Traditional loss given default (LGD) calculations are typically estimated by looking at historical averages, usually segregated by collateral type and seniority type. This article offers an approach that regresses LGD on several independent variables with the goal of understanding which factors drive bank loan losses and, further, the amount of losses to expect when loans do default. The model presented here incorporates macroeconomic, firm-specific, and loan-specific information. While the exact model presented here won’t fit every bank, the approach and methodology may help in estimating other LGD models.

Our model results show that several variables are both statistically and economically significant in modeling LGD. They include total debt over capital, profit margin, revenue, ratio of current liabilities over total liabilities, risk rating at default, industry type, GDP growth rate, and tightness of the loan covenants. Some variables, such as collateral type and seniority, appear to be insignificant, while outstanding balance at default and authorized balance at default are on the border of significance.

The next section discusses the dependent variable (LGD) definition as well as the econometric procedure to transform it from its original beta distribution to a normal distribution suitable for ordinary least squares (OLS) regressions. Following that we discuss the independent variables used in the analysis as well as their rationale. The fourth section discusses the final model and its results. Conclusions are drawn at the end to summarize the results and approach.

**Dependent Variable**

We defined LGD, the loss given default rate, as the percentage of the charge-off (net of charge-off recovery) over outstanding balance at default for each defaulted loan. The definition of LGD used requires careful attention since there are at least two ways of defining LGD. Theoretically, loss derived from net charge-off should be the same as that
A General Methodology for Modeling Loss Given Default

derived from the remaining balance of total cash outflows and cash inflows after default. Specifically,

Loss derived from net charge-off = charge-off - charge-off recovery
Loss derived from cash flows = total cash outflows - total cash inflows

While the loss data derived from these two sources should be the same, empirically some discrepancies are noticed and require deeper investigation to ensure consistency with other definitions used in the overall Basel-compliant risk analysis modeling framework within the bank, especially with the definitions used in any other data, internal or external. Since the loss data is used to arrive at the independent variable LGD, such discrepancies might significantly affect the modeling results.

As a normally distributed variable is required to utilize OLS regression techniques, we transform LGD from its original underlying beta distribution to a normal distribution. This is done in two steps.

1. First we calculate the $\alpha$ and $\beta$ parameters from the underlying beta distribution as follows.
   Define
   
   \[
   LGD = \frac{(\text{Charge-off} - \text{charge-off recovery})}{\text{Outstanding balance at default}}
   \]
   
   $\alpha$: The beta distribution’s center parameter and can be derived from equations below
   
   $\beta$: The beta distribution’s shape parameter and can be derived from equations below
   
   Min: Minimum of all cases
   
   Max: Maximum of all cases
   
   $\alpha$ and $\beta$ are then derived from the following equations:

   \[
   \alpha = \frac{\mu}{\text{Max}} \cdot \left[ \frac{\mu}{\text{Max}} - \frac{\text{Max} - \mu}{\text{Max}} \cdot \delta^2 - 1 \right]
   \]

   \[
   \beta = \alpha \cdot \left[ \frac{\text{Max}}{\mu} - 1 \right]
   \]

   where $\mu$, $\delta^2$ are population mean and variance respectively.

2. We then transform LGD from a beta to a normal distribution suitable for use in OLS regressions, using the definitions of $\alpha$ and $\beta$ as calculated above.

   \[
   \text{NLGD} = \text{Normal} (LGD, \mu', \sigma')
   \]

   \[
   \mu' = \frac{\alpha \cdot \text{Max}}{\alpha + \beta} \quad \text{and} \quad \sigma' = \sqrt{\frac{\alpha \cdot \beta}{(\alpha + \beta)^2 (1 + \alpha + \beta)}} \cdot \text{Max}
   \]

Independent Variables

Based on our experience, our reading of the literature and empirical analysis, we tried various explanatory factors in the model. These include variables capturing loan-specific information, firm-specific information, and macroeconomic information, as follows:

**Loan-specific information.**
- Ratio of collateral value at default / outstanding at default.
- Ratio of collateral value at one year before default / outstanding at default.
- Outstanding balance at default / one year before default.
- Authorized balance at default.
- Risk rating at default.
- Risk rating at one year before default.
- Collateral type.
- Facility type.
- Covenant structure.
- Seniority.

**Firm-specific information.**
- Leverage I: (total assets - net worth) / net worth.
- Leverage II: total debt / capital at default.
- Operating income/sales at default.
- Current liability/total liability at one year before default.
- Firm size: revenue, total assets, net worth.
- Industry type.

**General economy information.**
- GDP quarterly growth rate one year before default.

The collateral ratio, defined as a ratio of collateral value at default (or one year before default) to the outstanding balance at default (or one year before default), is used as one explanatory variable. This estimates a bank loan’s recovery rate as a function of its corresponding collateral ratio where collateral is involved. As empirically used in many credit risk models, the expected negative relationship between the recovery rate and the collateral rate proves useful in estimating recovery rates, in setting sufficient collateral requirements, and in pricing debt with stochastic collateral values.

Each loan was assigned a risk rating based on the bank’s internal risk-rating model. The ratings range from 10 to 100; the higher the rating, the greater the risk of the loan.

We assigned dummy variables for the facility type based on the total numbers of each facility type
in our dataset. For some facility types, such as committed revolving loans, committed term loans, or fixed term loans, there are quite a few data points in our dataset. For some facility types, on the other hand, there are small numbers of data points available in the dataset. We combined these small groups into “Others.” With these dummy variables in the model, we can easily see which facility type is significant in estimating LGD. Also, in separate runs, we assigned one single dummy variable to capture whether facility type, in general, contributes to the model. While some specific facility type might not be significant in the model, the overall facility-type factor may appear to be an important factor in predicting LGD.

Similarly, we assigned dummy variables for industry type based on the bank’s classification criteria and, in separate runs, one single dummy variable for the industry type in general. This is to see whether industry type, in general, is a significant contributor to the model, and then which specific industry type is a truly significant factor among all industry types.

Seniority of loans was also captured through a dummy variable. Senior bank loans are guaranteed senior claim positions when the borrowers default, while non-senior loans are either subordinated loans or other loans. As such, during the resolution process, senior bank loans will benefit from the liquidation procedure prior to other junior creditors and borrowers’ shareholders. Therefore, recovery rates from senior bank loans will be higher than those of non-senior loans, leading to a negative relationship between LGD and senior loans.

GDP quarterly growth rate is based on the percentage change from the preceding period, one year preceding default. The GDP rate one year before default is a direct and relevant factor in triggering events that ultimately result in the borrowers’ defaults a year later.

We experimented with several definitions of leverage as per the literature and settled on using the ratio of total debt over capital at default—a ratio capturing borrowers’ leverage at default—as the most appropriate definition. Other definitions of leverage gave counterintuitive signs.

In our preliminary analysis, we found that firm size, proxied by revenue, total assets, or net worth, is useful in estimating the loss given default. A negative relationship is expected between LGD and firm size, since larger firms are expected to have better tracking and enforcement mechanisms, as well as better recovery capabilities and hence lower LGDs.

Since covenant structures of each facility are able to protect against default losses to some degree, we expect covenant structures to contribute to the model in a positive way. We created three dummy variables for the covenant structures—representing no covenant, weak covenant, and tight covenant, respectively—based on individual judgment of the description of the covenants. The covenant dummy variables proved to have significant impact on the model results. In separate runs, we also assigned one dummy variable to capture covenant tightness in general. As covenants get tighter, LGD values typically go lower.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient Sign</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-</td>
</tr>
<tr>
<td>Collateral Value Outstanding at Default</td>
<td>-</td>
</tr>
<tr>
<td>Authorized Balance at Default</td>
<td>+</td>
</tr>
<tr>
<td>Risk Rating at Default</td>
<td>+ *</td>
</tr>
<tr>
<td>Revenue</td>
<td>. +</td>
</tr>
<tr>
<td>Leverage</td>
<td>+ *</td>
</tr>
<tr>
<td>Profitability</td>
<td>+ *</td>
</tr>
<tr>
<td>Current Ratio</td>
<td>. +</td>
</tr>
<tr>
<td>Interest Coverage</td>
<td>-</td>
</tr>
<tr>
<td>Agriculture/Food Industry</td>
<td>-</td>
</tr>
<tr>
<td>Construction Industry</td>
<td>-</td>
</tr>
<tr>
<td>Manufacturing Industry</td>
<td>+</td>
</tr>
<tr>
<td>Services Industry</td>
<td>- *</td>
</tr>
<tr>
<td>Wholesale Industry</td>
<td>+ *</td>
</tr>
<tr>
<td>Facility Type</td>
<td>-</td>
</tr>
<tr>
<td>GDP</td>
<td>- *</td>
</tr>
<tr>
<td>Collateral Type</td>
<td>+</td>
</tr>
<tr>
<td>Covenant Tightness</td>
<td>. +</td>
</tr>
<tr>
<td>Seniority</td>
<td>-</td>
</tr>
<tr>
<td>R-squared</td>
<td>62%</td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>51%</td>
</tr>
<tr>
<td>P-value of Regression’s F-statistic</td>
<td>0.0</td>
</tr>
</tbody>
</table>

*indicates coefficients that are significant at the 10% level or higher.
Some other financial ratios capturing borrowers’ liquidity and leverage were also experimented with given data availability. Profit margin and percentage of current liabilities in total liabilities are two useful variables in explaining the loss rate.

**Model Results**

We analyzed more than 20 specifications of the model over multiple iterations to arrive at the one that is best fitting, both economically and statistically, for our portfolio. (See Table 1.)

We see that the results from the regressions are not only a good fit, but also intuitive and follow economic sense. The positive coefficient on the risk rating variable indicates that as loans get riskier, the LGD expected increases. Similarly, with the positive sign on the leverage variable, as leverage increases, the LGD also increases. Conversely, as revenue increases, the expected LGD decreases, as indicated by the positive coefficient on the revenue variable. This negative relationship also holds for the current ratio, GDP, and covenant tightness. As the current ratio decreases, the overall economy slows, or as covenants are loosened, the expected LGD increases.

Some of the other variables did not give results as expected. This may have been due to peculiarities with our base modeling portfolio.

This model fits the data reasonably well and gives results that are intuitively sound. Hence it can be used for modeling LGD within the bank to not only improve performance, but also to satisfy Basel II’s requirements. The development, working, and testing of the model are quite open to inspection: There are no black-box components. The model can and should be tested using out-of-sample and out-of-time validation techniques to ensure satisfactory results.

**Conclusion**

We have reviewed a general methodology for developing a reliable model to predict loss given default for a bank portfolio. Great care needs to be taken in defining LGD, in converting raw data to a form suitable for standard regression techniques, as well as in selecting and defining the variables to be regressed. The model built is a multifactor statistical model incorporating data at three levels: loan level, firm level, and macroeconomic level, in the final form. The model built is totally transparent and can be validated using out-of-sample and out-of-time approaches to ensure it gives satisfactory results over time.

The authors can be contacted by e-mail at nadeem.siddiqi@bearingpoint.com and meiqingzhang@bearingpoint.net.

For more information, visit www.BearingPoint.com.