LGD - Tobit - Fractional

Shubham Kotal

August 2024

A Tobit model is a type of regression model used when the value you're predicting (in this case, Loss Given Default (LGD)) is censored. Censoring means that for some data points, you only know that the value is above or below a certain threshold, but you don't know the exact value.

Key Points:

1. Censoring: - Imagine you're predicting the percentage of loss a bank might face if a borrower defaults on a loan (LGD). - LGD is typically between 0 (no loss) and 1 (total loss). - Sometimes, you know that the loss is zero (because the borrower repaid fully), or you know the loss is 100

2. Tobit Model Logic: - The Tobit model works by predicting a latent (hidden) variable that can go below 0 or above 1, but then applies rules to bring it back within the 0 to 1 range. - For example, if the model predicts a value below 0, it sets it to 0 (because LGD can't be negative). If it predicts above 1, it sets it to 1 (because LGD can't be more than 100

3. Example: - Suppose you have data on loans and want to predict LGD based on factors like the loan-to-value (LTV) ratio and borrower's age. - If a loan has a high LTV (the loan amount is close to the value of the collateral), the Tobit model might predict a high LGD, meaning the bank could lose a lot if the borrower defaults.

Simple Example:

Imagine you have three loans:

- Loan 1: LTV = 90- Loan 2: LTV = 70- Loan 3: LTV = 95

Using the Tobit model:

- The model predicts latent values for each loan, say 0.25 for Loan 1, -0.1 for Loan 2, and 1.2 for Loan 3. - Then it adjusts these values to be within the 0-1 range: - For Loan 1: Predicted LGD remains 0.25. - For Loan 2: Predicted LGD is set to 0 (since the latent value was below 0). - For Loan 3: Predicted LGD is set to 1 (since the latent value was above 1).

Why Use Tobit?

The Tobit model is particularly useful when you often encounter situations where the outcome is at the boundary (e.g., 0 or 1 for LGD). It helps provide a better fit than traditional regression models, which might predict values outside the acceptable range.

Let's break down the Tobit model step by step in a very simple way using an example.

Step 1: Understanding the Problem

Imagine you work at a bank, and you're trying to predict how much money the bank will lose if a borrower defaults on a loan. This is called Loss Given Default (LGD). LGD can range from 0

Step 2: What is Censoring?

Censoring happens when you don't know the exact value for some of your data, but you know it falls within a certain range. For example: - If a borrower pays back everything, the LGD is 0- If a borrower doesn't pay back anything, the LGD is 100- Sometimes, the loss is somewhere in between, like 40

Step 3: Introducing the Tobit Model

The Tobit model helps us predict LGD when we know the values are censored at 0

Step 4: How Does Tobit Work? 1. Imagine a Hidden (Latent) Variable: - Think of there being a "hidden" variable called y^*, which could take any value, even outside the range of 0- For example, y^* could be -10

2. Predicting y^*: - The Tobit model predicts this hidden y^* using the information we have, like the loan amount and borrower's age.

3. Adjusting Predictions: - If y^* is predicted to be below 0- If y^* is predicted to be above 100

Step 5: Simple Example

Let's say you have three loans:

- Loan 1: Borrower is young, with a small loan. The Tobit model predicts $y^* = 20\%$, so the LGD is 20- Loan 2: Borrower is old, and the loan is fully paid back. The Tobit model predicts $y^* = -5\%$, but since LGD can't be negative, it's set to 0- Loan 3: Borrower is at high risk, and the loan is not paid back. The Tobit model predicts $y^* = 110\%$, but since LGD can't be more than 100

Step 6: Why Not Use Regular Regression?

If you just used regular regression (which doesn't know about the limits of 0

Conclusion

The Tobit model is like a regression model with a built-in understanding that LGD is between 0

Yes, you can use a fractional logit model to predict Loss Given Default (LGD), especially when LGD is a proportion between 0 and 1 (or 0

How the Fractional Logit Model Works:

1. Understanding the Data: - The fractional logit model is designed for dependent variables that are proportions or fractions (like LGD, which is between 0 and 1).

2. Logistic Transformation: - Instead of predicting the proportion directly, the fractional logit model uses a logistic transformation. This transformation ensures that the predicted values are always between 0 and 1. - The model estimates the probability of the outcome being at a certain level within the 0 to 1 range using a logit function, similar to logistic regression.

3. Model Equation: - The basic idea is similar to logistic regression but

applied to fractional outcomes:

$$logit(E(y|X)) = X\beta$$

- Here, y is your LGD, X represents your explanatory variables, and β are the coefficients to be estimated. - The logit function transforms the expected value of y so that it falls within the [0, 1] range.

4. Advantages of Fractional Logit: - No Censoring Assumption: Unlike the Tobit model, which is specifically useful when there's censoring at 0 or 1, the fractional logit model doesn't assume censoring. It's ideal when your data naturally falls within the [0, 1] range without actual observations at the boundaries being censored. - Predicts Within Bounds: The fractional logit model automatically keeps predictions within the [0, 1] range, making it more appropriate when dealing with proportions.

Example: Let's say you have the same loans:

- Loan 1: High LTV, younger borrower. LGD might be around 30- Loan 2: Low LTV, older borrower. LGD might be very low, say 5- Loan 3: Very high LTV, high-risk borrower. LGD might be close to 100

When to Use Fractional Logit vs. Tobit:

- Fractional Logit: Best when your LGD data is naturally within [0, 1] and you don't have to worry about censoring at 0 or 1. - Tobit Model: Better when there is actual censoring at the boundaries (like when some loans have an LGD exactly at 0 or 1, and you suspect that's due to a censoring mechanism).

Conclusion:

The fractional logit model is a flexible and appropriate choice for modeling LGD when the outcome is a proportion and you don't have censoring. It ensures predictions stay within the [0, 1] range and is simpler to use when censoring isn't a concern.

1. Purpose and Interpretation: Logistic Regression:

Purpose: Designed for binary outcomes (0 or 1). Interpretation: The output is a probability of the outcome being 1 (e.g., an 80Fractional Logit Model:

Purpose: Designed for continuous proportions directly (e.g., 0.3, 0.7, 0.95). Interpretation: The output is the expected value of the proportion itself, not just the probability of hitting 1. The model is explicitly built to handle data where the outcome is a proportion.

Sure! Let's walk through the process of using the fractional logit model with a FICO score example, including how to convert the results back to the original scale.

1. Scaling the FICO Score:

First, you need to scale the FICO score to fit within the [0, 1] range. The formula to scale a FICO score Y is:

$$Y_{scaled} = \frac{Y - 350}{850 - 350}$$

- 350 is the minimum FICO score. - 850 is the maximum FICO score.

2. Apply the Fractional Logit Model:

Once the FICO scores are scaled, you can apply the fractional logit model to predict the scaled FICO score.

3. Converting Predictions Back to Original Scale:

After applying the fractional logit model, you'll get predictions in the [0, 1] range. To convert these predictions back to the original FICO score scale, use the inverse of the scaling formula:

$$Y = Y_{scaled} \times (850 - 350) + 350$$

Example:

Let's go through an example step-by-step:

1. Original FICO Score: - Suppose you have a FICO score of 700.

2. Scale the FICO Score: - Use the scaling formula to convert it to the [0, 1] range:

$$Y_{scaled} = \frac{700 - 350}{850 - 350} = \frac{350}{500} = 0.7$$

3. Apply the Fractional Logit Model: - Suppose the fractional logit model gives you a predicted value of 0.3 for Y_{scaled}. This is a model output in the [0, 1] range.

4. Convert the Scaled Prediction Back to the Original FICO Scale: - Use the inverse scaling formula to convert the predicted value back to the original FICO score range:

$$Y = 0.3 \times (850 - 350) + 350$$

$$Y = 0.3 \times 500 + 350$$

$$Y = 150 + 350 = 500$$

So, a predicted scaled value of 0.3 corresponds to a FICO score of 500 on the original scale.

Summary:

- Scale the Original Score to fit within [0, 1] for modeling. - Apply the Fractional Logit Model to get predictions within [0, 1]. - Convert Predictions Back to the original FICO score scale using the inverse of the scaling formula.

This process allows you to use a fractional logit model effectively with FICO scores, even though the scores themselves are not naturally within the [0, 1] range.]

A Tobit model is a type of regression model used when the value you're predicting (in this case, Loss Given Default (LGD)) is censored. Censoring means that for some data points, you only know that the value is above or below a certain threshold, but you don't know the exact value.

Key Points:

1. Censoring: - Imagine you're predicting the percentage of loss a bank might face if a borrower defaults on a loan (LGD). - LGD is typically between 0 (no loss) and 1 (total loss). - Sometimes, you know that the loss is zero (because the borrower repaid fully), or you know the loss is 100

2. Tobit Model Logic: - The Tobit model works by predicting a latent (hidden) variable that can go below 0 or above 1, but then applies rules to

bring it back within the 0 to 1 range. - For example, if the model predicts a value below 0, it sets it to 0 (because LGD can't be negative). If it predicts above 1, it sets it to 1 (because LGD can't be more than 100

3. Example: - Suppose you have data on loans and want to predict LGD based on factors like the loan-to-value (LTV) ratio and borrower's age. - If a loan has a high LTV (the loan amount is close to the value of the collateral), the Tobit model might predict a high LGD, meaning the bank could lose a lot if the borrower defaults.

Simple Example:

Imagine you have three loans:

- Loan 1: LTV = 90- Loan 2: LTV = 70- Loan 3: LTV = 95

Using the Tobit model:

- The model predicts latent values for each loan, say 0.25 for Loan 1, -0.1 for Loan 2, and 1.2 for Loan 3. - Then it adjusts these values to be within the 0-1 range: - For Loan 1: Predicted LGD remains 0.25. - For Loan 2: Predicted LGD is set to 0 (since the latent value was below 0). - For Loan 3: Predicted LGD is set to 1 (since the latent value was above 1).

Why Use Tobit?

The Tobit model is particularly useful when you often encounter situations where the outcome is at the boundary (e.g., 0 or 1 for LGD). It helps provide a better fit than traditional regression models, which might predict values outside the acceptable range.

Let's break down the Tobit model step by step in a very simple way using an example.

Step 1: Understanding the Problem

Imagine you work at a bank, and you're trying to predict how much money the bank will lose if a borrower defaults on a loan. This is called Loss Given Default (LGD). LGD can range from 0

Step 2: What is Censoring?

Censoring happens when you don't know the exact value for some of your data, but you know it falls within a certain range. For example: - If a borrower pays back everything, the LGD is 0- If a borrower doesn't pay back anything, the LGD is 100- Sometimes, the loss is somewhere in between, like 40

Step 3: Introducing the Tobit Model

The Tobit model helps us predict LGD when we know the values are censored at 0

Step 4: How Does Tobit Work? 1. Imagine a Hidden (Latent) Variable: - Think of there being a "hidden" variable called y^*, which could take any value, even outside the range of 0- For example, y^* could be -10

2. Predicting y^*: - The Tobit model predicts this hidden y^* using the information we have, like the loan amount and borrower's age.

3. Adjusting Predictions: - If y^* is predicted to be below 0- If y^* is predicted to be above 100

Step 5: Simple Example

Let's say you have three loans:

- Loan 1: Borrower is young, with a small loan. The Tobit model predicts $y^* = 20\%$, so the LGD is 20- Loan 2: Borrower is old, and the loan is fully paid back. The Tobit model predicts $y^* = -5\%$, but since LGD can't be negative, it's set to 0- Loan 3: Borrower is at high risk, and the loan is not paid back. The Tobit model predicts $y^* = 110\%$, but since LGD can't be more than 100

Step 6: Why Not Use Regular Regression?

If you just used regular regression (which doesn't know about the limits of 0

Conclusion

The Tobit model is like a regression model with a built-in understanding that LGD is between 0

Yes, you can use a fractional logit model to predict Loss Given Default (LGD), especially when LGD is a proportion between 0 and 1 (or 0

How the Fractional Logit Model Works:

1. Understanding the Data: - The fractional logit model is designed for dependent variables that are proportions or fractions (like LGD, which is between 0 and 1).

2. Logistic Transformation: - Instead of predicting the proportion directly, the fractional logit model uses a logistic transformation. This transformation ensures that the predicted values are always between 0 and 1. - The model estimates the probability of the outcome being at a certain level within the 0 to 1 range using a logit function, similar to logistic regression.

3. Model Equation: - The basic idea is similar to logistic regression but applied to fractional outcomes:

$$logit(E(y|X)) = X\beta$$

- Here, y is your LGD, X represents your explanatory variables, and β are the coefficients to be estimated. - The logit function transforms the expected value of y so that it falls within the [0, 1] range.

4. Advantages of Fractional Logit: - No Censoring Assumption: Unlike the Tobit model, which is specifically useful when there's censoring at 0 or 1, the fractional logit model doesn't assume censoring. It's ideal when your data naturally falls within the [0, 1] range without actual observations at the boundaries being censored. - Predicts Within Bounds: The fractional logit model automatically keeps predictions within the [0, 1] range, making it more appropriate when dealing with proportions.

Example: Let's say you have the same loans:

- Loan 1: High LTV, younger borrower. LGD might be around 30- Loan 2: Low LTV, older borrower. LGD might be very low, say 5- Loan 3: Very high LTV, high-risk borrower. LGD might be close to 100

When to Use Fractional Logit vs. Tobit:

- Fractional Logit: Best when your LGD data is naturally within [0, 1] and you don't have to worry about censoring at 0 or 1. - Tobit Model: Better when there is actual censoring at the boundaries (like when some loans have an LGD exactly at 0 or 1, and you suspect that's due to a censoring mechanism).

Conclusion:

The fractional logit model is a flexible and appropriate choice for modeling LGD when the outcome is a proportion and you don't have censoring. It ensures predictions stay within the [0, 1] range and is simpler to use when censoring isn't a concern.

1. Purpose and Interpretation: Logistic Regression:

Purpose: Designed for binary outcomes (0 or 1). Interpretation: The output is a probability of the outcome being 1 (e.g., an 80Fractional Logit Model:

Purpose: Designed for continuous proportions directly (e.g., 0.3, 0.7, 0.95). Interpretation: The output is the expected value of the proportion itself, not just the probability of hitting 1. The model is explicitly built to handle data where the outcome is a proportion.

Sure! Let's walk through the process of using the fractional logit model with a FICO score example, including how to convert the results back to the original scale.

1. Scaling the FICO Score:

First, you need to scale the FICO score to fit within the [0, 1] range. The formula to scale a FICO score Y is:

$$Y_{scaled} = \frac{Y - 350}{850 - 350}$$

- 350 is the minimum FICO score. - 850 is the maximum FICO score.

2. Apply the Fractional Logit Model:

Once the FICO scores are scaled, you can apply the fractional logit model to predict the scaled FICO score.

3. Converting Predictions Back to Original Scale:

After applying the fractional logit model, you'll get predictions in the [0, 1] range. To convert these predictions back to the original FICO score scale, use the inverse of the scaling formula:

$$Y = Y_{scaled} \times (850 - 350) + 350$$

Example:

Let's go through an example step-by-step:

1. Original FICO Score: - Suppose you have a FICO score of 700.

2. Scale the FICO Score: - Use the scaling formula to convert it to the [0, 1] range:

$$Y_{scaled} = \frac{700 - 350}{850 - 350} = \frac{350}{500} = 0.7$$

3. Apply the Fractional Logit Model: - Suppose the fractional logit model gives you a predicted value of 0.3 for Y_{scaled}. This is a model output in the [0, 1] range.

4. Convert the Scaled Prediction Back to the Original FICO Scale: - Use the inverse scaling formula to convert the predicted value back to the original FICO score range:

$$Y = 0.3 \times (850 - 350) + 350$$

$$Y = 0.3 \times 500 + 350$$

$$Y = 150 + 350 = 500$$

So, a predicted scaled value of 0.3 corresponds to a FICO score of 500 on the original scale.

Summary:

- Scale the Original Score to fit within [0, 1] for modeling. - Apply the Fractional Logit Model to get predictions within [0, 1]. - Convert Predictions Back to the original FICO score scale using the inverse of the scaling formula.

This process allows you to use a fractional logit model effectively with FICO scores, even though the scores themselves are not naturally within the [0, 1] range.