Generalizing Censored Regression Models

Andrew Cherpeski, Michael Stafford, Benjamin McMullin

September 8, 2025

1 Introduction

For decades, the challenge of censored data has been a topic of interest for econometricians. Censoring in the econometric sense occurs when the underlying data of interest (y_i^*) is sometimes cropped above or below in our observed data (y_i). An example might be an unemployment survey that asks how long a worker was unemployed before getting a new job. If the worker has already found a new job, then our observed data is the correct value, but if they have not, all we know is a minimum amount of time for them to find a new job. This flaw in the data leads to attenuation bias in traditional estimation techniques.

In 1958, James Tobin used a truncated, Normal MLE to estimate the effects on a censored population. Unfortunately, this model requires the strong assumption that the true underlying variable. Conversely, Powell (1984) showed that quantile regression is an alternative that requires looser assumptions. However, it is less precise than MLE and also measures something different because it works with quantiles rather than expectations.

For our project, we will explore ways that we can use other MLE estimators to estimate effects when the dataset is censored but we have good reason to believe that its true underlying data follow a certain distribution. We will work with Gamma, Negative Binomial, and Poisson distributions.

2 Data

We create our data using computer-generated random values from numpy.

2.1 Gamma distribution

Consisting of two parameters α and β, the Gamma distribution can be thought of as the sum of α independent variables drawn from an exponential distribution with mean β. Its pdf can be expressed as the following:

$$\psi(y_i|\alpha,\beta) = \frac{\beta^{-\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}$$

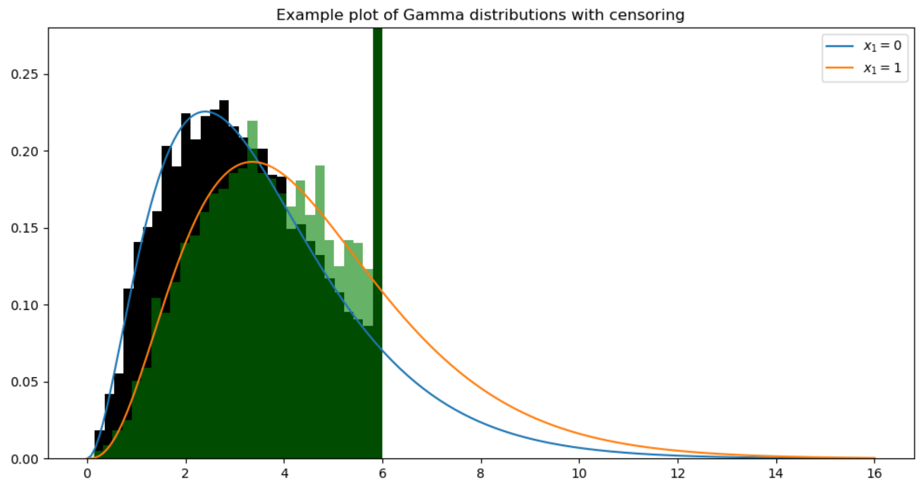

Figure 1: $10,000$ observation sample of values of y_i from the censored Gamma distribution. The values for those treated placed on top of those that were not treated. Observe the extreme censoring.

Observe how all variables drawn from a gamma distribution must be positive. Due to its relation to the exponential distribution, the Gamma distribution is often the distribution of choice when dealing with survival phenomena, such as time spent unemployed or longevity of a lightbulb. Because this theoretical motivation for Gamma distributions exists, the distribution could be a useful way to approach censored problems where survival phenomena are likely to occur.
Let us suppose that a survey has been conducted on workers who lost their jobs six months ago. Our aim is to see whether a benefits program affects the amount of time that they spend unemployed. We define our variables as follows:

$$y_i \sim \text{Gamma}(x_i^\top \alpha, \beta) : \text{the time it takes worker } i \text{ to find a new job.}$$

$$y_i^* = \min\{y_i, 6\} : \text{The censored value of } y_i \text{ that we observe.}$$

$$x_i = (1, x_{1i})^\top \text{The dependent variables.}$$

x_{1i} is an indicator of whether worker i was in the new benefits program.
For our simulation, we will choose the true parameter values to be $\alpha = (3, 0.8)^\top$ and $\beta = 1.2$. In this case, the mean effect of the program will be $(0.8) \cdot (1.2) = 0.96$. In other words, being part of the program will increase a worker's unemployment by about a month. Since the values of y_i are censored at $T = 6$, we will have the problems associated with censoring. See Figure 1 for a histogram of these data.

2.2 Negative Binomial

For a series of independent, identically distributed Bernoulli trials, the negative binomial distribution measures the number of failed trials that occur before k successes. As such, the negative binomial distribution has two parameters: p, representing the probability that any given Bernoulli trial is successful, and k, the predetermined number of successes. The probability mass function is given below:

$$g(y^*) = \binom{y^* + k - 1}{y^*} (1-p)^{y^*} p^k$$

$$\mathbb{E}[y^*] = \frac{kp}{1-p}, \quad \text{var}(y^*) = \frac{kp}{(1-p)^2}$$

As an extended example for this section, we will consider the effect of desired starting salary on the number of job applications required to get an offer. Let y_i^* denote the number of rejected job applications for person i before she receives her first job offer, while x_i represents her desired salary. In this example, $k = 1$ and p represents the probability that any given job application will lead to an offer. Therefore, the probability mass function will be in this form:

$$g(y^*) = p \cdot (1-p)^{y^*}$$

The parameter p must depend on x_i and its coefficients. A good choice to model p based on x_i would be to use a sigmoid function, which guarantees $p \in [0, 1]$:

$$p(x_i, \beta) = \frac{\exp(\beta_0 + \beta_1 x_i)}{\exp(\beta_0 + \beta_1 x_i) + 1}$$

For our job application example, we would anticipate that a higher desired starting salary (higher x_i) would make one less likely to receive an offer (lower $p(x_i, \beta)$). The partial derivative of $p(x_i, \beta)$ with respect to x is as follows:

$$\frac{\partial p}{\partial x_i} = \beta_1 \cdot \frac{\exp(\beta_0 + \beta_1 x_i)}{(\exp(\beta_0 + \beta_1 x_i) + 1)^2}$$

Given the negative relationship between x_i and p, $\frac{\partial p}{\partial x} < 0$ and we would expect $\beta_1 < 0$.

By combining the probability mass function with the definition of p, we get our likelihood function for any given observation:

$$\Pr(y_i^*|x_i; \beta) = f(y_i^*|x_i; \beta) = p(x_i, \beta) \cdot [1 - p(x_i, \beta)]^{y_i^*} = \frac{\exp(\beta_0 + \beta_1 x_i)}{(\exp(\beta_0 + \beta_1 x_i) + 1)^{y_i^* + 1}}$$

With any job search data, we would expect right-censoring, as some individuals surveyed have not yet received a job offer. Therefore, we would expect our censoring to have a different threshold for each individual (as opposed to a universal threshold). If y_i^* denotes the total (theoretical) number of job applications a person submits before receiving an offer, our observed variable, y_i, is given by $y_i = \min\{y_i^*, T_i\}$, where T_i is the idiosyncratic thresholding point. If $y_i = T_i$, person i has not yet received an offer at the time of the survey. To simulate job search data, we created 10000 data points and used the following values for x_i and β:

$$x_i \sim \text{Uniform}(30000, 130000)$$

$$\beta = [1, -4e-5]$$

x_i represents the salary expectations of 10000 applicants, ranging between \$30000 and \$130000. As mentioned above, we expect $\beta_1 < 0$, so we fixed β to match those expectations. With these values of x_i and β, we calculate $p(x_i, \beta)$. Then, we constructed y_i^* by drawing from the negative binomial distribution with parameters $p(x_i, \beta)$ and $k = 1$. The censoring was applied by randomly selecting 30% from y_i^* and censoring them by $y_i = y_i^* \cdot u_i$, where $u_i \sim$ Uniform(0,1). For uncensored values, $y_i = y_i^*$. The data were generated using Python's numpy and scipy.stats libraries. An example of the generated data is shown below:

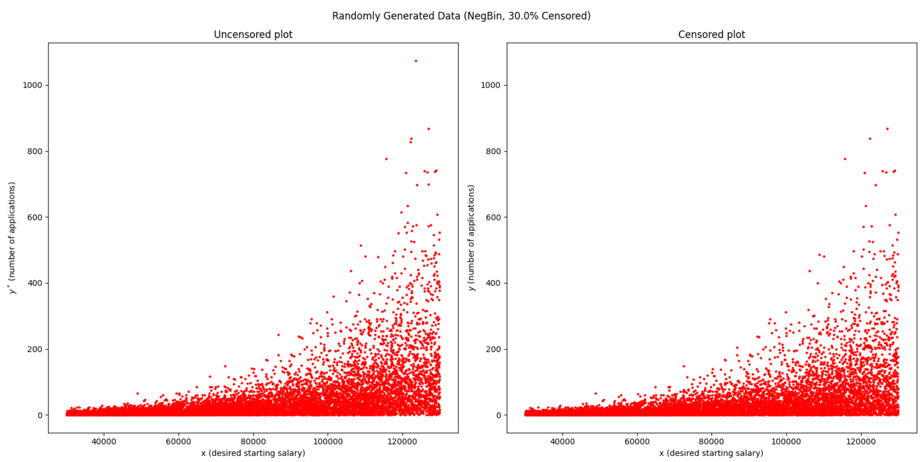

Figure 2: The left graph shows the uncensored data, while the right graph applies censoring to 30% of the data

2.3 Poisson

The Poisson distribution is commonly used to model count data, where outcomes represent the number of occurrences of an event within a fixed interval. Its probability mass function (PMF) is given by:

$$\psi(y_i|\lambda) = \frac{\lambda^{y_i} e^{-\lambda}}{y_i!},$$

where $\lambda > 0$ is the mean and variance of the distribution.

The Poisson distribution is particularly useful in econometrics for modeling discrete events such as the number of sales, arrivals, or clicks. Due to its versatility and simplicity, it often serves as the foundation for more advanced models, including those accounting for censoring or truncation. In this study, we explore the application of a censored Poisson model for cases where outcomes are restricted to a maximum value.

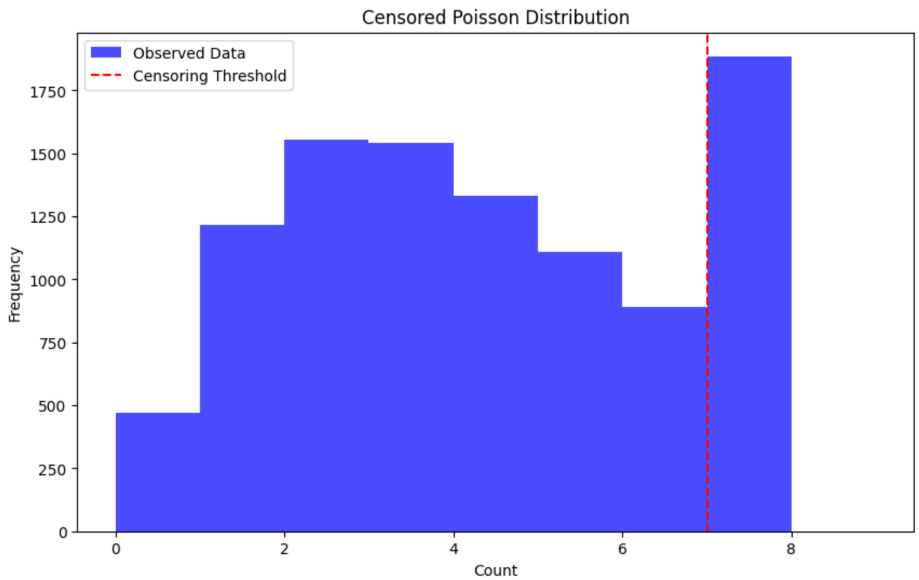

Figure 3: A sample of 10,000 observations drawn from the censored Poisson process. Censored values are capped at 7.

To simulate a realistic censored Poisson process, we generate data as follows. Let x_i represent the independent variable, drawn from a uniform distribution over the interval $[0, 2.5]$:

$$x_i \sim \text{Uniform}(0, 2.5).$$

The mean of the Poisson distribution for each observation is determined by a log-linear function of x_i:

$$\lambda_i = e^{\beta_0 + \beta_1 x_i},$$

where the true parameter values are set as $\beta_0 = 1$ and $\beta_1 = -4 \times 10^{-5}$.
The dependent variable y_i is then sampled from a Poisson distribution with mean λ_i:

$$y_i \sim \text{Poisson}(\lambda_i).$$

To introduce censoring, values of y_i greater than 7 are truncated:

$$y_i^* = \min(y_i, 7).$$

This censoring creates a dataset with a clear upper limit on observed outcomes, commonly encountered in empirical settings.
The true mean effect of x_i on y_i is $\partial \mathbb{E}[y_i]/\partial x_i = \beta_1 e^{\beta_0 + \beta_1 x_i}$. This will serve as the benchmark for evaluating the performance of different estimation methods.

3 Econometric methods

For each distribution, we will use Maximum Likelihood Estimation (MLE) to estimate the parameters. The likelihood functions for each distribution is the following:

In order for the MLE estimators to be consistent, the following conditions must be met:

- The sample likelihood functions converge uniformly in probability to the population expectation: This assumption holds for each distribution, due to the law of large numbers.
- The population expectation likelihood function is maximized at the true β: This assumption holds for each distribution, due to the likelihood principle.
- The set of reasonable β values is compact: By arbitrarily defining a compact set around each chosen β, this assumption is met.
- The population expectation likelihood function is continuous in b: This assumption holds for each distribution.

3.1 Gamma

For our censored Gamma distribution, we will perform a MLE estimate. Its log likelihood function is the following:

$$\ell(\alpha, \beta) = \frac{1}{n}\sum_{i=1}^{n}\Big[\log\psi(y_i|x_i^\top\alpha, \beta)\mathbf{1}(y_i < T) + \log\Psi(y_i|x_i^\top\alpha, \beta)\mathbf{1}(y_i = T)\Big]$$

Where Ψ is the survival function (1−cdf) of the corresponding Gamma distribution. We maximize this expression to estimate α and β. Note also that the estimated mean effect will be $\frac{\partial}{\partial x_{1i}}E[y_i] = \hat{\alpha}_1 \cdot \hat{\beta}$. From the delta test, the standard error of this estimate will be $\sqrt{\frac{-J^\top H^{-1} J}{n}}$ where $J = (0, \beta, \alpha_1)^\top$ and

$$H = -E\left[\begin{bmatrix} a^2 x_i x_i^\top & abx_i \\ abx_i^\top & b^2 \end{bmatrix}\right]\mathbf{1}(y_i < T) - E\left[\left(\frac{\psi(y_i|x_i^\top\alpha, \beta)}{\Psi(y_i|x_i^\top\alpha, \beta)}\right)^2 \begin{bmatrix} x_i x_i^\top & x_i \\ x_i^\top & 1 \end{bmatrix}\right]\mathbf{1}(y_i = T)$$

With $a = -\log(\beta)x_i - \frac{1}{\Gamma(x_i^\top\alpha)}x_i + \log(y_i)x_i$ and $b = -\frac{x_i^\top\alpha}{\beta} + \frac{y_i}{\beta^2}$.

3.2 Negative Binomial

Using our censored variable y_i, our likelihood function is

$$f(y_i|x_i;\beta) = \begin{cases} \phi(y_i, p(x_i, \beta)) & y_i < T_i \\ S(y_i, p(x_i, \beta)) & y_i = T_i \end{cases}$$

where $\phi(y_i, p) = p \cdot (1-p)^{y_i}$ is the pmf of the $k = 1$ negative binomial distribution, and $S(y_i, p) = 1 - \sum_{a=0}^{y_i} p \cdot (1-p)^a$ is the survival function.

Using maximum likelihood estimation, we can estimate β_0 and β_1 by maximizing the following log-likelihood function:

$$\ell(b|x,y) = \frac{1}{n}\sum_{i=1}^{n}[\mathbb{1}(Y_i < T_i)\cdot \ln[\phi(y_i, p(x_i, b))] + \mathbb{1}(Y_i = T_i)\cdot \ln[S(y_i, p(x_i, b))]]$$

The maximum likelihood estimates are found by maximizing this function, or equivalently, minimizing the negative.

$$\hat{b} = \arg\min -\ell(b|x,y)$$

3.3 Poisson

To estimate the parameters of the censored Poisson process, we employ a maximum likelihood approach tailored to account for censoring. The log-likelihood function for the censored Poisson model is given by:

$$\ell(\lambda) = \sum_{i=1}^{n} I_{(i\in T)}\left[y_i \log(\lambda) - \lambda - \log(y_i!)\right] + \sum_{i=1}^{n} I_{(i\notin T)} \log\left(1 - e^{-\lambda}\sum_{k=0}^{T}\frac{\lambda^k}{k!}\right),$$

.

The first term corresponds to the likelihood for uncensored observations, while the second term accounts for the probability mass beyond the censoring threshold for censored observations. The parameters β_0 and β_1 are estimated by maximizing this likelihood function.

4 Results

4.1 Gamma

When testing the results of the censored Gamma MLE model, we compared its results to a variety of different models. As seen in Figure 4, the other methods all underestimate the true effect of the program while the Gammit estimates appear approximately unbiased.

Moreover, the censored Gamma model is an MLE estimate, making it consistent, efficient, and asymptotically normal.

4.2 Negative Binomial

Minimizing the Negative Binomial likelihood function numerically, we obtained the following results:

$$\hat{b} = [7.900e-01, -4.007e-05]$$

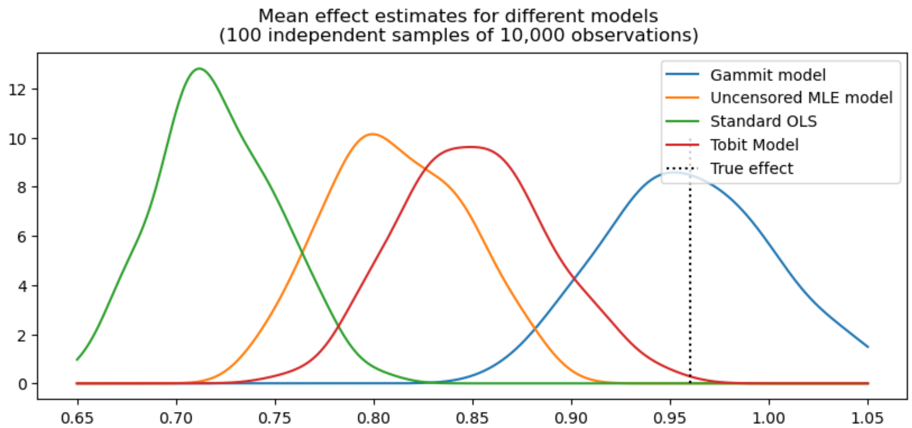

Figure 4: This figure compares the estimated distribution of point estimates from various models applied to the Gamma data. 100 independent samples were drawn each with 10,000 observations. Compare these estimate distributions with the true value represented by the dotted line.

In the maximum likelihood framework, $Var(\hat{b}) = -H(\hat{b})^{-1}$, where H is the Hessian of our log likelihood function. $H(\hat{b})$ was calculated numerically and gave us a variance matrix of:

$$Var(\hat{b}) = -H(\hat{b})^{-1} = \begin{bmatrix} 1.592e+01 & -1.707e-04 \\ -1.707e-04 & 2.034e-09 \end{bmatrix}$$

4.2.1 Marginal Effects

To observe the marginal effect of x on y, we must use the expected value

$$\mathbb{E}[y_i|x_i, \beta] = \frac{1-p(x_i,\beta)}{p(x_i,\beta)} = \frac{1}{\exp(\beta_0+\beta_1 x)} = \exp(-(\beta_0+\beta_1 x))$$

Deriving this with respect to x_i, we get:

$$\frac{\partial \mathbb{E}[y_i|x_i,\beta]}{\partial x_i} = -\beta_1 \cdot \exp(-(\beta_0+\beta_1 x))$$

If β_1 is negative (like in our example), then $\frac{\partial \mathbb{E}[y_i|x_i,\beta]}{\partial x_i} > 0$. This aligns with our intuition; a higher desired starting salary means a higher expected number of applications before an offer is extended.

4.2.2 Comparison with the Tobit Model

The Tobit model assumes that the data is normally distributed such that $y_i^* \sim N(\tilde{\beta}_0 + \tilde{\beta}_1 x_i, \sigma^2)$, for estimated parameters $\tilde{\beta}$ and σ^2. The expected value of y is given by $\mathbb{E}[y_i] = \tilde{\beta}_0 + \tilde{\beta}_1 x_i$. In this model, the marginal effect, $\tilde{\beta}_1$ is constant.

The graph below compares the marginal effects: one with the Negative Binomial underlying distribution, and the Tobit model. While the Tobit model's effect is constant (independent of x), the Negative Binomial method's marginal effect increases as x increases.

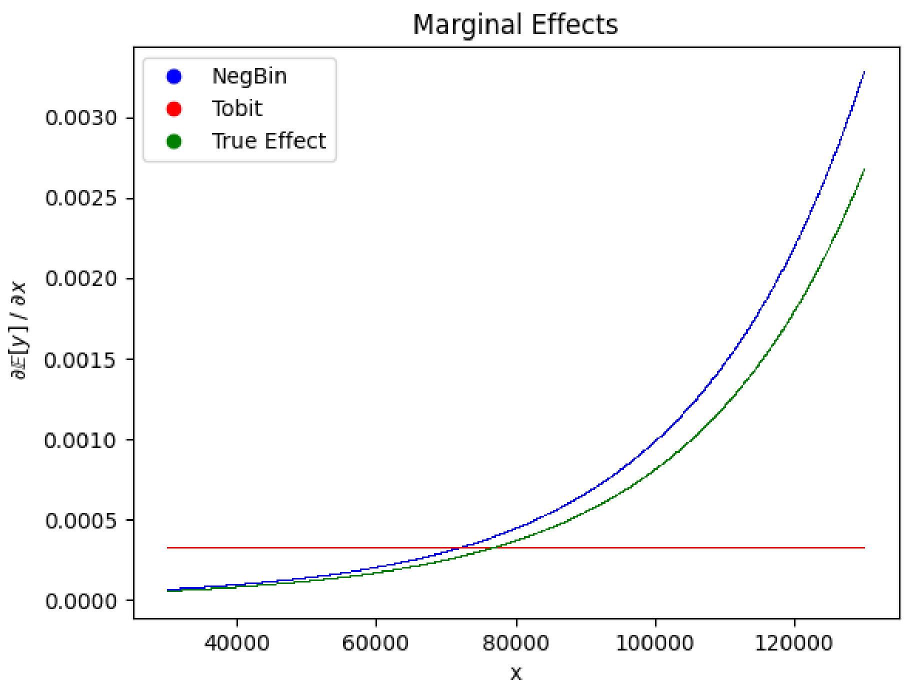

Figure 5: In this example, the estimated marginal effect overestimates the true effect, while the Tobit model assumes a constant marginal effect

4.2.3 Consistency

To test whether the negative binomial estimate $\hat{b}_1$ is consistent, we can repeat this test for increasing values of n and perform a Wald test. Because we have the true parameter β_1, we can use the following hypotheses:

$$H_0 : \hat{b_1} = \beta_1$$

$$H_a : \hat{b_1} \neq \beta_1$$

The results of these tests are summarized in the table below:
The table shows that as n increases, $\hat{b}_1$ becomes more accurate and the standard errors shrink. The p value remains high as n increases. This indicates that $\hat{b}_1$ consistently estimates β_1. The graph below also compares the true parameter to the estimates $\hat{b}_1$.

n	β_1	$\hat{b}_1$	$se(\hat{b}_1)$	**P value**	**Reject**
10	-0.00004	-0.000060	9.8724e-06	0.0397	True
100	-0.00004	-0.000041	4.3494e-06	0.8890	False
1000	-0.00004	-0.000040	4.7897e-07	0.3889	False
10000	-0.00004	-0.000040	4.4714e-07	0.3850	False
50000	-0.00004	-0.000040	1.9510e-07	0.4595	False
60000	-0.00004	-0.000040	1.8081e-07	0.0467	True
70000	-0.00004	-0.000040	1.7068e-07	0.3438	False
80000	-0.00004	-0.000040	1.5557e-07	0.2237	False
90000	-0.00004	-0.000040	1.5014e-07	0.7616	False
100000	-0.00004	-0.000040	1.3828e-07	0.1984	False
150000	-0.00004	-0.000040	1.1286e-07	0.7770	False
200000	-0.00004	-0.000040	1.0046e-07	0.6034	False

Table 1: Summary Statistics for Negative Binomial

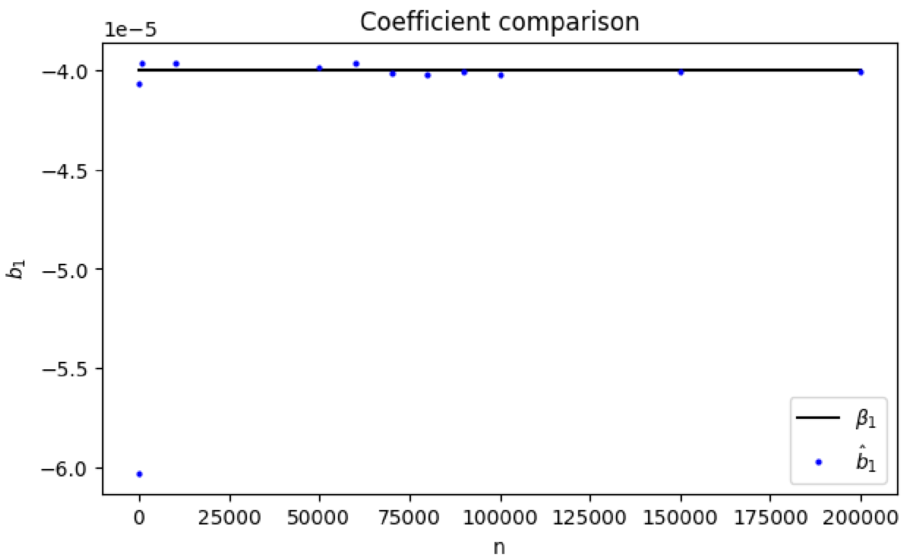

Figure 6: As n increases, $\hat{b_1}$ gets closer to the true parameter β_1.

4.3 Poisson

To evaluate the effectiveness of the censored Poisson model, we compare its performance to that of a Tobit model, which assumes normality in the dependent variable. The Poisson model aligns with the true DGP, while the Tobit model introduces a misspecification.

Table 2 summarizes the estimated parameters for both models. While the Poisson estimators align closely with the true parameter values, the Tobit estimators are biased, particularly for β_1, due to the incorrect assumption of normality.

As seen in Figure 7, the Poisson model produces estimates that are approximately

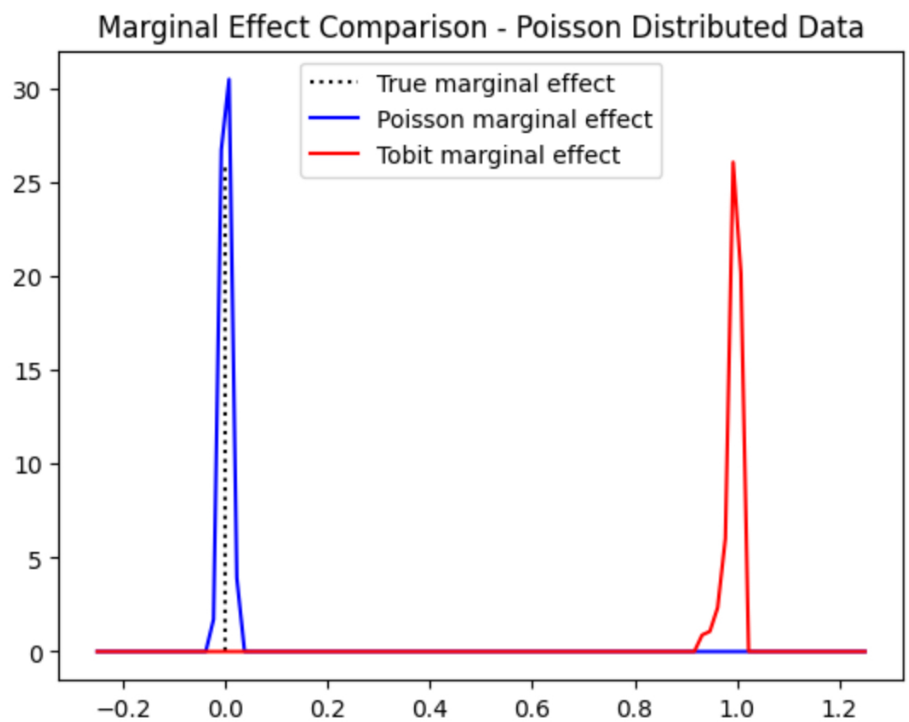

Figure 7: Distribution of estimated coefficients from 100 simulations with $10,000$ observations each. The true parameter values are marked by the dotted lines.

Table 2: Comparison of Parameter Estimates from Poisson and Tobit Models

Model	$\hat{\beta}_0$	$\hat{\beta}_1$	Additional Parameters
True Values	1	-0.00004	-
Poisson Estimators	1.0011	0.0070	-
Tobit Estimators	1.8454	1.0017	1.0000 (censoring parameter)

unbiased and consistent, closely aligning with the true parameter values. In contrast, the Tobit model underestimates the slope parameter due to its incorrect assumption of normality.

5 Conclusion

The models that we have presented here provide alternatives to the Tobit model and Quantile regression. Like all MLE estimators, they require the strong assumption that we know the conditional distribution of the data. However, as we have mentioned, these distributions have theoretical motivations that make them plausible in many real-life situations. While they will not apply in most situations, these models add a tool to the econometrician's toolkit.

6 References

Tobin, J. (1958). Estimation of Relationships for Limited Dependent Variables. Econometrica, 26(1), 24. https://doi.org/10.2307/1907382

Powell, J. L. (1984). "Least Absolute Deviations Estimation for the Censored Regression Model." Journal of Econometrics, 25, 303–325.