Creative Economy
2022, VOL. 6, NO. 6, 5-17
DOI: 10.47297/wspceWSP2516-251901.20220606

An Analysis of Different Factors' Contribution to Facebook Page Likes

Guodong Ma

The Australian National University, Canberra 2600, Australia

ABSTRACT
On the basis of big data regarding a publicly-available dataset about statistics of 500 posts published in 2014 on Facebook's page of a worldwide renowned cosmetic brand, this research aims to exhibit a comprehensive and rational deducing progress in character with the justified and formulated variables and visible results to be instrumental in the analysis of the factors' contribution to page total likes. In this paper, the generalized linear regression on the page total likes against the candidate covariates is drew to analysis, Bayesian lasso was used for selecting the essential variables and the Metropolis-Hastings algorithm was applied to obtain the posterior draws of the parameters.

KEYWORDS
Bayesian lasso; Metropolis-Hastings algorithm; Generalized linear regression; Facebook analysis

1 Introduction

Nowadays with the development of Internet technology, socializing online has become a more and more important and necessary part of we human beings ' life. We chat with friends and share our life stories and feelings on a variety of online social platforms. A typical example of social media is Facebook, an American online social media and social networking service owned by Meta, Inc.. According to Statista, the worlds ' first business platform, the website owns roughly 2.89 billion monthly active users as of June, 2021.[1] After registering an account, users can contact their family members, friends, colleagues, and even someone he has never met. The users are also able to post whatever articles, pictures, music, or videos they find interesting as well as their thoughts and opinions towards a certain event. Interactions between different users can be achieved by comments and private chats. Without a doubt, websites like Facebook make it much easier for modern people to stay connected and keep informed of news happening in their community and around the globe.

Considering its popularity, Facebook is certainly the most important media channel for companies to reach their potential clients and promote their products. Recovering the factors that contribute to page total likes is, therefore of great significance for these companies to design elaborate posts that cater to the consumers ' interest. Based on the valuable findings, advertisers can then make decisions on the type, time, and other features of what they wish to publish, proposing effective strategies that optimize the impacts of their pages.

Evaluating the exact contribution of each individual factor, unfortunately, is not easy. The

* Guodong Ma can be contacted at Guodong.Ma@anu.edu.au

popularity of a post is usually attributed to a number of different reasons. Some of the factors are qualitative rather than quantitative while some others are strongly correlated with each other. Therefore, variable selection shall be considered in the analyzing process.

In this project, by making use of a publicly-available dataset about statistics of 500 posts published in 2014 on Facebook's page of a worldwide renowned cosmetic brand,[2] I drew a generalized linear regression on the page total likes against the candidate covariates. Bayesian lasso was used for selecting the essential variables and the Metropolis-Hastings algorithm was applied to obtain the posterior draws of the parameters.

The structure of the rest of this passage is as follows. The next section describes the data I used and approaches to preprocess the data before making further analysis. The models I built and the methods I utilized for variable selection are also covered in the same section. In the third section, results are stated, including both the estimates of the model parameters and convergence diagnosis. In the last section, discussions are made with respect to the results in the previous sections as well as some prospects of further analyses and potential applications.

2 Methodology

2.1 Data and data processing

The original data set consists of 19 variables: page total likes, type, category, post month, post weekday, post hour, paid, lifetime post total reach, lifetime post total impressions, lifetime engaged users, lifetime post consumers, lifetime post consumptions, lifetime post impressions by people who have liked your page, lifetime post reach by people who like your page, lifetime people who have liked your page and engaged with your post, comment, like, share and total interactions.

Among the different factors, 'type' is qualitative while 'category' looks numeric and quantitative but in fact, different numbers represent different categories of the posts. To deal with this problem, I introduced some more indicator variables so as to mark the differences between each type or category of the posts. I set photo and action as the baselines, as they are the majority type and category of the posts (Table 1).

Table 1. Counts of different types or categories of posts in the dataset.

Type	Photo	Status	Link	Video
Counts	426	45	22	7

Category	Action	Product	Inspiration
Counts	215	130	155

For the six missing values in the original dataset, I treated them all like 0, since it is hard to 'predict' their true values from other entries. Since they only account for a trivial part of the whole dataset, they are more likely to be regarded as outliers and such imputation would not bring much bias to our estimations.

Other changes I made to the original dataset include deleting time variables 'post month' and 'post hour' while changing 'post weekday' into a binary one with 0 indicating it was posted during the weekdays and 1 for one posted at weekends. I did this based on the assumption that people are generally more likely to respond to the posts during the time when they are not working and changes in either the month or the exact time in one day would not lead to many differences.

Note that 'paid' is also a binary variable with 1 representing the post was paid by the

company to be published (advertised).

'Page total likes' indicates the popularity of each post. It varies greatly from post to post, with a mean of 123194 and a variance of 16272. Other statistics regarding this variable are listed as follows (Table 2) and plots are given to show the distribution of the points (Figure 1).

Table 2. Five-number summary of values of 'page total likes'.

Minimum	25% quantile	Median	75% quantile	Maximum
81370	112676	129600	136393	139441

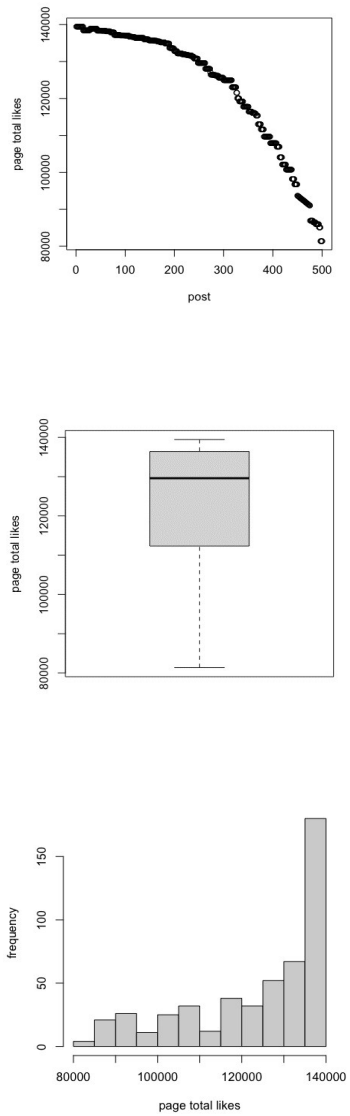

Figure 1. Up: Scatter plot of values of 'page total likes'; Middle: Boxplot of values of 'page total likes'; Down: Histogram of values of 'page total likes'

The descriptions of the rest of the features are as follows (Table 3), which was directly

taken from the original research paper that used this dataset.

Table 3. List of part of the features in the model.

Feature	Description
Lifetime post total reach	The number of unique users who saw a page post.
Lifetime post total impressions	Impressions are the number of times a post from a page is displayed, whether the post is clicked or not. People may see multiple impressions of the same post. For example, someone might see a Page update in News Feed once, and then a second time if a friend shares it.
Lifetime engaged users	The number of unique users who clicked anywhere in a post.
Lifetime post consumers	The number of people who clicked anywhere in a post.
Lifetime post consumptions	The number of clicks anywhere in a post.
Lifetime post impressions by people who have liked a page	Total number of impressions just from people who have liked a page.
Lifetime post reach by people who like a page	The number of users who saw a page post because they have liked that page.
Lifetime people who have liked a page and engaged with a post	The number of unique users who have liked a Page and clicked anywhere in a post.
Comments	Number of comments on the publication.
Likes	Number of "Likes" on the publication.
Shares	Number of times the publication was shared.
Total interactions	The sum of "likes," "comments," and "shares" of the post.

Before fitting the data to the model, I normalized all the features, making the mean for each covariate as 0 and variance as 1. I did this because the measure for each feature varies significantly and it is hard to directly compare the effects without assigning the features 'a uniform standard'. Also, it would be more convenient to see effects of variable selection.

2.2 Model of Page total likes

Since page total likes only take non-negative integer values, I assumed it to follow a Negative binomial distribution. For each individual post *i*, the Negative binomial distribution of its page total likes, N_i, has a mean $\mu_i = \exp\left(\mathbf{X}_{i\cdot}\boldsymbol{\beta} + \alpha\right)$ and variance $\sigma_i^2 = \tau\mu_i$, where $\alpha, \boldsymbol{\beta} = \left(\beta_1, \beta_2, \cdots \beta_p\right)^T, \tau$ are parameters to be estimated.

A general linear regression model is utilized, as I believe diverse factors can account for

the popularity of a post. Since parameters of the negative binomial distribution are positive, the mean μ is always greater than zero. Therefore, a 'log' link is used to ensure this property to be satisfied. For simplicity, I did not include a noise in the regression formula. Instead, I allowed more variance for the total page likes by introducing the τ parameter, which allows for a inflation of the variance of the distribution. Note that a requirement for this parameter is that $\tau > 1$, thus proposed new draws with $\tau \leq 1$ are rejected directly.

2.3 Bayesian Lasso

Following Park and Casella's work,[3] I used the Bayesian Lasso approach to select the important variables in this model.

For a regression problem, the Lasso is used to estimate $\boldsymbol{\beta} = \left(\beta_1,\beta_2,\cdots\beta_p\right)^T$ in the model

$$\boldsymbol{y} = \mu 1_n + \boldsymbol{X\beta} + \boldsymbol{\epsilon},$$

where $\boldsymbol{y}$ is the $n \times 1$ vector of responses (in my model $y_i = \log N_i$, $i = 1,2,\cdots n$), μ is the overall mean, $\boldsymbol{X}$ is the $n \times p$ matrix of standardized regressors and $\boldsymbol{\epsilon}$ is the noise vector (in my model it equals to a constant 0). Lasso estimates are usually regarded as L_1-penalized least squares estimates, i.e.

$$\arg\min_{\boldsymbol{\beta}} (\tilde{\boldsymbol{y}} - \boldsymbol{X\beta})^T (\tilde{\boldsymbol{y}} - \boldsymbol{X\beta}) + \lambda \sum_{i=1}^{p} |\beta_j|$$

for some $\lambda > 0$, where $\tilde{y} = y - \bar{y} 1_n$.

Under this framework, the Bayesian Lasso assume an independent and identical conditional Laplace prior for each individual $\beta_j, j = 1,2,\cdots p$, with mean 0 and variance $2\lambda^{-2}$, i.e.

$$\beta_j|\lambda \sim Laplace\left(\lambda^{-1}\right)$$

and the joint prior distribution for $\boldsymbol{\beta}$ is then

In my model, I set $\lambda = 10$, believing it can provide satisfactory variable selection results.

2.4 Prior selection for other parameters

Since there is no prior information regarding parameter α and τ, I assigned a flat improper uniform distribution for both of them.

2.5 Likelihood calculation

The posterior distribution for each set of parameters,

$$\pi\left(\boldsymbol{\beta}|\lambda\right) = \prod_{i=1}^{p} \frac{\lambda}{2} e^{-|\lambda|\beta_j|}. \theta = \{\alpha,\boldsymbol{\beta},\tau\} = \{\alpha,\left(\beta_1,\beta_2,\cdots\beta_p\right)^T,\tau\}$$

can be written as

$$\begin{aligned}\pi\left(\theta|\boldsymbol{N},\boldsymbol{X},\lambda\right) &\propto \pi\left(\boldsymbol{N},\boldsymbol{X}|\theta,\lambda\right)\pi\left(\alpha\right)\pi\left(\boldsymbol{\beta}|\lambda\right)\pi\left(\tau\right)\\ &\propto \pi\left(\boldsymbol{N},\boldsymbol{X}|\theta,\lambda\right)\pi\left(\boldsymbol{\beta}|\lambda\right)\\ &\propto \pi\left(\boldsymbol{N},\boldsymbol{X}|\theta,\lambda\right)\prod_{i=1}^{p}\pi\left(\beta_i|\lambda\right),\end{aligned}$$

where $N = \left(N_1,N_2,\cdots N_n\right)^T$, $n = 500$ and $\pi\left(\boldsymbol{N},\boldsymbol{X}|\theta,\lambda\right)$ is the likelihood function for θ:

$$\pi\left(\boldsymbol{N},\boldsymbol{X}|\theta,\lambda\right) = \prod_{i=1}^{n}\pi\left(N_i|\boldsymbol{X}_{i\cdot},\theta,\lambda\right),$$

$\pi\left(N_i|X_{i\cdot},\theta,\lambda\right) = f_{\mu_i,\sigma_i^2}\left(N_i\right)$ is the density function at value N_i of the Negative Binomial distribution

with mean $\mu_i = \exp\left(\mathbf{X}_{i.}\boldsymbol{\beta} + \alpha\right)$ and variance $\sigma_i^2 = \tau\mu_i$.

2.6 Metropolis-hastings algorithm

Since Laplace distribution is not a conjugate prior for Negative binomial distribution, there is no closed from of the marginal distribution of each parameters, I made use of the Metropolis-Hastings Algorithm to obtain the posterior draws of the parameters.

I chose a symmetric distribution to propose a new draw of parameters θ^* in an arbitrary iteration n so that $p\left(\theta^*|\theta^{(n-1)}\right) = p\left(\theta^{(n-1)}|\theta^*\right)$. I set the proposal distribution to be a multinomial one with mean $\theta^{(n-1)}$ and covariance matrix $\hat{\Sigma}$, i.e.

$$\theta^* \sim N\left(\theta^{(n-1)}, \hat{\Sigma}\right),$$

where $\hat{\Sigma}$ was set as a diagonal matrix in the first round of iterations and in later rounds proportional to covariance matrix of the posterior draws obtained in previous rounds so as to obtain a satisfactory acceptance rate. In this way, the acceptance ratio is the same in value as the posterior likelihood ratio between the proposed set of parameters and the original one.

The algorithm was fast to run, so I set the number of iterations in each round to be no smaller than 100,000. Several rounds of iterations were done to make sure I achieved the target posterior distributions. In each iteration, I first proposed a new set of parameters, θ^* from the distribution $N\left(\theta^{(n-1)}, \hat{\Sigma}\right)$. Then, I calculate the posterior likelihood ratio r between the proposed set of parameters and the previous draw. In the next step, I randomly sampled a value u from the Uniform distribution $Unif\,(0,1)$ and compared r with it. If $u \le r$, I accepted the proposal and store $\theta^{(n)} = \theta^*$. Otherwise, I rejected the proposal and store $\theta^{(n)} = \theta^{(n-1)}$.

Meanwhile, when coding with R software, to reduce computation error, I calculated the log posterior ratio instead, i.e.

$$\log r = \log \pi\left(\theta^*|\boldsymbol{N}, \boldsymbol{X}, \lambda\right) - \log \pi\left(\theta^{(n-1)}|\boldsymbol{N}, \boldsymbol{X}, \lambda\right).$$

To compare with this log ratio, the random value u is then sampled from the distribution $-Exp\,(1)$ (negative of a random value from the exponential distribution with mean 1). The new proposal θ^* is rejected if and only if $u > \log r$.

The following flow chart (Figure 2) provides a summary of the algorithm.

3 Results

A plot of the credible intervals and point estimates of the coefficients is given as Figure 3 while Figure 4 shows different variables' impact on the page total like after some scaling (a value greater than 1 means a positive contribution and a value smaller than 1 means a negative contribution). Obviously, the Bayesian Lasso managed to 'push' the majority of the posterior medians of the coefficients, $\boldsymbol{\beta}$, to 0. It is also clear that different factors have diverse contributions to the page total likes. The effects are to be analyzed in the next section.

I also included one plot of the results generated in the scenario when the penalty $\lambda = 1$ (Figure 5). Compared to the previous version, results are similar in terms of whether they are smaller or greater than 1, whilst the posterior medians deviate more from 1, which again illustrates the effects of the Bayesian Lasso.

Convergence of the Markov chains is assessed through traceplots and stationary plots. Since there are too many variables included in this model and they all share similar patterns, here I just show the traceplots and stationary plots of some of the coefficients as well as the

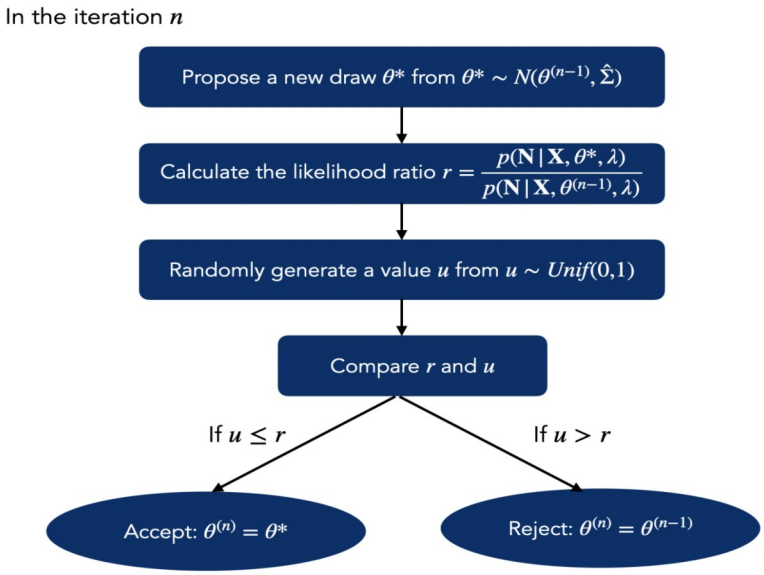

Figure 2. Flow chart of the Metropolis-Hastings Algorithm.

log-posterior values for all the draw, indicating convergence of the model (Figure 6, 7).

4 Conclusions

This study revealed the relationship between page total likes and a variety of candidate factors.

Most of the factors had little impact on the page total likes, including but not limited to whether the page was paid to advertise and whether the page was published at weekends. These results are quite surprising and are against our previous hypotheses that paid posts were more likely to be noticed by the users and that people were more likely to pay attention to posts during the weekends as they had more free time. Lifetime engaged users, lifetime post consumers, and lifetime people who liked the page and engaged with the post failed to make a page more appreciated. Interactions generally did little change to the popularity of a page, despite more comments tending to make the page more popular.

Types and categories, by contrast, had relatively significant influence on the page total likes. Compared to photos, status and videos made the average page total likes 5.27% (1.00% – 9.98%) and 11.10 (1.07% – 21.02%) higher, whilst links lowered the average page total likes by 6.25% (1.03% – 9.80%). For categories, however, product increased the average page total likes by 2.67% (-0.70% – 6.13%) compared to action while inspiration made the number drop by 3.58% (0.86% – 5.89%). Therefore, it can be concluded that the users preferred product showed in the form of videos most, which could possibly explain by the fact that products were the things they concerned most about a company and that videos are more vivid and easy to comprehend. As I mentioned in the previous sections, such user preferences would be of great help for designers to plan a more appealing page.

There are some further analyses of this dataset to be done if given more time. The following are the 3 possible extensions I have come up with.

Predictions of page total likes can be made through the posterior predictive distribution of the parameters, which can be derived through Monte Carlo sampling from the posterior draws

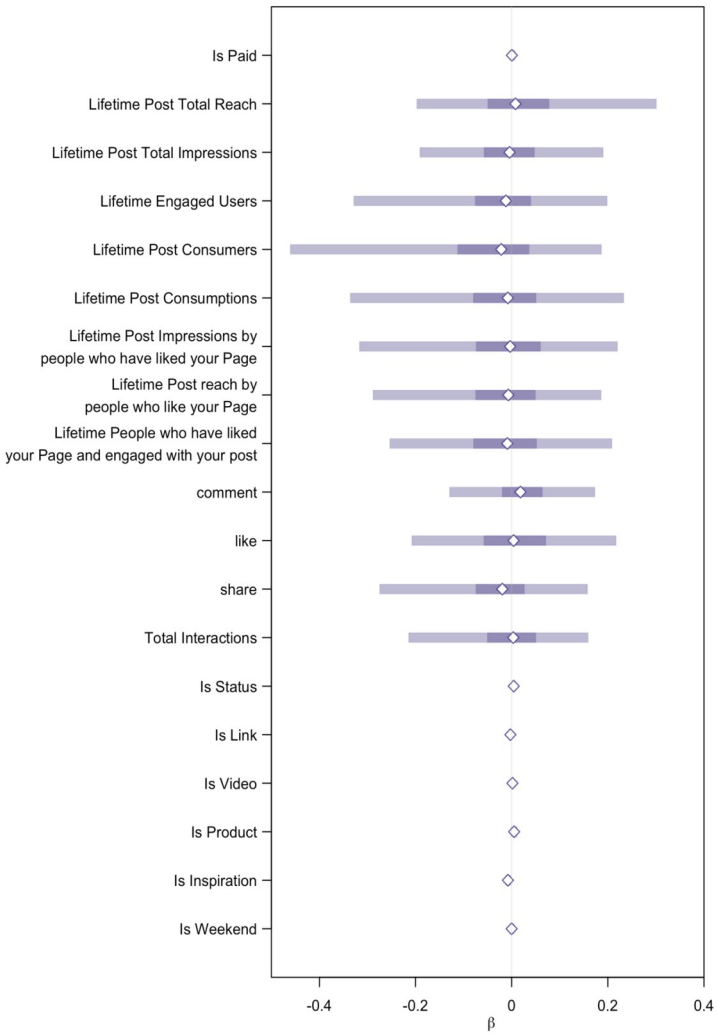

Figure 3. Estimates of different β coefficients. Diamonds are the posterior medians. 50% and 95% credible intervals are darker and lighter blue rectangles respectively.

of the parameters. To be more specific, for a page with values of the p different factors, $\boldsymbol{x} = \left(x_1, x_2, \cdots x_p\right)$, available, the mean and variance of the Negative Binomial distribution from the draw n can be presented as

$$\mu^{(n)} = \exp\left(\alpha^{(n)} + \boldsymbol{x}\boldsymbol{\beta}^{(n)}\right)$$

and

$$\sigma^{2(n)} = \tau^{(n)}\mu^{(n)}.$$

Then we randomly sample from this Negative Binomial distribution many times (at least 100) and repeat this process for each draw after burn-in. Finally, we combine all the samples to get the posterior predictive distribution of the page total likes for this post. Predictions of this kind are meaningful in that it helps forecast the popularity of a page and hence estimate

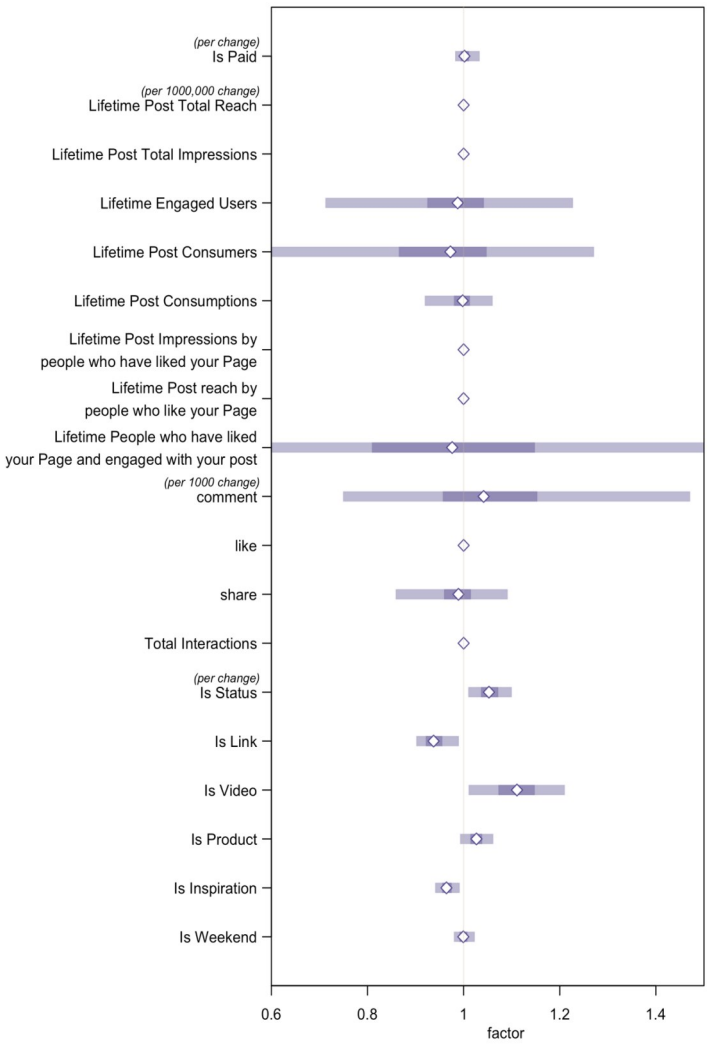

Figure 4. Impacts of different covariates on the page total likes after scaling. Diamonds are the posterior medians. 50% and 95% credible intervals are darker and lighter blue rectangles respectively.

its impact on the potential costomers.

Validation can be done in the same way as predictions, where we separate the 500 samples into 2 groups: training set and testing set. Samples in the training set are used to derive the posterior draws of the parameters, which are then used to calculate the posterior predictive distributions of the page total likes of the posts in the testing set. We compare the true values with the point estimates (such as posterior means, posterior medians, and posterior modes) as well as the credible intervals to evaluate how well the model is fit.

Models involving other discrete distributions, such as Poisson, can also be constructed to compare with the current model. Theoretically, allowing a smaller variance for the page total likes will not make the fitting as good as the current version, but a test on the dataset will better substantiate this claim. In addition, since the page total likes are generally large in number and the variance between different samples are relatively small, a continuous

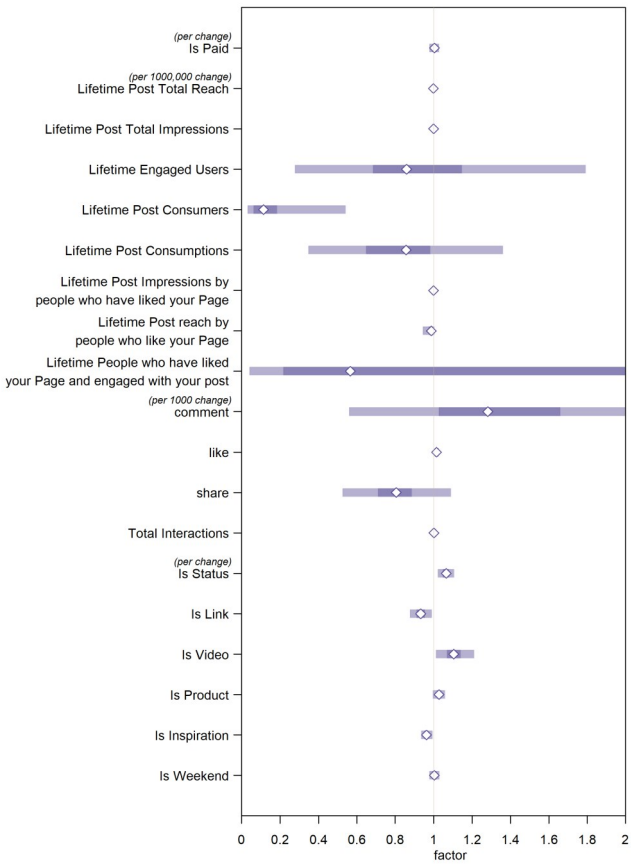

Figure 5. Impacts of different covariates on the page total likes after scaling in the scenario when the penalty λ was set to be 1. Diamonds are the posterior medians. 50% and 95% credible intervals are darker and lighter blue rectangles respectively.

distribution like Normal or Gamma distribution may as well be included in the model. Approaches are similar, as we focus the regression on the mean and variance of page total likes. The only extra work is to recalculate the parameters required for each distribution from the given formulae regarding mean and variance.

The method I derived in this study has the following merits. First, regression results are convergent, which has already been demonstrated in the previous result section. Second, by introducing the parameter τ to inflate the variance, I allowed the total page likes to deviate more from the mean value, making the model more flexible and durable to outliers. Third, the Bayesian Lasso method I introduced in this study generated desirable results with respect to variable selection, where the posterior medians are similar to the estimates given by the classic Lasso. The Laplace prior distribution assigned for each $\beta_i, i = 1,2,\cdots p$, enables most of the $\beta_i{}'$ s to concentrate around zero, thus realizing the variable selection function. Furthermore, compared to the method utilizing gprior introduced in class, the Bayesian Lasso saves the trouble of introducing another random vector $z = \left(z_1,z_2,\cdots z_p\right)^T$, where $z_i's$ are binary random variables with an independent and identical prior Binomial distribution $Binom(1,p)$, $0< p < 1$. Lastly, the model I constructed is easy to understand and the code is fast to run. Despite closed forms of marginal

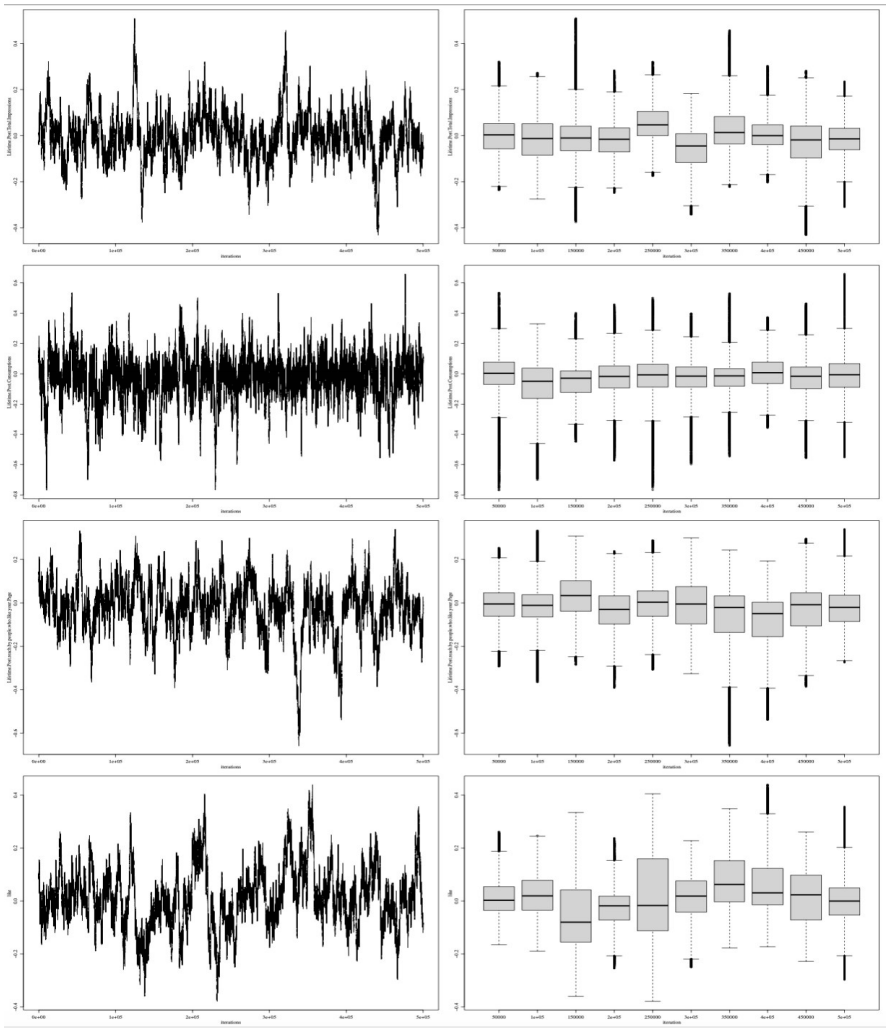

Figure 6. Assessment of convergence. Up: Traceplot of 4 different coefficients for each iteration; Down: The corresponding stationary plot of the 4 coefficients for all the iterations.

posterior distributions cannot be deduced, the standard to accept or reject a new proposal is provided by the Metropolis-Hastings algorithm and the threshold is computable.

As a matter of fact, this approach can also be generalized to other datasets. The negative binomial model I established in this study could be readily applied to data with non-negative integer valued responses and multiple predictors, such as the number of products a salesman sells or the daily case counts of a certain disease. For continuous responses, similar models with other distributions like gamma or log-normal can be applied and parameters can still be expressed via mean and variance. The Bayesian Lasso still remains a powerful method to select variables as long as the number of factors to be analyzed is large enough and the covariate matrix X satisfies the consistency property

$$\frac{1}{n} X^T X \to C$$

as the number of samples n goes to infinity, where C is some positive definite matrix. Interpretation of the outputs are also similar to what I did in this study.

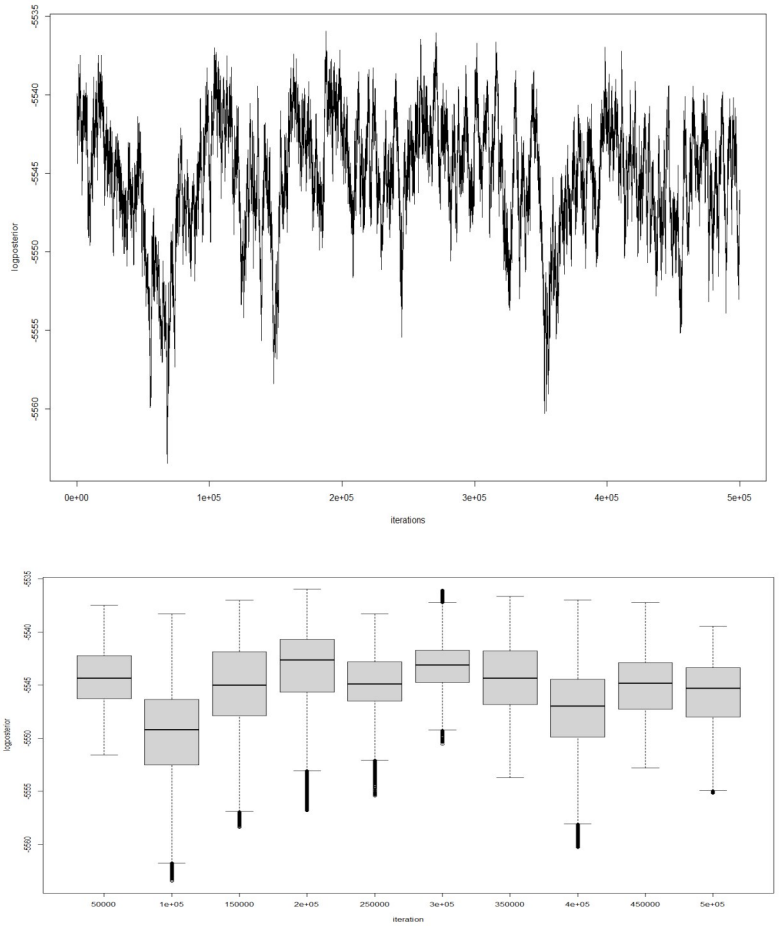

Figure 7. Assessment of convergence. Up: Traceplot of the log−posterior value for each iteration; Down: Stationary plot of the log−posterior values for all the iterations.

Limitations of this study include the static assumption of the users' responses at different times during a day or in different months throughout the year. Quantifying these time-related variables would be complicated and results would be hard to interpret. I made no prior assumption about the collinearity of some of the factors and instead analyzed their individual effects, which might be not so close to reality as results are usually a complex interplay of various factors. Besides, page likes might be subject to other factors that were not included in this dataset and the model I built might be too simple to explain all the causes, for mankind's' behaviors are often unpredictable, making the variance in estimation results large. Lastly, the limited time also made it difficult for me to perform cross-validation on each possible λ's involved in the Bayesian Lasso, thus λ = 10 might not be the optimal value of Lasso penalty, but I did choose one that allowed variable selection without making the regressed values too deviated from the real ones.

References

[1] Facebook MAU worldwide 2021. Statista . https://www. statista. com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/.

[2] Moro, S., Rita, P. & Vala, B. Predicting social media performance metrics and evaluation of the impact on brand building: A data mining approach. J. Bus. Res. 69, 3341–3351 (2016).
[3] Park, T. & Casella, G. The Bayesian Lasso. J. Am. Stat. Assoc. 103, 681–686 (2008).