The Instability of Mean Target Encoding in LGD and EAD Modeling

Andrija Djurovic

www.linkedin.com/in/andrija-djurovic

Standard Approaches to LGD and EAD Modeling

- In IRB credit risk modeling, Ordinary Least Squares (OLS) regression is one of the most commonly used methods for modeling Loss Given Default (LGD) and Exposure at Default (EAD) risk parameters.
- When developing these models, practitioners often work with a mix of numeric and categorical risk factors. Although not mandatory, binning is commonly used in risk factor engineering.
- After binning, practitioners must select an appropriate encoding method to run the OLS regression. Commonly discussed encoding methods in LGD and EAD modeling include mean target encoding, Weights of Evidence (WoE) encoding adjusted for a continuous target, and dummy encoding.
- Practitioners often debate using one encoding method over another, while some advocate combining multiple methods within the final model. This typically represents a trade-off between degrees of freedom, explainability, and accuracy.
- This presentation investigates risk factor encoding from a different perspective. Specifically, it examines whether a model using dummy and mean target encoding can be perfectly replicated given a perfect underlying model. This work extends earlier research on WoE encoding instability in Probability of Default (PD) modeling conducted in collaboration with Dr. Alan Forrest.
- The results of this simulation reinforce conclusions from previous PD modeling exercises that dummy encoding provides perfect replication, whereas mean target encoding does not. Practitioners should therefore be aware of this instability in all methods that pre-process risk factors to reduce degrees of freedom. The extent of this instability, and whether it affects model monitoring, should be assessed on a case-by-case basis.
- The following slides outline the simulation design and present results based on simulated LGD data.

Simulation Design

The following simulation uses data available at this link, where the column `lgd` is the target variable and the remaining columns are categorical risk factors.

The simulation design consists of the following steps:

1. Fix a factor encoding method.
2. Estimate the LGD model using OLS regression on the training dataset.
3. Predict LGD values using the model estimated in step 2.
4. Extract and use the predicted LGD values as a new target variable.
5. Check whether the model replicates by comparing the mean target values across risk-factor modalities.
6. Repeat the above steps N times and graphically present the differences in mean target values across modalities for one or more risk factors over consecutive cycles (simulations), varying the number of risk factors included in the final model.

Note

Practitioners should note that this simulation design does not directly compare the estimated OLS coefficients but instead evaluates the mean target values across risk factor modalities. These two approaches are equivalent for confirming model replicability.

The following slides present the results for dummy encoding and mean target encoding separately, assessing model replicability under the assumption that the final model includes four risk factors.
The last slide graphically illustrates the differences in mean target values across modalities over iterations for one selected risk factor, under varying numbers of risk factors included in the final model.

Dummy Encoding - Does the Model Replicate?

The following table presents the model estimates for the assumed final model based on four risk factors using dummy encoding. Specifically, the estimated model is of the form `lgd ~ rf_05 + rf_16 + rf_02 + rf_13`.

```
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    0.4410     0.0657  6.7115   0.0000
## rf_0502 [38.408,177.9077)     -0.2056     0.0341 -6.0239   0.0000
## rf_0503 [177.9077,198.5536)   -0.2871     0.0483 -5.9404   0.0000
## rf_0504 [198.5536,Inf)        -0.3564     0.0415 -8.5986   0.0000
## rf_05SC                       -0.2828     0.1241 -2.2788   0.0229
## rf_1602 [0.0634,34.6935)       0.2074     0.0390  5.3231   0.0000
## rf_1603 [34.6935,43.6851)      0.3147     0.0543  5.7982   0.0000
## rf_1604 [43.6851,Inf)          0.3796     0.0463  8.1914   0.0000
## rf_16SC                        0.1544     0.0402  3.8414   0.0001
## rf_0202 [47.8446,125.0093)    -0.1918     0.0260 -7.3856   0.0000
## rf_0203 [125.0093,Inf)        -0.2628     0.0425 -6.1842   0.0000
## rf_02SC                       -0.1408     0.0291 -4.8447   0.0000
## rf_1302 [1,2.6617)             0.0301     0.0446  0.6755   0.4995
## rf_1303 [2.6617,5.0394)        0.0867     0.0449  1.9311   0.0537
## rf_1304 [5.0394,Inf)           0.1302     0.0444  2.9310   0.0034
## rf_13SC                       -0.0515     0.0612 -0.8414   0.4003
```

After estimating the model, the next step is to generate predictions and compare the mean target values of the original `lgd` column (`mt_initial`) with the predicted values (`mt_predict`) across the modalities of the risk factor `rf_05`.

```
##                    rf_05   n mt_initial mt_predict
## 1        01 (-Inf,38.408) 111     0.6436     0.6436
## 2    02 [38.408,177.9077) 815     0.3822     0.3822
## 3 03 [177.9077,198.5536)  91     0.2654     0.2654
## 4      04 [198.5536,Inf) 175     0.1925     0.1925
## 5                     SC   8     0.2032     0.2032
```

Conclusion

As shown in the results, the mean target values for risk factor `rf_05` are consistent across the two iterations. Practitioners can therefore conclude that the model using dummy encoding replicates perfectly.

Mean Target Encoding - Does the Model Replicate?

The following table presents the model estimates for the assumed final model based on four risk factors using mean target encoding. Specifically, the estimated model is of the form `lgd ~ rf_05 + rf_16 + rf_02 + rf_13`.

```
##             Estimate Std. Error  t value Pr(>|t|)
## (Intercept)  -0.6854     0.0584 -11.7377        0
## rf_05         0.7789     0.0868   8.9684        0
## rf_16         0.7711     0.0887   8.6923        0
## rf_02         0.7182     0.0922   7.7934        0
## rf_13         0.5911     0.1088   5.4348        0
```

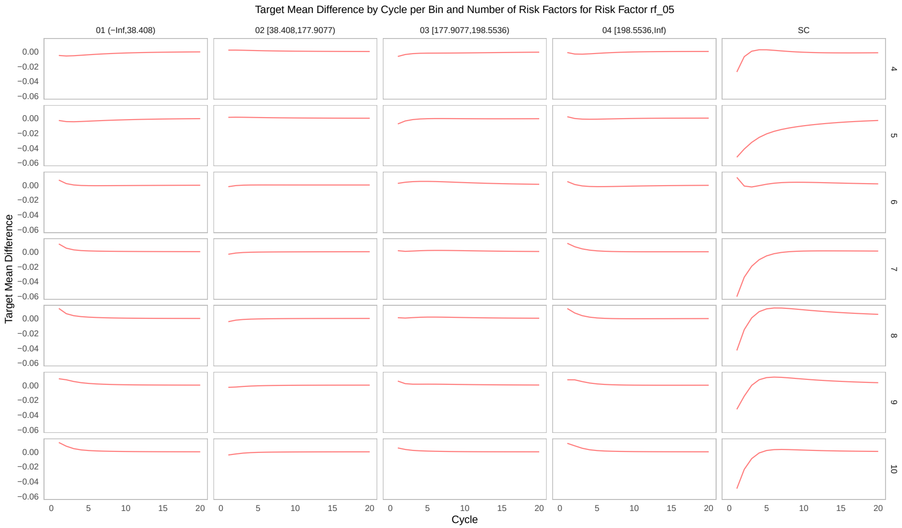
After estimating the model, the next step is to generate predictions and compare the mean target values of the original `lgd` column (`mt_initial`) with the predicted values (`mt_predict`) across the modalities of the risk factor `rf_05`.

```
##                    rf_05   n mt_initial mt_predict
## 1       01 (-Inf,38.408) 111     0.6436     0.6384
## 2   02 [38.408,177.9077) 815     0.3822     0.3842
## 3 03 [177.9077,198.5536)  91     0.2654     0.2588
## 4      04 [198.5536,Inf) 175     0.1925     0.1912
## 5                     SC   8     0.2032     0.1757
```

Conclusion

As shown in the results, the mean target values for risk factor `rf_05` are not consistent across the two iterations. Practitioners can therefore conclude that the model using mean target encoding does not replicate.

Visualization of Differences in Mean Target Encoding



Discussion Points

- Precisely what model risks are present in this mean target encoding instability?
- How far should model validators and model risk managers be concerned about this?
- How should this concern be expressed, the risk managed? What practices should we change or adopt?
- What performance outcomes would justify keeping the model unchanged? In some cases it seems that a perfect performance is not good enough: do we seek new kinds of model monitoring?
- How does the mean target instability in OLS-based LGD/EAD modeling compare to WoE instability in logistic regression and PD modeling?