Transactions, SMiRT-26
Berlin/Potsdam, Germany, July 10-15, 2022
Division VII

APPLICATION OF ZERO-MODIFIED PROBABILITY DISTRIBUTION TO TSUNAMI FRAGILITY ASSESSMENT

Toshiaki Kondo,[1] Yoshinori Mihara,[2] Yoshiyuki Takahashi,[3] Naoto Kihara,[4] Hideki Kaida[5]

[1] Kajima Corporation, Tokyo, Japan (kondtosh@kajima.com)
[2] Senior Manager, Kajima Corporation, Tokyo, Japan (y-mihara@kajima.com)
[3] Deputy Manager, Kajima Corporation, Tokyo, Japan (takyoshi@kajima.com)
[4] Senior Research Scientist, Central Research Institute of Electric Power Industry, Chiba, Japan (kihara@criepi.denken.or.jp)
[5] Research Scientist, Central Research Institute of Electric Power Industry, Chiba, Japan (h-kaida@criepi.denken.or.jp)

ABSTRACT

In tsunami fragility assessments, the evaluated values of the inundation depth or debris impact speed usually contain many zero values. In this research, we investigated methods that can be applied to data with many zeros. In addition, we applied these methods to the actual analysis results at a virtual site. We modeled the relationship between tsunami height and inundation depth by a two-part model and debris impact speed by k-means clustering and the Gaussian mixture model. Both methods were able to partially model the simulation results, but they are not complete and further research is required.

INTRODUCTION

In tsunami fragility assessments, the evaluated values of inundation depth or debris impact speed usually contain many zero values. This is a unique characteristic of tsunami fragility assessments since the evaluated values for seismic fragility assessments, such as peak ground acceleration, are always positive.

Kihara et al. (2019) conducted tsunami risk assessments at a virtual site. The virtual site is a non-existent nuclear power plant located on the Pacific coast of the Tohoku region of Japan (Figure 1). Kihara et al. (2019) evaluated the earthquake occurrence probability at seven earthquake source regions (Figure 1) and the annual frequency of exceedance of tsunami height.

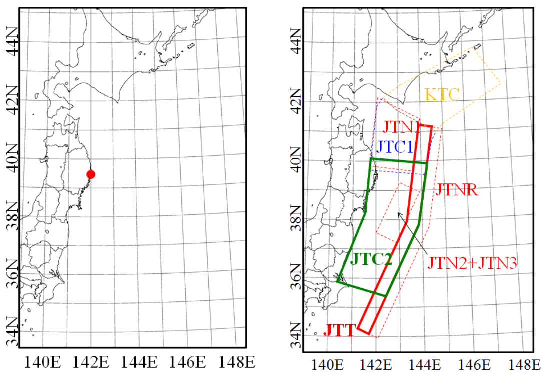

Figure 1. Virtual site location (left) and earthquake source regions (right).

APPLICATION TO INUNDATION DEPTH

Objective

Takahashi et al. (2020) conducted tsunami inundation simulations at the virtual site. The virtual site has two ground levels (T.P. (Tokyo Peil) +7 m and T.P. +15 m), and it has a seawall with a height of T.P. +27 m (Figure 2). Takahashi et al. (2020) numerically evaluated tsunami inundation depths at several locations (Figure 2) at the virtual site under five tsunami heights (T.P.+31 m, 35 m, 39 m, 43 m, and 47 m) with 7 to 21 earthquake source regions for each tsunami height.

The numerical results of Takahashi et al. (2020) contain many zero values (Figure 3). We investigated and formulated a model that can be applied for modeling these inundation depth datasets with many zero values.

In this research, only inundation depths at locations A to N were used (Figure 2). Therefore, the elevations of the locations used in this research are all T.P. +15 m. Also, we only use the inundation depths with tsunami heights of 31 m to 39 m since there was no location where the inundation depth was zero when the tsunami heights were 39 m to 47 m.

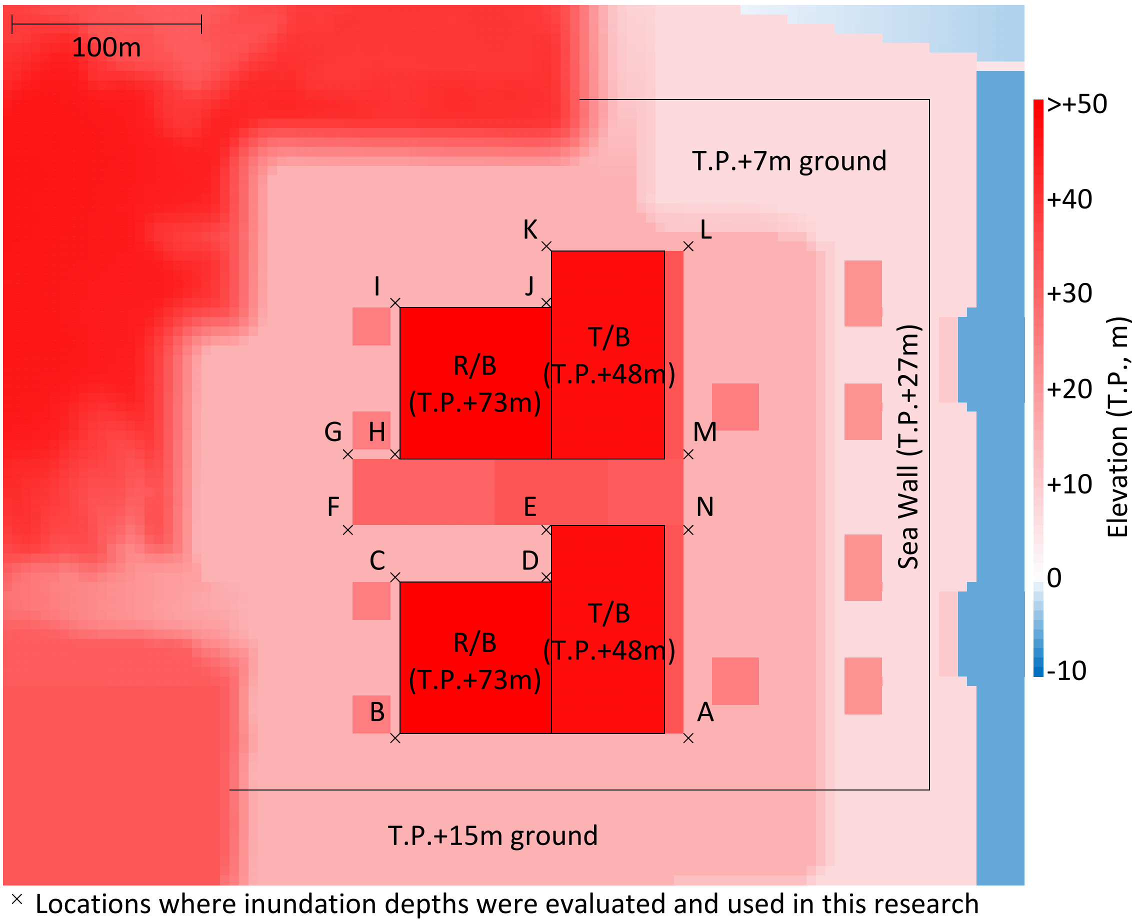

Figure 2. Virtual site topography and initial position of debris (left). Locations where inundation depths were evaluated. In this research, only inundation depths at locations A to N were used (right).

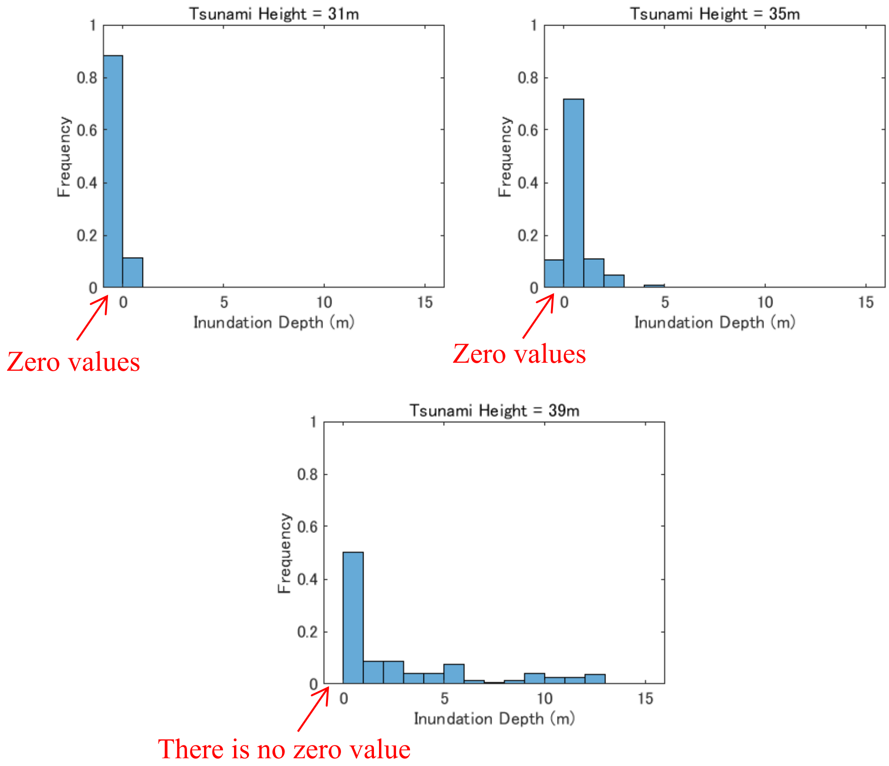

Figure 3. Histogram of inundation depth for each tsunami height using all data from locations A to N.

Method

According to Min and Agresti (2002), there are two major methods for modeling continuous data with many zeros. The first is the Tobit model (Tobin, 1958) and the second is the two-part model (Duan et al., 1983). Since inundation depths can be modeled with the log-normal distribution (Takahashi et al. 2020), we adopted the two-part model in which we can model the positive data with the log-normal distribution.

The probability density function (PDF) of the two-part model is the sum of the zero data part and the positive data part. The positive data part can be modeled with the log-normal distribution. The ratio of the zero data part and the positive data part can be modeled by the logit function using logistic regression. Equation (1) shows the PDF of inundation depth, and equation (2) shows the probability that the inundation depth is greater than 0.

$$f(D) = (1-p)\delta(D) + p\frac{1}{\sqrt{2\pi}\beta D}\exp\left(-\frac{\left(\ln D - R_m(H)\right)^2}{2\beta^2}\right) \quad (1)$$

whereas
D: Inundation depth (m)
$f(D)$: PDF of D
p: Probability that $D > 0$
$\delta(D)$: Dirac's delta
$R_m(H), \beta$: Parameters of positive D distribution (log-normal distribution)
H: Tsunami height (m)

$$p = \frac{1}{1 + e^{-k(\ln(H) - \ln(H_0 \exp(-\beta^u \cdot X)))}} = \frac{1}{1 + (H/(H_0 \exp(-\beta^u \cdot X)))^{-k}} \quad (2)$$

whereas
k, H_0: Parameters of the model (logit function) of p
β^u: Logarithmic standard deviation of epistemic uncertainty factors
X: Standard normal probability variate for logistic regression confidence

Result

Figure 4 shows the regression results at locations J, K, M, and N. Locations J and K have similar curves, probably because the two locations are close together. On the other hand, locations M and N have different curves even though they are close to each other. This is probably because these places are in front of the seawall, and the results differ depending on whether the tsunami enters from the north side or the south side of the seawall.

Regression calculations did not converge for locations A to I and L. This is because the inundation rate at the tsunami height of 31 m is 0%, or the inundation rate at the tsunami height of 35 m is 100%. This suggests that the tsunami heights used for this research were not sufficient for evaluating these locations. Also, further research is needed on the validity of modeling with the logit function, the effect of the site topography, and generalization.

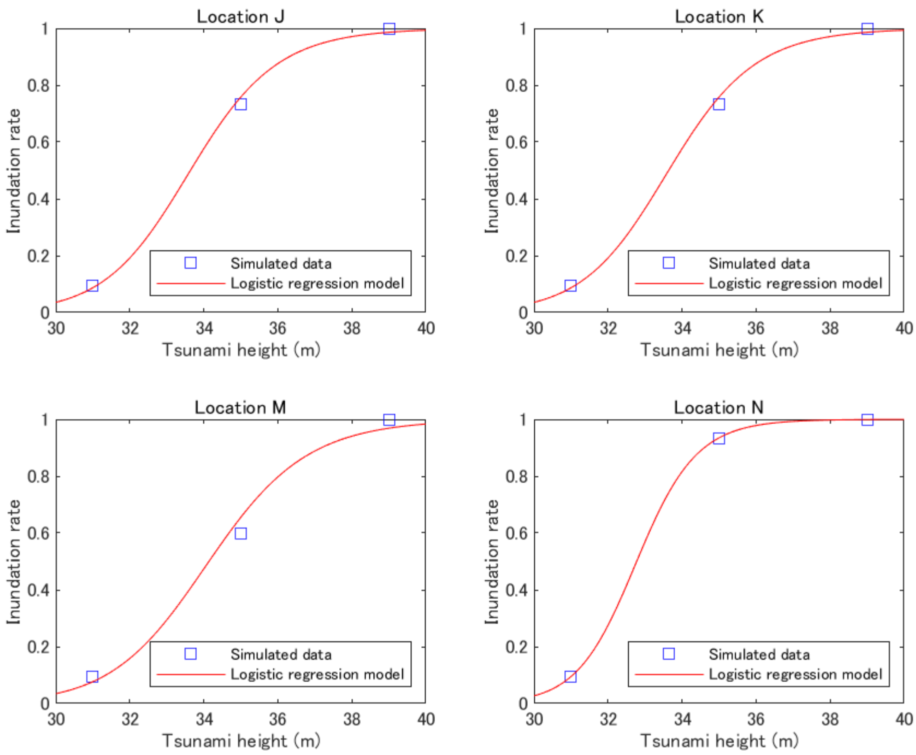

Figure 4. Regression results at locations J, K, M, and N.

APPLICATION TO DEBRIS IMPACT SPEED

Objective

Kaida and Kihara (2020) conducted numerical simulations of debris motion tracking (hereinafter "debris tracking simulation") at the virtual site. The tsunami flow field was calculated using a two-dimensional shallow water model. A vehicle (height 2.0 m, 2.0 m, length 4.8 m, mass 2500 kg) is considered tsunami-borne debris. Six hundred debris elements (200 debris elements for each of three initial positions) were considered for each debris tracking simulation, and 540 cases (54 epistemic uncertainty cases with 10 Monte Carlo simulations) were calculated for each tsunami scenario. (Figure 5)

The debris tracking simulation result of Kaida and Kihara (2020) contains many zero values with some near-zero values (Figure 6). To properly model the debris impact speed distribution, we investigated methods of distinguishing near-zero values from other positive values.

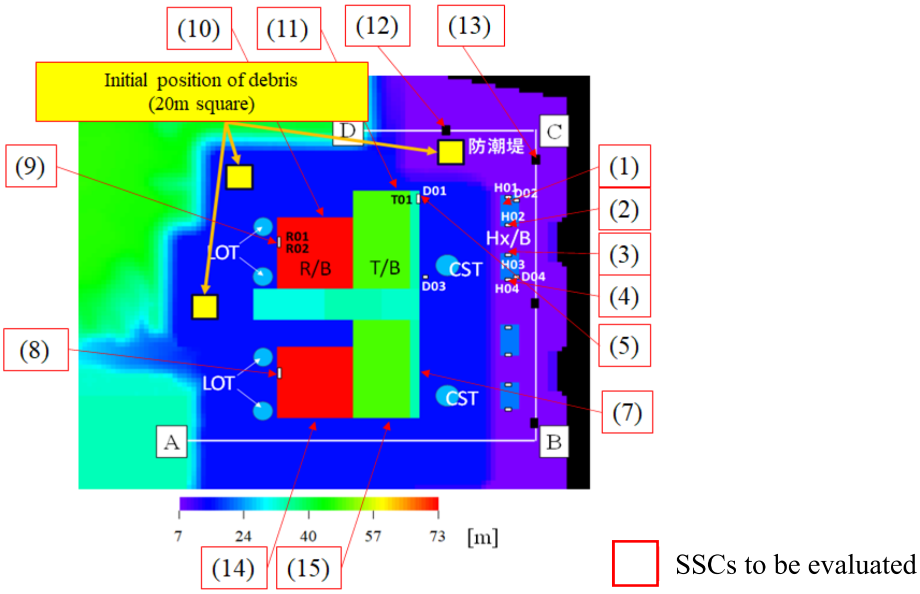

Figure 5. Initial positions of debris (yellow rectangular) and the SSCs (Structures, Systems, and Components) to evaluate debris impact speed (red rectangular with numbers).

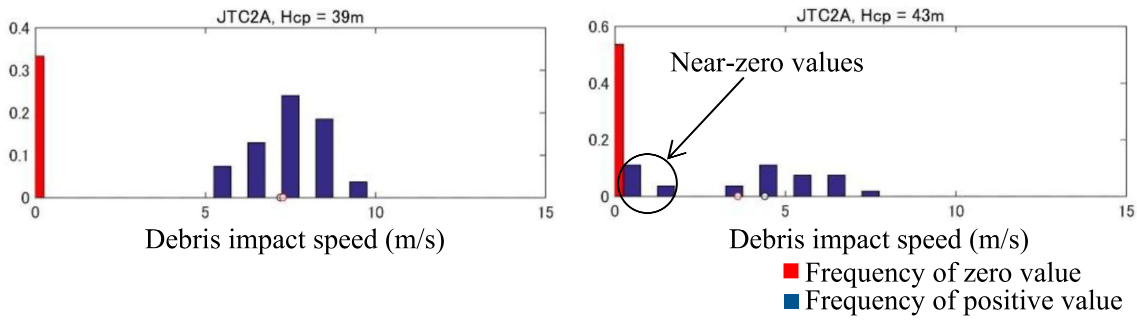

Figure 6. Normalized debris impact speed distribution at SSC (1) in Figure 5 when the tsunami height is 39 m (left) and 43 m (right).

Method

In this research, we applied k-means clustering and the Gaussian mixture model (GMM) for the debris impact speed distribution. The k-means clustering is a method for clustering data based on the distance from the geometric center of each cluster. GMM is a model that expresses and clusters data as a weighted average of a Gaussian distribution. (Figure 7)

Clustering was performed in the four cases shown in Table 1. In each case, both k-means clustering and GMM were used.

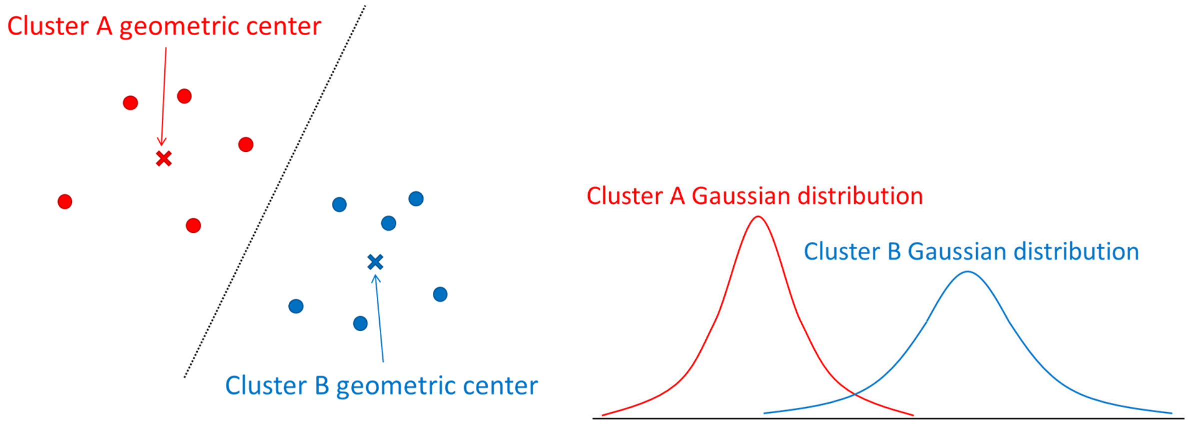

Figure 7. Image diagram of k-means clustering (left) and Gaussian mixture model (right).

Table 1. Clustering cases.

Case	Axis for clustering	Value to replace with zero value (m/s)
1	Linear axis	-
2	Logarithmic axis	10^{-10}
3	Logarithmic axis	10^{-1}
4	Logarithmic axis	0.5

Result

Figure 8 shows the clustering results of debris impact speed at SSC (1) with the tsunami height of 43 m. Case 1 shows good result with both k-means clustering and GMM, but assuming a linear normal distribution for the debris impact is not appropriate. Cases 2 and 3 show almost the same result for GMM and a slightly different result for k-means clustering, and case 3 has a better result. In case 4, the mode value of the log-normal distribution is shifted to the right compared to case 3, which is more consistent with the data. Case 4 shows a good result with both k-means clustering and GMM, but it is necessary to explain the reason to replace the zero value with 0.5 m/s. In addition, it is necessary to analyze the reason why near-zero values are observed in the debris tracking simulation.

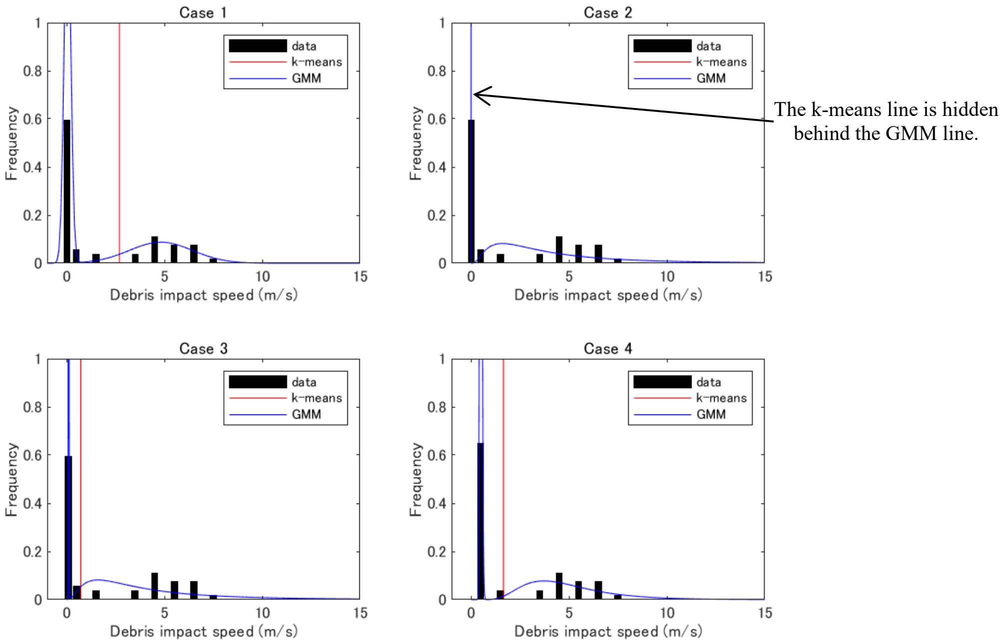

Figure 8. Clustering results at SSC (1) when the tsunami height is 43 m. The red lines denote the perpendicular bisector of each geometric centers of k-means clusters.

CONCLUSION

We investigated and formulated methods to model tsunami fragility assessment-related data containing many zero values, such as tsunami inundation depth and debris impact speed. In addition, the model was applied to the actual analysis result at the virtual site. The inundation depth was modeled with the two-part model with the logit function. The debris impact speed was modeled by k-means clustering and GMM. Both methods were able to model the simulation results partially, but they are not complete and further research is required.

REFERENCES

Duan, N., Manning, W. G., Morris, C. N. and Newhouse, J. P. (1983). "A Comparison of Alternative Models for the Demand for Medical Care," *Journal of Business & Economic Statistics*, 1:2, 115-126.

Kaida, H and N. Kihara. (2020). "Fragility Evaluation Methodology for Tsunami-Borne Debris Impact," *Coastal Engineering Proceedings*, 36, Currents.9

Kihara, N., Kaida, H., Kimura, T., Nishi, A., Masuko, M. and Fujii, N. (2019). "A Tsunami Scenario Determination Methodology for Risk Assessment of Coastal Industrial Facilities," *The Proceedings of the 29th International Ocean and Polar Engineering Conference*, 3292-3296.

Min, Y. and Agresti, A. (2002). "Modeling Nonnegative Data with Clumping at Zero: A Survey," *Journal of The Iranian Statistical Society*, 1, 7-33.

Takahashi, Y., Mihara, Y., Kaida, H. and Kihara, N., (2020). "Uncertainty Quantification of Tsunami Parameters for Tsunami Fragility Assessment Based on Tsunami Hazard Assessment and Inundation Simulation," *Proceedings of the 30th European Safety and Reliability Conference and the 15th Probabilistic Safety Assessment and Management Conference*, 2444-2451

Tobin, J. (1958). "Estimation of Relationships for Limited Dependent Variables," *Econometrica: Journal of the Econometric Society*, 26, 24-36.