Predicting Credit's Default Payment

Vivian Phan
Alexandr Pimikov
Thilini Panagoda

Aalto University

Control of default risk

- Accurate prediction of default payments enables banks and credit issuers to effectively mitigate lending risks, manage their credit portfolios more efficiently, and minimize potential financial losses.
- Predictive models aid in identifying consumers at high risk of default, facilitating early intervention strategies like debt restructuring and financial counseling, which can help consumers avoid defaulting on their loans.

Introduction

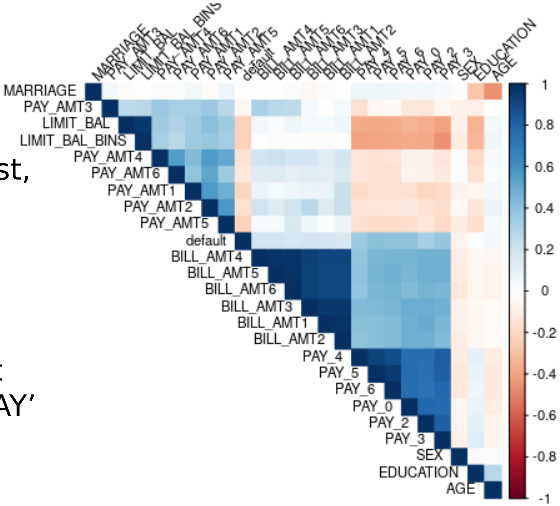
- This report presents a comparative study of two statistical models designed to predict credit default.
  - Logistic regression model
  - Hierarchical model

- Our goal is to determine whether the consumers' education affect their ability to make payments

- We examine whether the hierarchical model has any added benefit in predicting the ability of making default payments.

Description of the data and the analysis

- Relationships between different features in the dataset.
- The attribute of interest, ’default’, which represents the outcome we aim to predict
- It exhibits a significant correlation with the ’PAY’ attributes,
- We consider first two months behavior of the clients PAY_0, PAY_2, and ’EDUCATION’



Description of the data and the analysis

- Our analysis focuses on a small sampled subset of the data, according to different educational backgrounds

  1 - graduate school,

  2 - university,

  3 - high school

  4 - others

- default outcomes

  0 - no default

  1 - default.

For each category within education and default status, we sampled 40 data points

Original data fitted data

30000 * (22 + 1) -> 240 * (3 + 1)

Description of two models

- Pooled model for default d

$$\mu = \mathrm{logistic}(\beta_0 + \beta_1 * \mathrm{PAY\_0} + \beta_2 * \mathrm{PAY\_2}),$$
$$d \sim \mathrm{Bernoulli}(\mu)$$

logistic function

$$\mathrm{logistic}(x) = \frac{1}{1 + exp(-x)},$$

- Hierarchical model

has one extra layer for intercept grouped by education

$$d_i \sim \mathrm{Bernoulli}(\mu_i)$$
$$\mu_i = \mathrm{logistic}(\beta_0 + \beta_1 * \mathrm{PAY\_0} + \beta_2 * \mathrm{PAY\_2} + \alpha_i),$$
$$\alpha_i \sim \mathcal{N}(a_i, \sigma).$$

Description of two models

- Pooled model
    - 4 chains 2000 iterations and 1000 warm-up iterations

- Hierarchical model
    - 4 chains 5000 iterations and 2500 warm-up iterations

Prior Selection

- Non informative prior – $\mathcal{N}(0, 100)$

The PDF for this prior is:

$$\mathcal{N}(0, 100) : P(\theta) = \frac{1}{\sqrt{2\pi \times 100^2}} \exp\left(-\frac{(\theta - 0)^2}{2 \times 100^2}\right)$$

$$\beta_i \sim \mathcal{N}(0, 100) \qquad \text{where } i = 0, 1, 2.$$

- For the Hierarchical model additional parameters

$$a_i \sim \text{Exponential}(0.02)$$

$$\sigma \sim \text{Exponential}(0.02)$$

Prior Selection

- Weakly informative prior – $\mathcal{N}(\log(\frac{0.7}{0.3}), 1)$

The PDF for this prior is:

$$\mathcal{N}(log(\frac{0.7}{0.3}), 1) : P(\theta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(\theta - \log(\frac{0.7}{0.3}))^2}{2}\right)$$

$$\beta_i \sim \mathcal{N}(\frac{0.7}{0.3}, 1)\,,$$

$$\beta_0 \sim \mathcal{N}(0, 10)\,.$$

Table 1: Dataset descriptions

Dataset	Nof. Entries	Nof. 0 entries	Nof. 1 entries
Short Data	240 entries	120 (50%)	120 (50%)
Original Data	30000 entries	23364 (77.88%)	6636 (22.12%)

Sensitivity Analysis

Table 4: Predictive Performance of Models with prior $\mathcal{N}(0, 100)$. The last column is the noninformative rate, the ratio of zeros to number of data points.

Model	Data Type	Accuracy	95% CI	Rate
Pooled	Short Data (240 entries)	0.73	(0.668, 0.784)	0.5
Pooled	Original Data (30000 entries)	0.80	(0.806, 0.814)	0.7788
Hierarchical	Short Data (240 entries)	0.75	(0.695, 0.807)	0.5
Hierarchical	Original Data (30000 entries)	0.80	(0.795, 0.804)	0.7788

Table 5: Predictive Performance of Models with prior $\mathcal{N}(\log(\frac{0.7}{0.3}), 1)$

Model	Data Type	Accuracy	95% CI	Rate
Pooled	Short Data (240 entries)	0.73	(0.673, 0.788)	0.5
Pooled	Original Data (30000 entries)	0.81	(0.804, 0.813)	0.7788
Hierarchical	Short Data (240 entries)	0.77	(0.712, 0.822)	0.5
Hierarchical	Original Data (30000 entries)	0.80	(0.804, 0.813)	0.7788

Rhat and ESS value for convergence

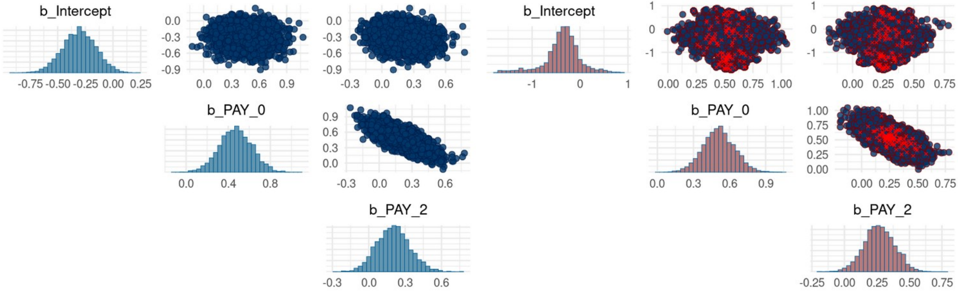
	variable	mean	median	sd	mad	q5	q95	rhat	ess_bulk	ess_tail
1	b_Intercept	-0.21	-0.21	0.15	0.15	-0.46	0.03	1.00	2592.41	2472.17
2	b_PAY_0	0.52	0.52	0.15	0.16	0.28	0.78	1.00	2058.43	2543.86
3	b_PAY_2	0.16	0.16	0.13	0.13	-0.05	0.38	1.00	2074.70	2361.45
4	lprior	-16.57	-16.57	0.00	0.00	-16.57	-16.57	1.00	2417.79	
5	lp__	-158.24	-157.93	1.21	0.98	-160.77	-156.90	1.00	1876.25	2311.99

Table 2: Pooled model

	variable	mean	median	sd	mad	q5	q95	rhat	ess_bulk	ess_tail
1	b_I	-0.42	-0.39	0.74	0.36	-1.50	0.63	1.01	505.62	232.77
2	b_PAY_0	0.22	0.23	0.17	0.16	-0.05	0.51	1.00	1088.72	281.44
3	b_PAY_2	0.59	0.59	0.15	0.15	0.34	0.84	1.01	818.61	305.30
4	sd_E__I	0.94	0.56	1.11	0.53	0.06	3.39	1.01	738.04	291.92
5	r_E[1,I]	0.32	0.21	0.77	0.39	-0.64	1.50	1.01	482.44	232.87
6	r_E[2,I]	-0.06	-0.04	0.74	0.35	-1.14	0.98	1.01	520.36	237.73
7	r_E[3,I]	-0.20	-0.16	0.76	0.36	-1.39	0.80	1.01	544.38	239.68
8	lprior	-20.50	-20.50	0.02	0.01	-20.55	-20.49	1.01	736.76	292.70
9	lp__	-158.56	-158.26	2.45	2.36	-163.03	-155.07	1.00	1481.31	1861.92

Table 3: Hierarchical model.

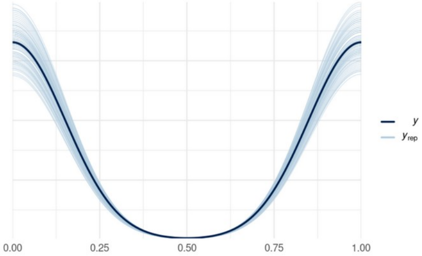
Comparative MCMC diagnostics for pooled and hierarchical models



(a) MCMC diagnostics for the pooled model **(b)** MCMC diagnostics for the hierarchical model

- Diagonal histograms-marginal posterior distributions
- Red points- divergent transitions

Posterior predictive checks- Pooled model

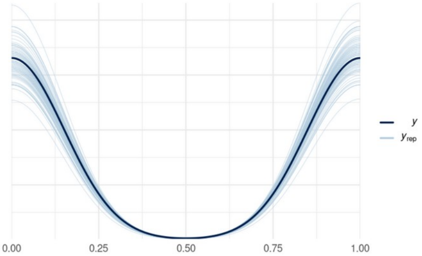


(b) Posterior Predictive Check for Predicted Probabilities

- Dark line represents-observed proportions(y)

  Lighter lines represents- densities of predicted proportions (y_rep)

Posterior predictive checks- Hierarchical model



(b) Posterior Predictive Check for Predicted Probabilities

- Hierarchical model behave similarly as pooled model

Model Comparison

Table 6: Comparison between pooled and hierarchical model

	elpd_diff	se_diff	elpd_loo	se_elpd_loo	p_loo	se_p_loo	looic	se_looic
fit2	0.00	0.00	-136.16	7.88	4.92	0.43	272.32	15.77
fit1	-7.72	12.17	-143.88	9.39	4.27	0.98	287.76	18.77

- The hierarchical model performs slightly better
- However, the standard error indicates that the elpd difference between two models are not significant.

Confusion Matrix - Stan model PAY_0 + PAY_2

```
Prediction  0  1
         0 94 36
         1 26 84

               Accuracy : 0.7417
                 95% CI : (0.6814, 0.7958)
    No Information Rate : 0.5
    P-Value [Acc > NIR] : 1.785e-14
```

```
          Reference
Prediction     0     1
         0 20637  3452
         1  2727  3184

               Accuracy : 0.794
                 95% CI : (0.7894, 0.7986)
    No Information Rate : 0.7788
    P-Value [Acc > NIR] : 7.634e-11
```

- False negative: 0.15
- False positive: 0.11

- False negative: 0.12
- False positive: 0.09

Confusion Matrix -Stan model EDUCATION Hierarchical

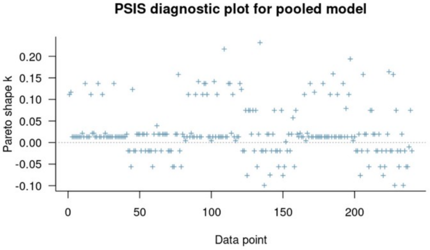
```
          Reference
Prediction  0  1
         0 95 37
         1 25 83

               Accuracy : 0.7417
                 95% CI : (0.6814, 0.795
    No Information Rate : 0.5
```

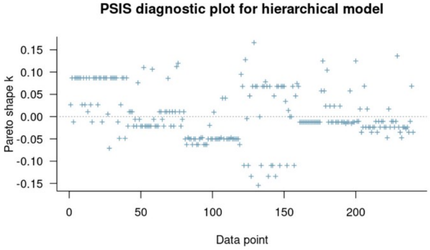
```
          Reference
Prediction     0     1
         0 20409  3372
         1  2955  3264

               Accuracy : 0.7891
                 95% CI : (0.7844, 0.7937)
    No Information Rate : 0.7788
    P-Value [Acc > NIR] : 8.003e-06
```

- False negative: 0.15
- False positive: 0.10
-
- False negative: 0.11
- False positive: 0.10

Predictive performance assessment

Pareto Smoothed Importance Sampling (PSIS) diagnostics for pooled and hierarchical models





- All the Pareto k-values seem to be well below 0.7, which is generally considered good.
- LOO diagnostics -> how well IS perform for LOO posterior.

Discussion

- Model Accuracy vs. No Information Rate:
  Both models’ accuracies do not significantly outperform the no information rate.

- Divergences in the Hierarchical Model:
  The presence of divergent transitions in the hierarchical model

- Comparative Performance of Models:
  hierarchical model does not significantly outperform the pooled model in terms of accuracy

Thanks!

Take away

- Considered models perform comparably in terms of accuracy.
- In the hierarchical model, presence of divergent transitions during the sampling process.

- Education does not influence on credit default.
- BDA allows to learn from small dataset

vivian.phan@aalto.fi
pimikov@gmail.com
thilini.panagodaarachchige@aalto.fi

Stan code for pooled model

```
fit1 <- brm(default ~ PAY_0 + PAY_2,
             data = data_credit,
             refresh = 0,
             prior=c(
                     prior(normal(0,100),
class="Intercept"),
prior(normal(0, 100), class = b)),
            family = bernoulli(),
            file = "pooled",
            backend = "cmdstanr",
            seed = 123
)
```

Stan code for hierarchical model

```
fit2 <- brm(
           default ~  PAY_0 + PAY_2 + (1 |
 EDUCATION),
           data = data_credit,
           prior=c(
               prior(normal(0,100),
class="Intercept"),
               prior(normal(0,100),  class="b"),
               prior(exponential(.02),
class="sd")),
           family = bernoulli(),
           backend = "cmdstanr",
           iter = 5000,  warmup = 2500,
           seed = 123
```