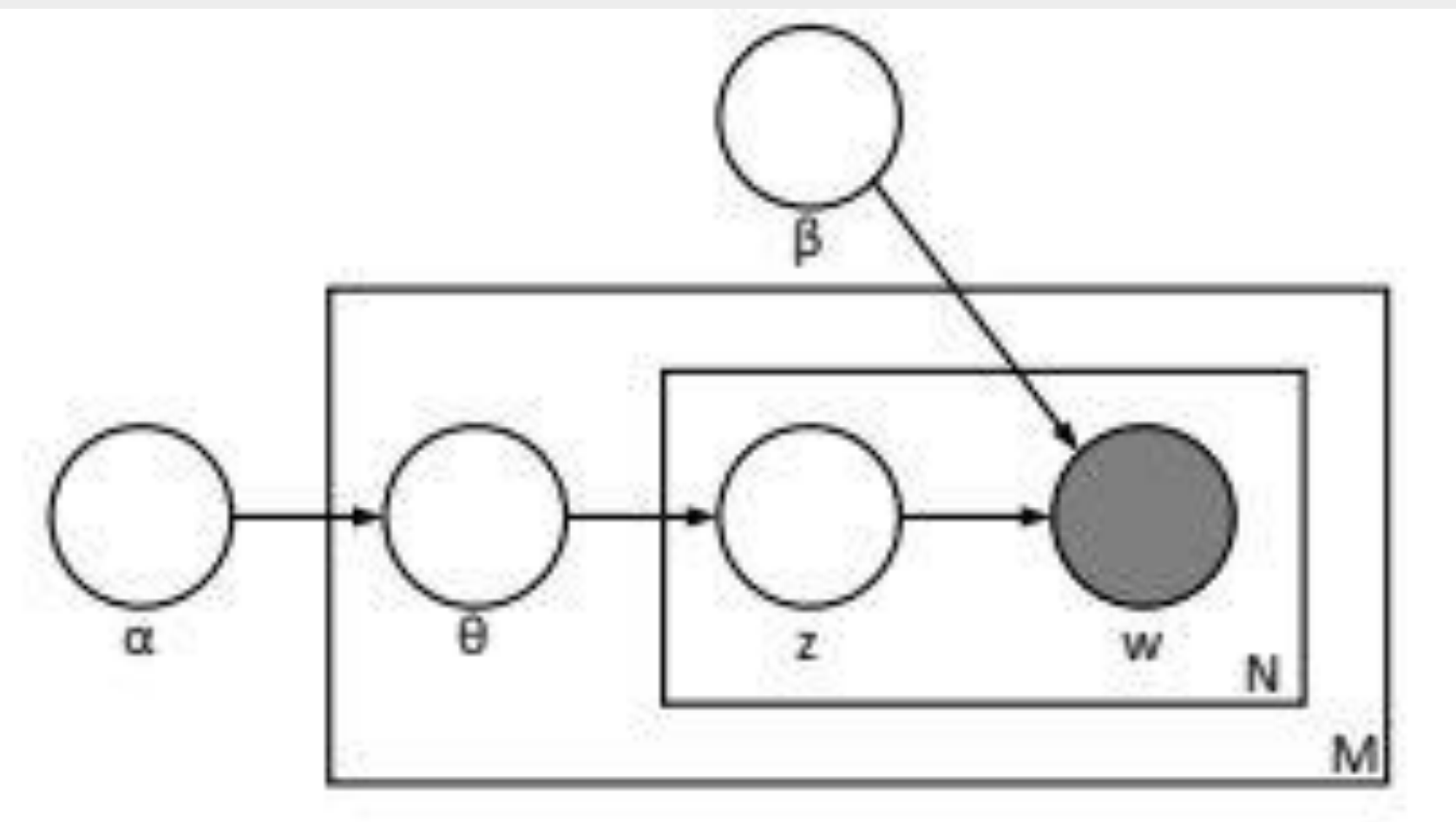
Source : (Feuerriegel, Ratku, & Neumann, 2016)

Event Study Methodology

- **Event of Interest** : analyze aviation news announcements that might have impact on airline stock returns
- **Event window** : measure focus on pre-announcement short-term stock performance. The day of publication is our chosen event window by using daily US airline stock price changes. To account for information leaks and insider trading, we calculate the cumulative abnormal return CAR(n, k). We use various windows from n days prior to the event announcement to k days after the event announcement
- **Normal returns**: expected return in the absence of a news announcement based on the market model. The market model is as follow:
  - $R(t) = \alpha + \beta Rm(t) + \varepsilon t \ \forall t \in T$
  - εt, zero mean disturbance term
  - $Rm(t)$, market return
  - T, event window
- **Abnormal returns**: is the difference between the actual and normal return of a security at time τ. The normal return comes from the market model.
  - $AR(\tau) = R(\tau) - E[R(\tau)| \neg X\tau]$
  - $R(\tau)$, actual return, measured by change of price of a security
  - $AR(\tau)$, abnormal return
  - $E[R(\tau)| \neg X\tau]$, expected return in the absence of an event $X\tau$

Source: (Feuerriegel, Ratku, & Neumann, 2016)

Current Status

- Identify the factors that influence stock price changes
- Identify with annual financial reports if textual part of the report contains some indication about future financial performance
- Which topics can be identified in Aviation news announcement?
- Once topics identified, analyze which news topics translate into non-zero stock market returns
- How do abnormal stock returns react to news belonging to the identified topics?
- Calculate the distribution of abnormal stock market returns following each news disclosure.
- Abnormal returns for each stock on disclosure date? DJ US Airlines Index?

References

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. The Journal of Machine Learning Research, 3, 993-1022.

Feuerriegel, S., Ratku, A., & Neumann, D. (2016). Analysis of how underlying topics in financial news affect stock prices using latent dirichlet allocation. In 2016 49th Hawaii International Conference on System Sciences (HICSS), 1072-1081.

For more Information

Contact : anarouaa@my.erau.edu