

# **A Bayesian logistic regression model for Credit Default Model of Small and Medium Business Firms**

김소정

# □ Index

- Purpose of the Study
- Theoretical Background
- Description of data
- Analysis
- Conclusion
- Discussions

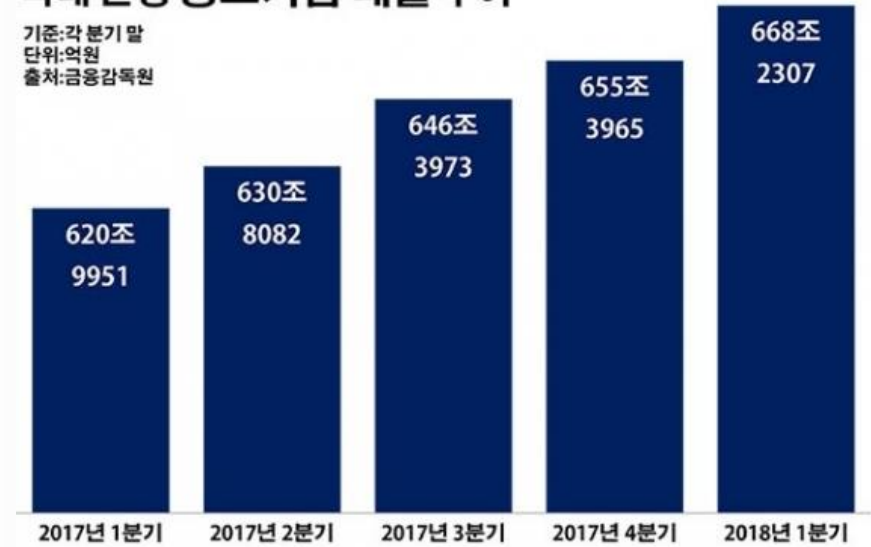
## □ Purpose of the Study

- 국내 기업 회사채 현황과 중소기업 신용평가에 대한 문제



**국내은행 중소기업 대출 추이**

기준: 각 분기 말  
단위: 억원  
출처: 금융감독원



- 기업대출이 지속적으로 증가-> 지역 경제 활성화를 위해 중소기업의 육성이 필요한 상황, 중소기업 대출의 확대 및 활성화 필요
- 중소기업의 신용대출 확대에 있어 가장 중요한 요인 : 신용위험의 최소화 -> 부도예측모델링의 정교화 필요
- 중소기업에 대한 현 신용평가 체제의 문제점 : 신용정보의 문제로 객관성과 정확성이 부족하여 정보의 획득, 축적, 평가가 어려움
  - > 중소기업의 특성에 맞는 독자적이고 신용평가 모형 빈약

# □ Background

1. 회사채 : 기업이 일시적으로 거액의 자금 조달을 목적으로 발행하며, 액면 금액, 만기, 이자율 등이 명시되어 있는 일종의 차용증서로, 이 회사채를 발행한 기업이 만기에 원리금을 지급할 수 있는지를 평가하는 것이 신용평가이다.
2. Literature review : <기업신용평가 관련 연구>
  - 회사채의 신용을 평가하는 행위는 채무자와 채권자 간의 정보비대칭을 완화하고 투자자의 정보탐색비용을 낮추는 등 금융시장의 효율성을 높인다는 점에서 중요성을 찾을 수 있다. (강경순, 이준서, 2011)
  - Altman\*의 부도예측모형 연구를 시작으로 평가자 주관에 의한 오류 가능성을 최소화하고 평가과정을 표준화 객관화하기 위해 통계 방법론에 의한 신용평가모형 개발이 본격화(Altman, 1968)
  - 회사채 신용등급 예측 연구에는 주로 해당 기업의 재무 지표를 이용한 연구가 주로 이루어져 왔으며, 재무적인 지표를 포함하여 비재무지표(김진선, 최영문, 2006), 기업의 사회적 책임활동 지수(박영규, 2015) 등을 변수로 하여 신용등급을 예측하는 연구가 진행되어 옴

# □ Description of data

## • 데이터 설명 – 변수

### 1. 데이터

- 출처 : 중소기업통합관리시스템(SIMS)  
부처·지자체 중소기업지원사업의 일환으로 정보 및 지원이력을 통합 관리하는 중소기업통합관리시스템
- NICE data 중 SIMS 에서 관리하고 있는 기업 중 비금융업종에 속한 기업
- Data size : 20만개로 구성되어 있으나 랜덤 샘플링을 통해 약 5000개의 data 를 추출함
- 결산 기준 : 2008년 3월 ~ 2017년 말

### 1. 변수

- 부도기준 : 결산년도, 부도기준일자 기준으로 부도관측기간 내 부도관측일자 존재 여부에 따라 '정상', '부도' 로 선정
- 설명변수 : 총 16개의 변수 선정  
논문(이건창 외 2명, 1996; 김성진, 안현철, 2016; 김진선, 최영문, 2006; 김성태 외 2명, 2006; 김태정 외 2명, 2003)과 한국기업평가 보고서와 한국신용정보원을 참고하여 언급된 변수들로 선정

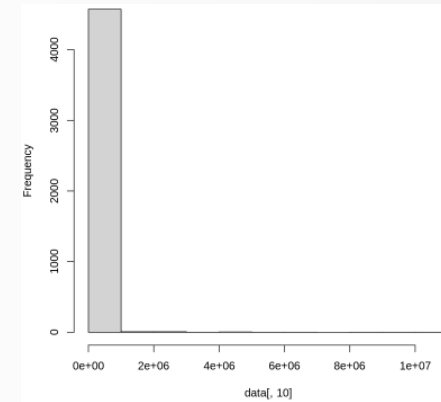
| 카테고리 | 변수명                      | 상세설명                       |
|------|--------------------------|----------------------------|
| 수익성  | EBIT/매출액 (%) (KK050)     | EBIT/매출액                   |
|      | 매출액순이익률 (%) (KK043)      | 당기순이익/매출액                  |
|      | 이자보상비율 (%) (KK103)       | (총금융비용+세전이익)/총금융비용         |
|      | ROA (%) (KK106)          | 당기순이익/자산총계                 |
|      | ROE (%) (KK039)          | 당기순이익/자기자본                 |
| 현금흐름 | 순차입금/EBITDA (배) (KK018)  | 순차입금/EBITDA                |
|      | EBIT/총금융비용 (배) (KK016)   | EBIT/총금융비용                 |
|      | EBITDA/총금융비용 (배) (KK015) | EBITDA/총금융비용               |
| 안정성  | 부채비율 (%) (KK057)         | 부채총계/자본총계                  |
|      | 단기차입금의존도 (%) (KK112)     | 단기성차입금/자산총계                |
|      | 차입금의존도 (%) (KK060)       | 총차입금/자산총계                  |
|      | 부채상환계수 (%) (KK022)       | (OCF+총금융비용)/(단기성차입금+총금융비용) |
| 성장성  | 매출액증가율 (%) (KK030)       | (당기매출액-전기매출액)/전기매출액        |
|      | 총자산증가율 (%) (KK025)       | (기말자산총계-기초자산총계)/기초자산총계     |
| 활동성  | 총자산회전율 (KK137)           | 매출액/자산총계                   |
|      | 재고자산회전율 (KK081)          | 매출액/재고자산                   |

# □ Description of data

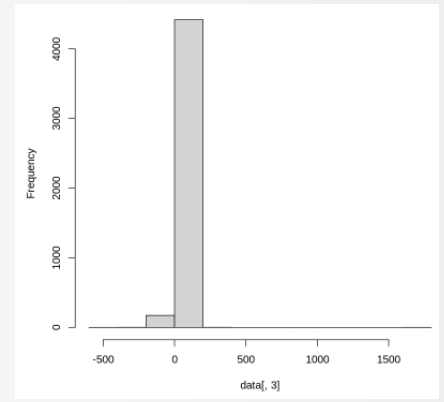
- 데이터 설명 – 변수 기초통계 및 히스토그램

|        | EBIT/매출액 (%) | 매출액순이익률 (%)  | 이자보상비율 (%) | ROA (%)    | ROE (%)    | 순차입금/EBITDA (배) | EBIT/총금융비용 (배) | EBITDA/총금융비용 (배) |
|--------|--------------|--------------|------------|------------|------------|-----------------|----------------|------------------|
| Min    | -286.53      | -434.5       | -523       | -169.56    | -592676    | -1975.67        | -51033         | -51033           |
| Median | 8.49         | 6.2          | 4.68       | 8.6        | 30.9       | 11.1            | 635            | 635              |
| Mean   | 1.08         | 7.28         | 7.48       | 1.34       | 1.21       | 1.29            | 1.06           | 1.06             |
| Max    | 2038         | 1716         | 140108     | 571        | 5521       | 96.6            | 15187000       | 15187000         |
| SD     | 3.35         | 0.9          | 2.24       | 2.43       | 2.29       | 3.48            | 3.16           | 3.16             |
|        | 부채비율 (%)     | 단기차입금의존도 (%) | 차입금의존도 (%) | 부채상환계수 (%) | 매출액증가율 (%) | 총자산증가율 (%)      | 총자산회전율         | 재고자산회전율          |
| Min    | 0            | 0            | 0          | -6860      | -96        | -95.9           | 0.00898        | 0.096            |
| Median | 238          | 4.46         | 54.6       | 172        | 2.95       | 4.07            | 1.41           | 45.4             |
| Mean   | 2.43         | 1.78         | 5.28       | 8.81       | 4.22       | 4.11            | 2.14           | 2.53             |
| Max    | 10019394     | 488          | 1489       | 16662800   | 43734      | 16449           | 55.2           | 17769439         |
| SD     | 3.18         | 2.6          | 3.75       | 3.18       | 7.8        | 3.03            | 2.54           | 9.49             |

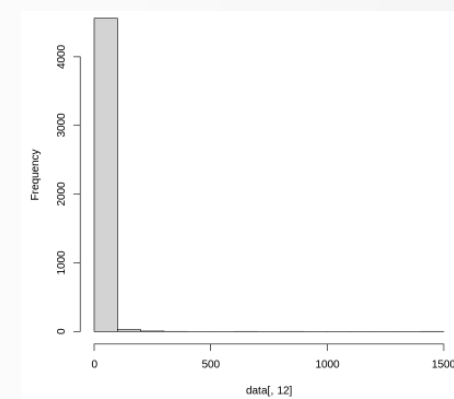
<부채비율>



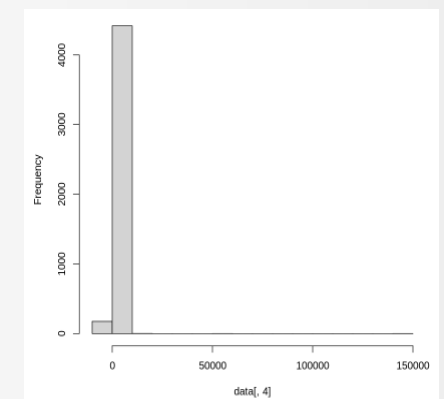
<매출액순이익률>



<차입금의존도>

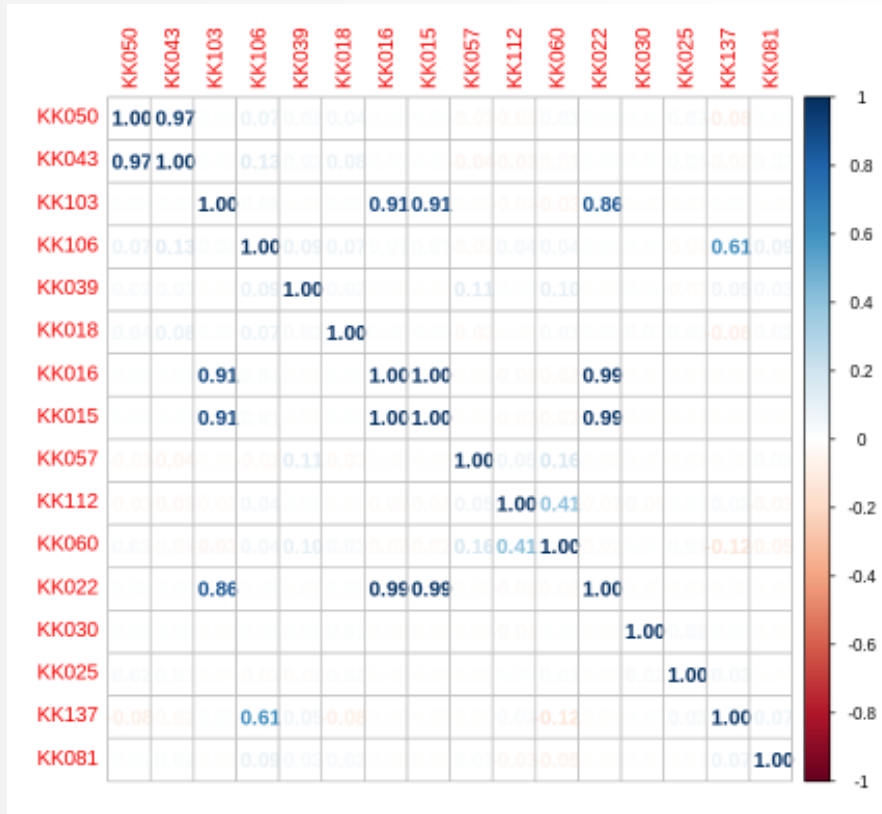


<이자보상비율>



# □ Description of data

- 데이터 설명 – 분석(상관관계)



## 1. 상관관계

- EBIT/매출액, 매출액순이익률 (양)  
-> 매출액순이익률 채택
- 이자보상비율, EBIT/총금융비용, EBITDA/총금융비용, 부채상환계수 (양)  
-> EBITDA/총금융비용 채택

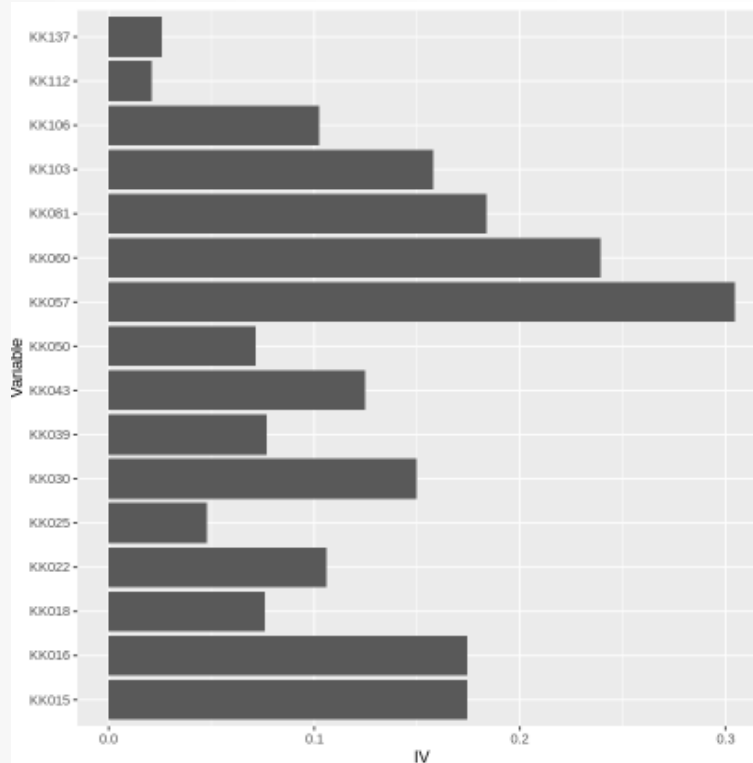
※ 논문(이건창 외 2명, 1996; 김성진, 안현철, 2016; 김진선, 최영문, 2006; 김성태 외 2명, 2006; 김태정 외 2명, 2003) 과 한국기업평가 보고서와 한국신용정보원에서 더 많이 언급된 변수들을 사용함

# □ Description of data

- 변수 선택 – Information Value(IV)

A data.frame: 16 × 2

|    | Variable | IV         |
|----|----------|------------|
|    | <chr>    | <dbl>      |
| 9  | KK057    | 0.30452505 |
| 11 | KK060    | 0.23947331 |
| 16 | KK081    | 0.18363346 |
| 7  | KK016    | 0.17455499 |
| 8  | KK015    | 0.17455499 |
| 3  | KK103    | 0.15762417 |
| 13 | KK030    | 0.14961917 |
| 2  | KK043    | 0.12482813 |
| 12 | KK022    | 0.10577673 |
| 4  | KK106    | 0.10225698 |
| 5  | KK039    | 0.07705543 |
| 6  | KK018    | 0.07589843 |
| 1  | KK050    | 0.07161973 |
| 14 | KK025    | 0.04759279 |
| 15 | KK137    | 0.02596700 |
| 10 | KK112    | 0.02109764 |



- WOE(Weights of Evidence)  
=  $\ln(\% \text{ of Good} / \% \text{ of Bads})$

-> Dist Good/Bad는 개별 등급 구간에서 연체건, 비 연체 건이 전체 총합계 비율로 얼마인지를 보는 것

- $IV = \sum_{i=1}^k (Dist\ Good_i - Dist\ Bad_i) * WOE_i$

-> 각 변수에서 개별 등급별로 계산된 WOE로 IV를 계산한 후 모든 구간의 IV를 더하면 Information value가 된다.

<0.02: unpredictive

0.02–0.1: weakly predictive

0.1–0.3: moderately predictive

+0.3: strongly predictive

- 0.02값에 가까운 총자산회전율, 단기차입금의존도 변수 제거

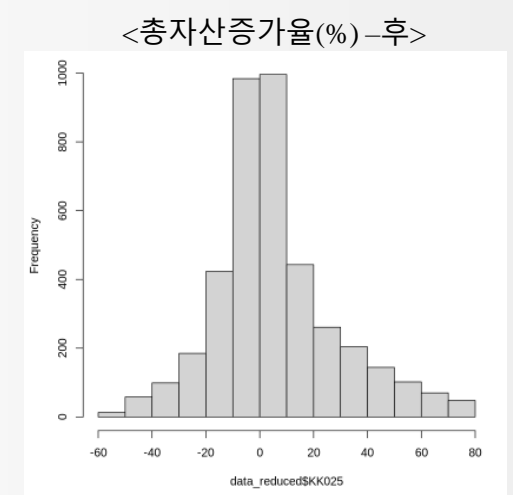
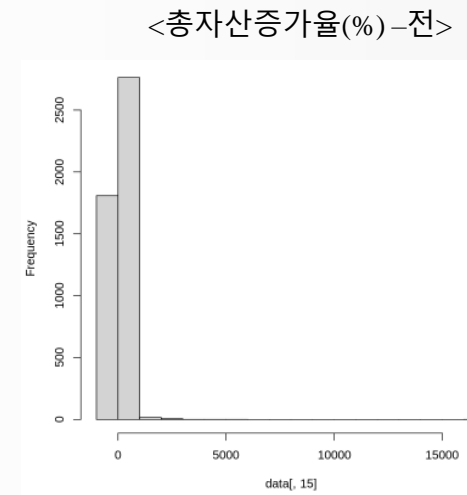
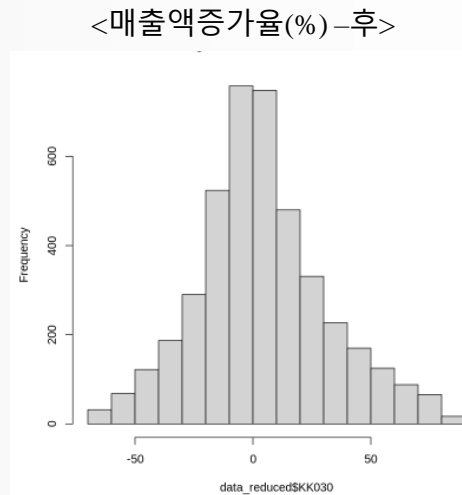
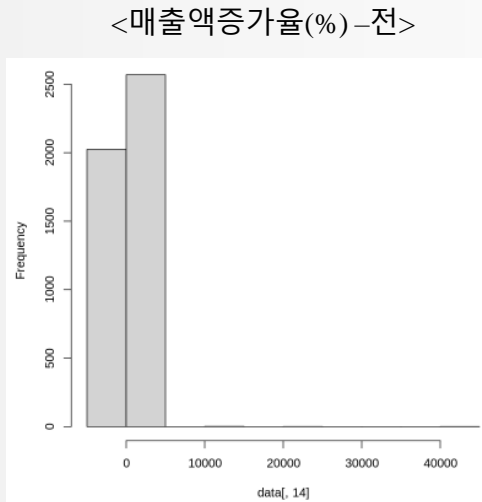


# □ Description of data

- 최종 변수

|         |        |        |             |              |
|---------|--------|--------|-------------|--------------|
| KK043   | KK106  | KK039  | KK018       | KK015        |
| 매출액순이익률 | ROA    | ROE    | 순차입금/EBITDA | EBITDA/총금융비용 |
| KK057   | KK060  | KK030  | KK025       | KK081        |
| 부채비율    | 차입금의존도 | 매출액증가율 | 총자산증가율      | 재고자산회전율      |

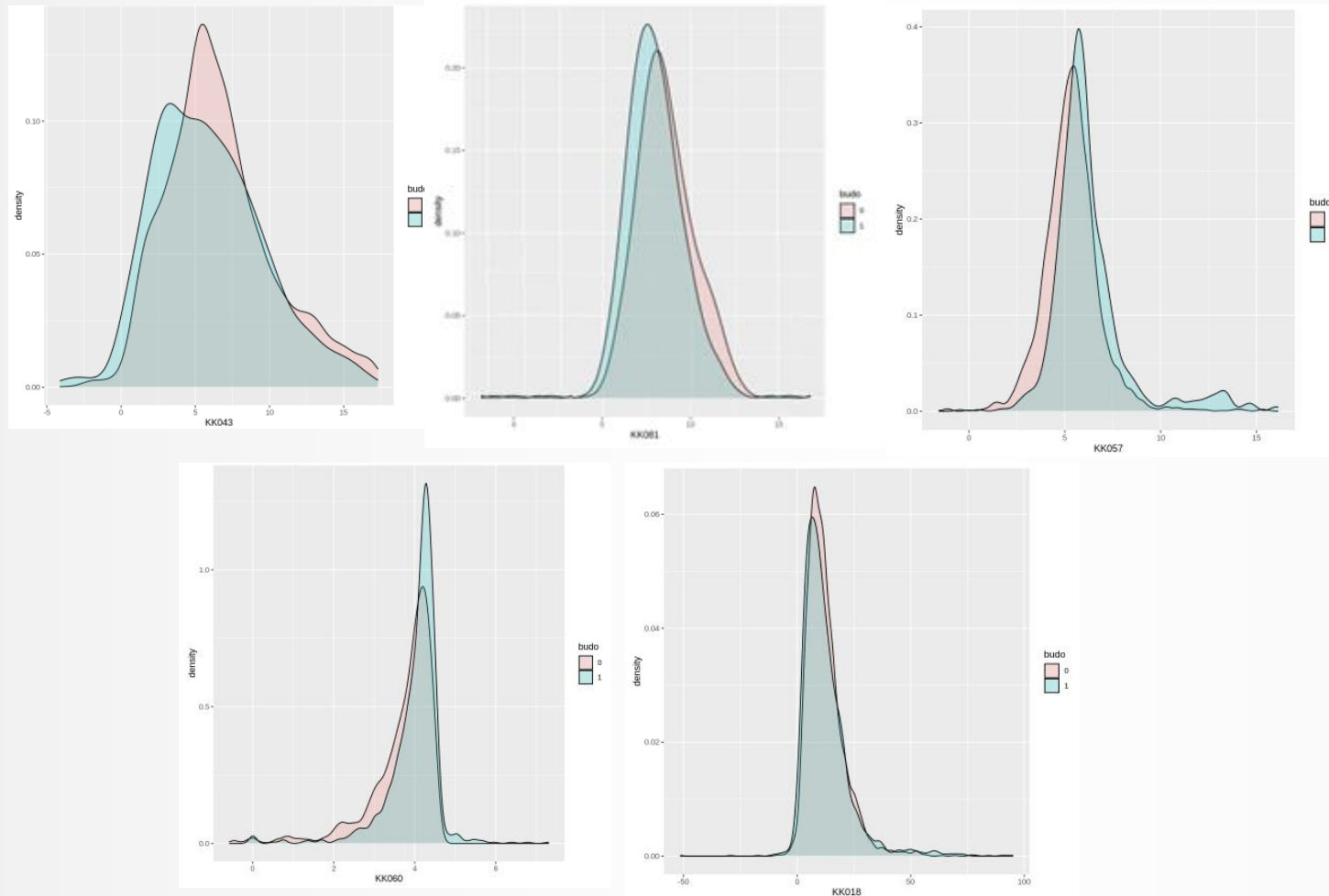
- 이상치 제거



- 예외적으로 이상치 제거 후에도 비대칭적인 분포 모양을 가지고 있는 변수 -> log 변환을 실시함
- 이상치 제거 후 변수 : 4119개 , NA : 478개

# □ Description of data

## • 데이터 설명 – 분석(EDA)



매출액순이익률이 낮을수록,  
재고자산회전율이 낮을수록,  
부채비율이 높을수록,  
차입금의존도가 높을수록  
순차입금/EBITDA(배)가 높을수록

기업 부도에 대한 특징이라고 볼 수 있음

# □ Analysis

- 분석 – 로지스틱 회귀분석과 베이지안 로지스틱 회귀분석 비교

## 1. 모델링 : 로지스틱 회귀모형

<추정된 회귀계수에 대한 요약 통계>

|                     |            |            |         |              |
|---------------------|------------|------------|---------|--------------|
| Deviance Residuals: |            |            |         |              |
| Min                 | 1Q         | Median     | 3Q      | Max          |
| -1.8908             | -0.7597    | -0.6091    | -0.3470 | 2.8569       |
| Coefficients:       |            |            |         |              |
|                     | Estimate   | Std. Error | z value | Pr(> z )     |
| (Intercept)         | -3.312e+00 | 2.672e-01  | -12.394 | < 2e-16 ***  |
| KK043               | -3.633e-02 | 1.345e-02  | -2.701  | 0.006905 **  |
| KK106               | 1.947e-03  | 1.901e-03  | 1.024   | 0.305838     |
| KK039               | -3.121e-06 | 2.035e-06  | -1.534  | 0.125109     |
| KK018               | 4.141e-03  | 4.991e-03  | 0.830   | 0.406678     |
| KK015               | -1.232e-06 | 2.107e-06  | -0.585  | 0.558815     |
| KK057               | 2.561e-01  | 2.876e-02  | 8.903   | < 2e-16 ***  |
| KK060               | 2.772e-01  | 7.537e-02  | 3.678   | 0.000235 *** |
| KK030               | -1.330e-05 | 1.216e-04  | -0.109  | 0.912898     |
| KK025               | 3.802e-02  | 2.206e-02  | 1.723   | 0.084885 .   |
| KK081               | -6.163e-02 | 1.080e-02  | -5.708  | 1.14e-08 *** |

<Odds 와 CI>

|             | OR      | 2.50%   | 97.50%  | P-value |
|-------------|---------|---------|---------|---------|
| (Intercept) | 0.03646 | 0.02055 | 0.06160 | 0.00000 |
| KK043       | 0.96432 | 0.94136 | 0.98987 | 0.00691 |
| KK106       | 1.00195 | 0.99791 | 1.00594 | 0.30584 |
| KK039       | 1.00000 | 0.99999 | 1.00000 | 0.12511 |
| KK018       | 1.00415 | 0.99474 | 1.01328 | 0.40668 |
| KK015       | 1.00000 | 0.99999 | 1.00000 | 0.55881 |
| KK057       | 1.29183 | 1.21898 | 1.37832 | 0.00000 |
| KK060       | 1.31947 | 1.12989 | 1.53602 | 0.00023 |
| KK030       | 0.99999 | 0.99950 | 1.00018 | 0.91290 |
| KK025       | 1.03802 | 1.00000 | 1.08248 | 0.08489 |
| KK081       | 0.94024 | 0.92020 | 0.95982 | 0.00000 |

- $\log \frac{\theta}{1-\theta} = -0.04x_1 + 0.002x_2 - 0.000003x_3 + 0.004x_4 - 0.000001x_5 + 0.2561x_6 + 0.2772x_7 - 0.00001x_8 + 0.04x_9 - 0.06x_{10}$
- 매출액순이익률(KK043), 부채비율(KK057), 차입금의존도(KK060), 재고자산회전율(KK081) 이 Credit default 를 설명하는 데 있어 유의한 변수임을 확인

# □ Analysis

- 분석 – 로지스틱 회귀분석과 베이지안 로지스틱 회귀분석 비교

## 2. 모델링 : 베이지안 로지스틱 회귀모형

<추정된 회귀계수에 대한 요약 통계>

|             | Mean       | SD        | 2.5%       | 97.5%      |
|-------------|------------|-----------|------------|------------|
| (Intercept) | -3.302e+00 | 2.732e-01 | -3.885e+00 | -2.787e+00 |
| KK043       | -3.633e-02 | 1.315e-02 | -6.043e-02 | -1.018e-02 |
| KK106       | 1.965e-03  | 2.040e-03 | -2.091e-03 | 5.925e-03  |
| KK039       | -3.260e-06 | 2.196e-06 | -7.982e-06 | 8.180e-07  |
| KK018       | 3.891e-03  | 4.783e-03 | -5.273e-03 | 1.319e-02  |
| KK015       | -3.395e-06 | 2.713e-06 | -9.662e-06 | -9.794e-08 |
| KK057       | 2.582e-01  | 2.963e-02 | 1.980e-01  | 3.209e-01  |
| KK060       | 2.720e-01  | 7.742e-02 | 1.221e-01  | 4.292e-01  |
| KK030       | -8.377e-05 | 1.754e-04 | -5.019e-04 | 1.782e-04  |
| KK025       | 3.985e-02  | 2.185e-02 | -2.230e-02 | 8.244e-02  |
| KK081       | -6.178e-02 | 1.097e-02 | -8.317e-02 | -4.101e-02 |

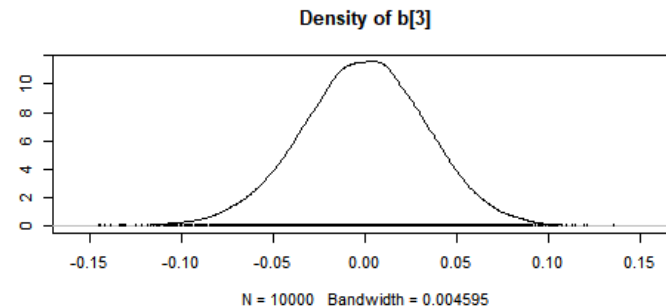
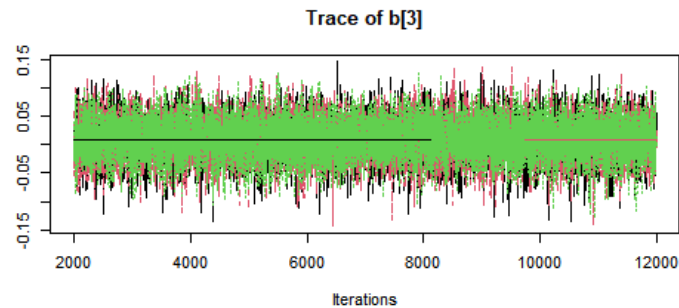
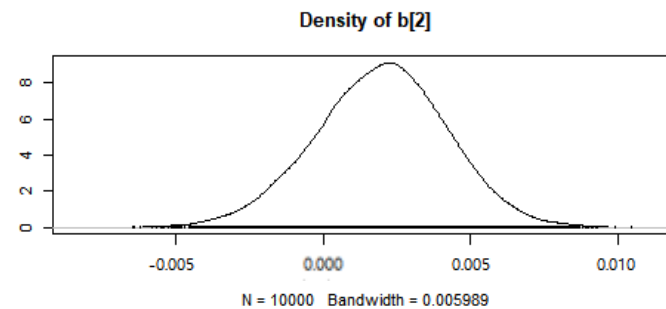
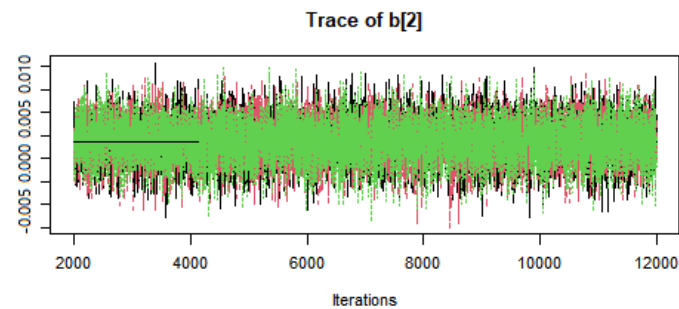
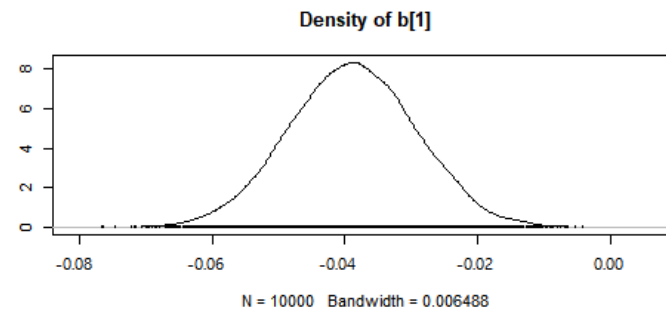
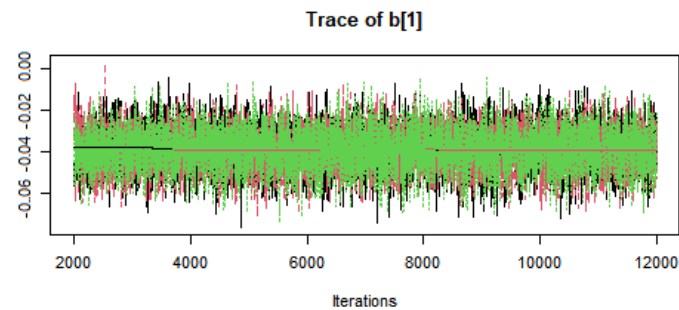
<Odds 와 CI>

|             | OR      | 2.50%   | 97.50%  |
|-------------|---------|---------|---------|
| (Intercept) | 0.03681 | 0.02132 | 0.06082 |
| KK043       | 0.96432 | 0.93920 | 0.99007 |
| KK106       | 1.00197 | 0.99811 | 1.00582 |
| KK039       | 1.00000 | 0.99999 | 1.00000 |
| KK018       | 1.00390 | 0.99421 | 1.01389 |
| KK015       | 1.00000 | 0.99999 | 1.00000 |
| KK057       | 1.29460 | 1.22181 | 1.36770 |
| KK060       | 1.31259 | 1.14218 | 1.53508 |
| KK030       | 0.99992 | 0.99956 | 1.00018 |
| KK025       | 1.04065 | 1.00000 | 1.08168 |
| KK081       | 0.94009 | 0.92027 | 0.96008 |

- 각 모수의 추정치와 95% 신용구간(credible interval)을 보면, 대부분의 설명변수의 회귀계수의 95% 신용구간이 0을 포함하지 않음으로써 credit default에 영향을 미치고 있다는 것을 알 수 있다.
- 차입금의존도(KK060)의 credit default 에 대한 odds가 1.312(=exp(0.272))을 나타내고 있는데, 이는 차입금의존도가 credit default 에 끼치는 영향이 (오즈(odds)) 1.312배 가량 높다는 것을 의미한다.
- 가장 높은 odds를 보이는 설명변수는 차입금의존도(KK060), 부채비율(KK057), ROE(KK039) 순

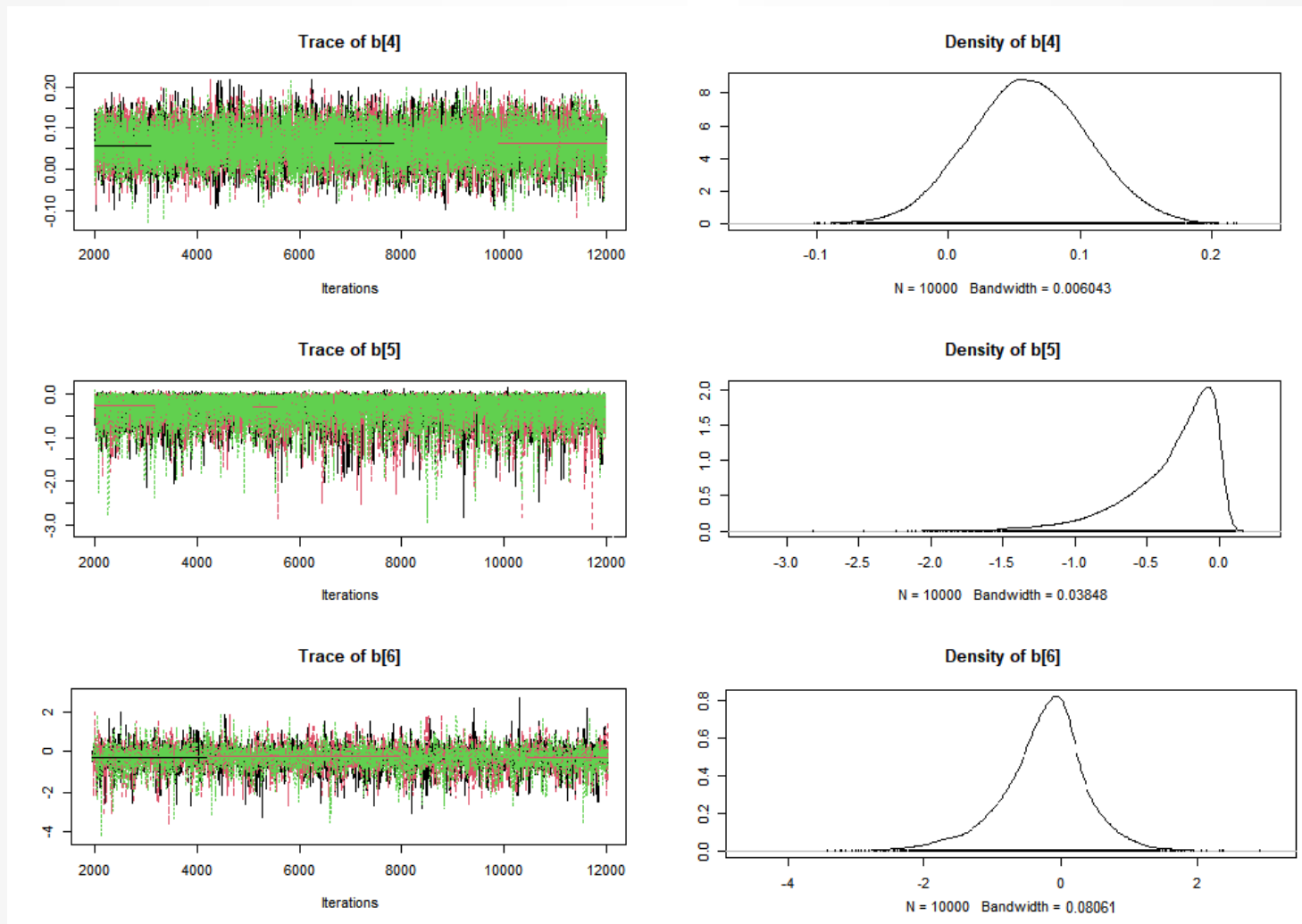
# □ Analysis

- **MCMC로 추출한 모수의 사후분포 도표** MCMC를 이용하여 10000번의 반복, 초기 2000번 제거(burn-in)



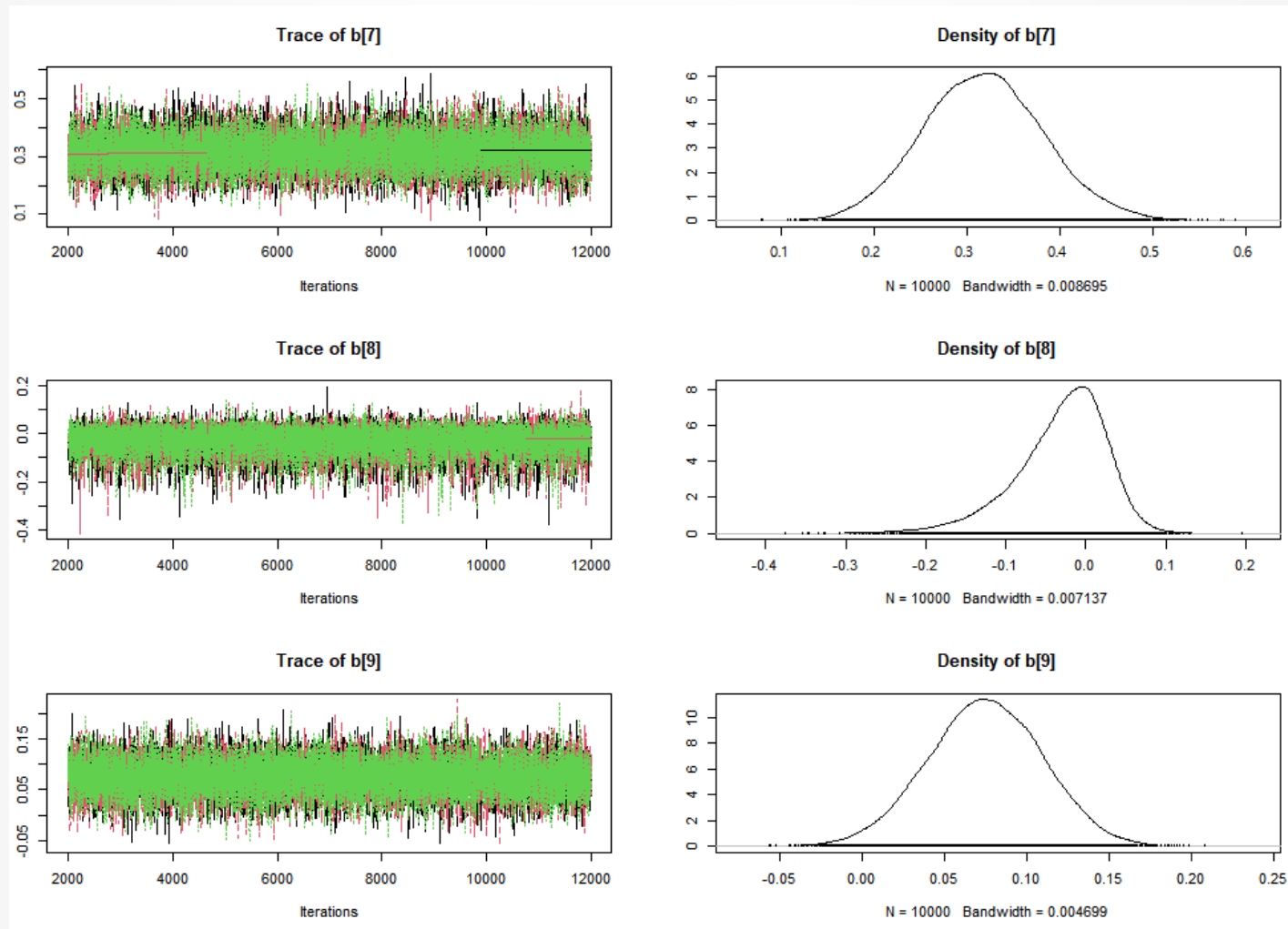
# □ Analysis

- MCMC로 추출한 모수의 사후분포 도표



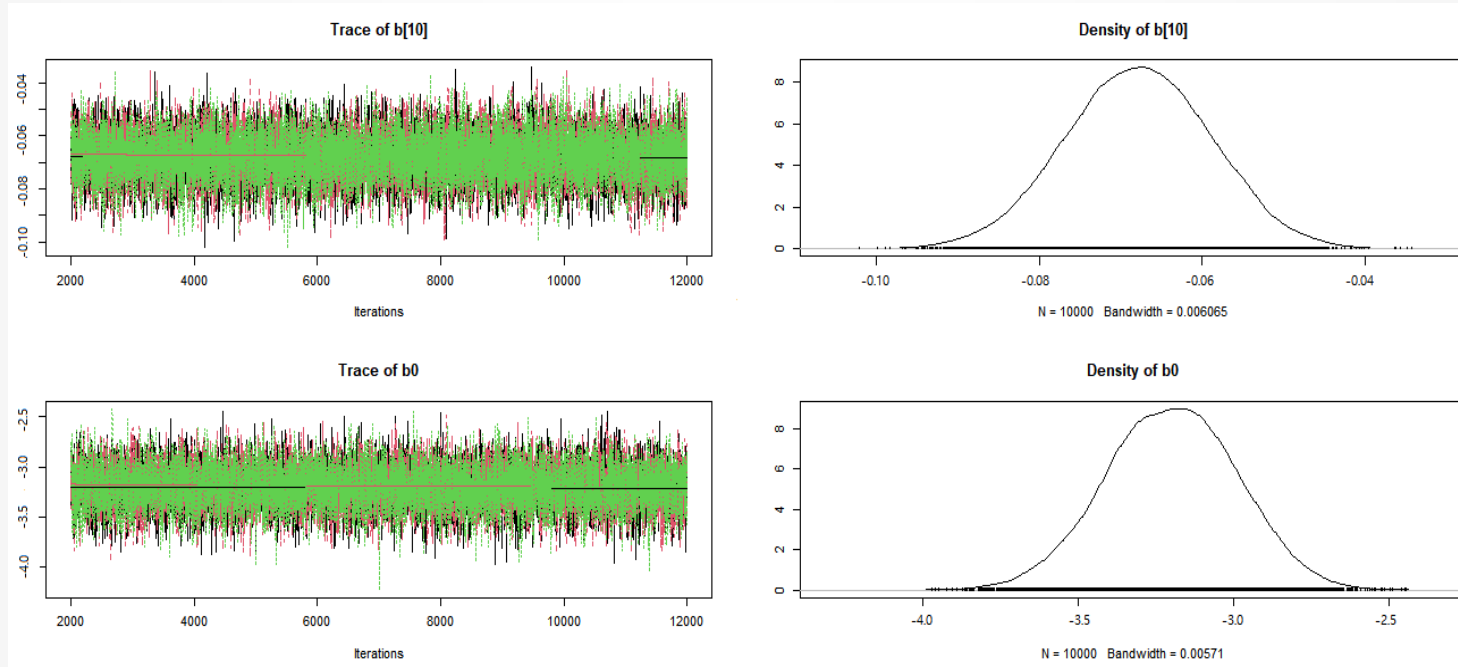
# □ Analysis

- MCMC로 추출한 모수의 사후분포 도표



# □ Analysis

- MCMC로 추출한 모수의 사후분포 도표





# □ Analysis

- MCMC로 추출한 모수의 사후분포 도표
- 모든 마르코프 체인이 추세(Trend)를 보이지 않으며 수렴하고 있음

Potential scale reduction factors:

|       | Point est. | Upper C.I. |
|-------|------------|------------|
| b[1]  | 1          | 1          |
| b[2]  | 1          | 1          |
| b[3]  | 1          | 1          |
| b[4]  | 1          | 1          |
| b[5]  | 1          | 1          |
| b[6]  | 1          | 1          |
| b[7]  | 1          | 1          |
| b[8]  | 1          | 1          |
| b[9]  | 1          | 1          |
| b[10] | 1          | 1          |
| b0    | 1          | 1          |

Multivariate psrf

1

## 3. 각 모델의 결과 비교 : 변수

- 0으로부터 벗어난 형태를 보이고 있는 모수는 매출액순이익률(KK043), 순차입금/EBITDA (KK018), 차입금의존도(KK060), 총자산증가율(KK025), 재고자산회전율(KK081)이며, 그 이외의 모수는 0으로부터 크게 벗어나지 못한 것을 볼 수 있음  
-> 모수의 분포가 0으로부터 크게 벗어나지 못했다는 의미는 해당 모수의 변수가 종속변수에 크게 영향을 주지 않는 것으로 해석할 수 있음
- 로지스틱 회귀변수의 유의미한 변수와 어느정도 겹치는 것을 확인할 수 있음  
-> 매출액순이익률(KK043), 차입금의존도(KK060), 재고자산회전율(KK081)

# □ Analysis

- 분석 – 로지스틱 회귀분석과 베이지안 로지스틱 회귀분석 비교

## 3. 각 모델의 결과 비교 : C.I 비교

<각 모델 별 CI 구간 차이표>

|             | logit model |          |         | bayesian model |         |         |
|-------------|-------------|----------|---------|----------------|---------|---------|
|             | 2.50%       | 97.50%   | 구간 차이   | 2.50%          | 97.50%  | 구간 차이   |
| (Intercept) | 0.020554    | 0.061596 | 0.04104 | 0.02055        | 0.06160 | 0.03950 |
| KK043       | 0.941356    | 0.989867 | 0.04851 | 0.94136        | 0.98987 | 0.05087 |
| KK106       | 0.997911    | 1.005942 | 0.00803 | 0.99973        | 1.00101 | 0.00772 |
| KK039       | 0.999992    | 1.000001 | 0.00001 | 1.00000        | 1.00000 | 0.00001 |
| KK018       | 0.994740    | 1.013278 | 0.01854 | 0.99902        | 1.00236 | 0.01968 |
| KK015       | 0.999990    | 1.000000 | 0.00001 | 1.00000        | 1.00000 | 0.00001 |
| KK057       | 1.218983    | 1.378322 | 0.15934 | 1.21898        | 1.37832 | 0.14589 |
| KK060       | 1.129891    | 1.536023 | 0.40613 | 1.12989        | 1.53602 | 0.39290 |
| KK030       | 0.999498    | 1.000178 | 0.00068 | 0.99996        | 1.00003 | 0.00061 |
| KK025       | 1.000000    | 1.082477 | 0.08248 | 0.99998        | 1.00082 | 0.08168 |
| KK081       | 0.920198    | 0.959816 | 0.03962 | 0.92020        | 0.95982 | 0.03981 |

- 추정치들의 구간 폭에 대한 비교
  - > 크게 차이가 나진 않았지만 절반 이상 부분이 Bayesian model 의 구간이 더 좁은 것을 확인 가능
  - > 이는 일반 logit model 보다 좀 더 안전하고 Efficient한 추정치를 이용할 수 있는 것을 의미

# □ Analysis

- 분석 – 로지스틱 회귀분석과 베이지안 로지스틱 회귀분석 비교

## 3. 각 모델의 결과 비교 : 예측 결과

<로지스틱 회귀분석 모델>

Confusion Matrix and Statistics

|   | 0    | 1  |
|---|------|----|
| 0 | 3149 | 26 |
| 1 | 894  | 50 |

Accuracy : 0.7766  
95% CI : (0.7636, 0.7893)  
No Information Rate : 0.9815  
P-Value [Acc > NIR] : 1

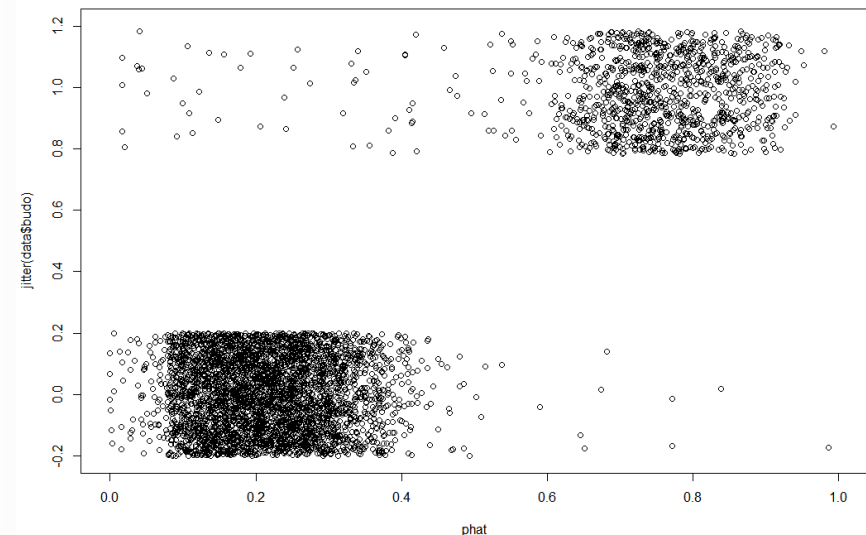
Kappa : 0.0661

McNemar's Test P-Value : <2e-16

Sensitivity : 0.77888  
Specificity : 0.65789  
Pos Pred Value : 0.99181  
Neg Pred Value : 0.05297  
Prevalence : 0.98155  
Detection Rate : 0.76451  
Detection Prevalence : 0.77082  
Balanced Accuracy : 0.71839

'Positive' Class : 0

<베이지안 로지스틱 회귀분석 모델>



실제 결과가 0인 경우의 phat은 0근처에 모여 있고, 1인 경우의 phat은 1근처에 모여 있는 것을 확인할 수 있다.

```
> sum((data$budo[phat >= 0.5] == 1), (data$budo[phat < 0.5] == 0))/nrow(data)
[1] 0.7776811
```

## □ Conclusion

- MCMC 알고리즘을 통해 추출된 각 모수의 표본값들은 trace plot을 통해 대체적으로 잘 수렴되고 있음을 확인할 수 있었고, Gelman-Rubin diagnostic  $\hat{R}$  역시 1에 근사한 값을 얻어서 표본값의 수렴에 큰 문제가 없음을 확인하였다.
- 로지스틱 회귀모형과 베이지안 모형의 결과로 최종적으로 Credit default에 유의한 영향을 주는 변수는 매출액순이익률(KK043), 차입금의존도(KK060), 재고자산회전율(KK081) 이나, 두 모형의 결과는 크게 다르지 않음을 확인하였다.
- 두 모형의 예측의 정확도 역시 비슷한 수치를 보임
- 중소기업 신용평가 중심으로 가장 큰 문제는 적은 데이터 표본 뿐만 아니라 질적인 데이터의 수도 부족하다는 것
  - 소표본 데이터에 적용한다면 베이지안 추론 방법의 장점을 보여줄 수 있는 향후 과제 중 하나가 될 것이 라는 의미를 가짐

## □ Discussions

- 두 모델 비교 방법에 대한 추가적인 검증 필요
  - > ROC 커브나 AUC 값에 대한 확인이 필요
- 각 모델링 중 모델 선택 과정을 거치지 않았다는 점
  - > 로지스틱 회귀분석에서는 AIC, 베이지안 모델에서는 DIC 비교를 통해 보완
- 비대칭 데이터의 문제
  - > 극단적으로 한쪽으로 치우쳐서 나오는 불균형적인 자료의 특성을 가지고 있기 때문에, 해당 모델 분석을 사용하게 되면 모수의 추정치가 편향될 가능성이 있음
  - > 비대칭 연결함수를 사용한 적합한 모델의 분석 필요 : Complementary log-log 연결 모형 등



**Thank You**