Effects of Ignoring Truncation in Poisson Count Models

Abdalhalim M. Ali Suaiee* Trent L. Lalonde†

Abstract
Count response data situations arise often in practice, and are typically modeled using a generalized linear model with a Poisson response distribution. A Poisson distribution imposes the assumption that the data are observed on the interval of all non-negative integers. However, practical applications often involve restrictions that reduce the domain of possible response values. Such data are referred to as "truncated" count responses. While zero-truncated data are well-recognized and often accounted for, little attention has been paid to general left-truncation, right-truncation, or double-truncation of counts. It is useful to be aware of the consequences of misspecification of a model such that truncation of any type is ignored. In this paper we compare model performance when truncation is not accounted for, when truncation is partially accounted for, and when truncation is completely accounted for. All three cases of left-truncation, right-truncation, and double-truncation are investigated.

Key Words: Double-Truncation; Model Misspecification; Poisson Regression; Truncated Counts

1. INTRODUCTION

Count response data are commonly encountered in statistical applications (Cameron and Trivedi, 2001), and consequently have received a significant amount of attention in the generalized linear model literature (Cameron and Sohansson, 1997), (Cameron and Trivredi, 1998) and (Cameron and Trivedi, 2001). Many practical applications involving count data have restricted domains of observable counts. For example, studies of hospital length of stay typically sample current or recent hospital patients, eliminating the possibility of an observation of zero days. Studies of the number of occurrences of some event per month or per year, for example usage of public transportation, exclude the possibility of more events than days considered. In both cases the count data are said to be truncated.

Left-truncated count data occur when all counts must be greater than a given value, or $Y_i \geq l$ for some count response variables Y_i and lower-bound l. Right-truncated data occur when all counts must be less than a given value, or $Y_i \leq r$ for some upper bound r. Double-truncated count data occur when all counts are restricted to a given interval, or $l \leq Y_i \leq r$ (Moore, 1954), (Creel and Loomis, 1990), (Brannas, 1992) and (Cameron and Trivedi, 2001). The case of left-truncated count data has received considerable attention in the literature, specifically with respect to the special case of zero-truncated counts (Grogger and Carson, 1991) and (Winkelmann, 2008). It is well documented that applying a model that does not account for zero-truncation can lead to the effects of overdispersion (Gurmu, 1991). Left-truncation at zero has been known as the most common form of truncation (Winkelmann, 2008).

There has been less attention paid to the cases of right-truncation (Cameron and Trivedi, 2001) and double-truncation (Cohen, 1954) for count responses. One possible explanation is that left-truncation is simply more common than right-truncation. In this paper we thoroughly characterize the effects of ignoring truncation of any type within a count regression model. In section 2 we describe the models and estimation methods of all three types of

*Department of Statistics, University of Zawia, Zawia City, Libya
†Department of Applied Statistics and Research Methods, University of Northern Colorado, Greeley, CO, USA

truncation. In section 3 we provide a simulation study to empirically evaluate the effects of ignoring each type of truncation or of partially ignoring double-truncation within count regression models, considering parameter estimate bias, standard errors, and hypothesis testing power and type I error rates. Some concluding remarks are made in section 4.

2. POISSON REGRESSION MODELS

2.1 Ordinary Poisson Regression Models

The Poisson regression model is the standard approach to count data analysis. The Poisson regression model is derived from the Poisson distribution by parameterizing the relation between the mean parameter λ and covariates $\boldsymbol{x}$ (Cameron and Trivedi, 2001). Consider a sample of n responses which can be treated as realizations of independent Poisson random variables, $Y_1, ..., Y_n$, with $Y_i \sim Poi(\lambda_i)$, and suppose that the mean λ_i depends on a vector of explanatory variables $\boldsymbol{x_i}$. The standard assumption is to use the exponential mean parametrization,

$$\lambda_i = exp(\boldsymbol{x}_i^T \boldsymbol{\beta}),$$

where $i = 1, 2, ..., n$. The distribution of Y_i will be conditional on the regressors $\boldsymbol{x}$, so the conditional distribution is $Y_i|\boldsymbol{x_i} \sim^{iid} Poi(\lambda_i)$, and the probability mass function is

$$f(Y_i|\lambda_i) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}, \tag{1}$$

where $y_i = 0, 1, 2, \ldots$. The conditional mean and variance of the distribution are given by

$$\lambda(\boldsymbol{X_i}) = E[Y_i|\boldsymbol{X_i}] = Var(Y_i|\boldsymbol{X_i}) = \exp(\boldsymbol{x_i^T}\boldsymbol{\beta}).$$

The coefficients $\boldsymbol{\beta}$ can be interpreted as average proportionate change in $E[Y_i|\boldsymbol{x_i}]$ for a unit change in $\boldsymbol{x_i}$, (Grogger and Carson, 1991). Using Equation (1) and the assumption that the responses $(Y_i|\boldsymbol{x_i})$ are independent, the likelihood function is

$$L(\boldsymbol{\beta}; y_i) = \prod_{i=1}^{n} \frac{e^{\lambda_i}\lambda_i^{y_i}}{y_i!}.$$

Thus, the log-likelihood function is

$$\begin{aligned} l(\beta; \boldsymbol{y_i}) &= \sum_{i=1}^{n} (y_i \ln \lambda_i - \lambda_i - ln(y_i!)) \\ &= \sum_{i=1}^{n} \left(y_i \boldsymbol{x_i^T}\boldsymbol{\beta} - \exp(\boldsymbol{x_i^T}\boldsymbol{\beta}) - ln(y_i)! \right). \end{aligned} \tag{2}$$

To estimate the parameters using the maximum likelihood estimation (MLE) method, we take derivatives of Equation (2) with respect to $\boldsymbol{\beta}$. Setting the derivatives to zero, the first order maximum likelihood condition is

$$\sum_{i=1}^{n} \left(y_i - \exp(\boldsymbol{x}_i^T\boldsymbol{\beta}) \right) \boldsymbol{x}_i = 0.$$

The advantages of using MLE are that maximum likelihood provides consistent estimators, are asymptotically normally distributed, and are asymptotically minimum variance unbiased estimators as the sample size increases. The most common hypothesis testing follows by using Wald tests Greene (2010).

2.2 Truncated Poisson Regression Models

When count data are observed over only part of the range of the response variable, then the data are called truncated count data (Cameron and Trivedi, 2001). Examples of truncated data arise in many contexts. The length of stay at a hospital is an example of zero-truncated data or generally left-truncated data, as all patients included will necessarily stay at the hospital for at least one day. Modeling the time between events reported through follow-back are right-truncated by the amount of time requested in the follow-back process. Double-truncation would occur when modeling monthly hospital length of stay, as the number of days per month provides an upper limit on counts. As another example of double-truncated data, observations on the number of items purchased when there is a limit per customer would exclude the zero class (since customers who do not buy the item will not be identified), and would exclude observations above the limit.

2.3 Left-Truncated Poisson Model at $k = l$

Count data for which $Y_i < l$ counts cannot be observed are called left-truncated count data at $k = l$. For the Poisson probability function, a model for counts truncated on the left at the value $k = l$ can be posited as

$$Pr(Y_i = y_i | Y_i \geq l) \quad = \quad \frac{\lambda_i^{y_i}}{\left[e^{\lambda_i} - \sum_{k=0}^{l-1} \frac{\lambda_i^k}{k!}\right] y_i!}.$$

In the case of the left-truncated Poisson model, the mean and variance of the distribution are readily shown to be

$$E[Y_i | x_i, Y_i \geq l] \quad = \quad \frac{\lambda_i e^{-\lambda_i}}{1 - \sum_{k=0}^{l-1} e^{-\lambda_i} \frac{\lambda_i^k}{k!}} \sum_{y_i = l}^{\infty} \frac{\lambda_i^{y_i - 1}}{(y_i - 1)!}, \tag{3}$$

and

$$Var(Y_i | X_i, Y_i \geq l) = \frac{\lambda_i e^{-\lambda_i}}{1 - F_p(l-1)} \left[\lambda_i \sum_{y_i = l+1}^{\infty} \frac{\lambda_i^{y_i - 2}}{(y_i - 2)!} + \sum_{y_i = l}^{\infty} \frac{\lambda_i^{y_i - 1}}{(y_i - 1)!}\right] - \left[\frac{\lambda_i e^{-\lambda_i}}{[1 - F_p(l-1)]} \sum_{y_i = l}^{\infty} \frac{\lambda_i^{y_i - 1}}{(y_i - 1)}\right]^2, \tag{4}$$

where $F_p(l-1)$ indicates the CDF of the ordinary Poisson distribution, evaluated from 0 to $l-1$. Equations (3) and (4) show that the mean of the left-truncated random variable exceeds the corresponding mean of the untruncated distribution model, whereas the variance of left-truncated random variable is smaller than the corresponding variance of the untruncated distribution model. (Cameron and Trivredi, 1998) expressed the relationship between the truncated mean and untruncated distribution mean as

$$E[y_i | y_i \geq l] = E[y_i] + \delta_i,$$

where $\delta_i > 0$ is the difference between the truncated and untruncated means. The joint likelihood and log-likelihood functions for the left-truncated Poisson model parameters are

$$L(\boldsymbol{\beta}; y_i) \quad = \quad \prod_{i=1}^{m} \frac{\lambda_i^{y_i}}{y_i!} \left[\frac{1}{e^{\lambda_i} - \sum_{k=0}^{l-1} \lambda_i^k}\right],$$

and

$$l(\boldsymbol{\beta}; y_i) \quad = \quad \sum_{i=1}^{m} \left[y_i \boldsymbol{x_i^T} \boldsymbol{\beta} - \ln(y_i!) - \ln\left(e^{\lambda_i} - \sum_{k=0}^{l-1} \frac{\lambda_i^k}{k!}\right)\right]. \tag{5}$$

The left-truncated Poisson maximum likelihood estimators (LTMLE's) can be obtained from solving the first order condition of maximum likelihood, obtained from Equation (5),

$$\sum_{i=1}^{m} x_{ih}\left[y_i - \frac{e^{\boldsymbol{x_i^T}\boldsymbol{\beta}}e^{e^{\boldsymbol{x_i^T}\boldsymbol{\beta}}} - \sum_{k=0}^{l-1}\frac{ke^{\boldsymbol{x_i^T}\boldsymbol{\beta}k}}{k!}}{e^{e^{\boldsymbol{x_i^T}\boldsymbol{\beta}}} - \sum_{k=0}^{l-1}\frac{e^{\boldsymbol{x_i^T}\boldsymbol{\beta}k}}{k!}}\right] = 0,$$

with solutions obtained through nonlinear optimization.

2.3.1 Right-Truncated Poisson Model at $k = r$

Count data for which $Y_i > r$ counts cannot be observed are called right-truncated count data at $k = r$. Estimating a Poisson regression model without considering this truncation will cause biased estimates of the parameter vector $\boldsymbol{\beta}$ and erroneous inferences will be drawn (Liu and Pitt, 2012). For the Poisson probability function, a model for count data truncated on the right at $k = r$ can be posited as

$$Pr(Y_i = y_i|Y_i \leq r) = \frac{\lambda_i^{y_i}}{\left(\sum_{k=0}^{r}\frac{\lambda_i^k}{k!}\right)y_i!},$$

where $i = 1, \ldots, m$. In the case of the right-truncated Poisson model, the mean and variance of the distribution are readily shown to be

$$E[Y_i = y_i|x_i, y_i \leq r] = \frac{\lambda_i}{\sum_{k=0}^{r}\frac{\lambda_i^k}{k!}}\sum_{y_i=1}^{r}\frac{\lambda_i^{y_i-1}}{(y_i-1)!}, \tag{6}$$

and

$$Var(Y_i|X_i, Y_i \leq r) = \frac{1}{\sum_{k=0}^{r}\frac{\lambda_i^k}{k!}}\left[\lambda_i^2\sum_{y_i=2}^{r}\frac{\lambda_i^{y_i-2}}{(y_i-2)!} + \lambda_i\sum_{y_i=1}^{r}\frac{\lambda_i^{y_i-1}}{(y_i-1)!}\right] - \left[\frac{\lambda_i}{\sum_{k=0}^{r}\frac{\lambda_i^k}{k!}}\sum_{y_i=1}^{r}\frac{\lambda_i^{y_i-1}}{(y_i-1)!}\right]^2. \tag{7}$$

It can be seen from Equations (6) and (7) that the right-truncated Poisson distribution results in a smaller mean and variance compared to the standard Poisson distribution (Cameron and Trivedi, 2001). The joint likelihood and log-likelihood functions for the right-truncated Poisson model parameters are

$$L(\boldsymbol{\beta}; y_i) = \prod_{i=1}^{m}\frac{\lambda_i^{y_i}}{\left(\sum_{k=0}^{r}\frac{\lambda_i^k}{k!}\right)y_i!},$$

and

$$l(\boldsymbol{\beta}; y_i) = \sum_{i=1}^{m}\left[y_i\boldsymbol{x_i^T}\boldsymbol{\beta} - \ln(y_i!) - \ln\left(\sum_{k=0}^{r}\frac{e^{\boldsymbol{x_i^T}\boldsymbol{\beta}k}}{k!}\right)\right].$$

To obtain the right-truncated Poisson maximum likelihood estimators (RTMLE's), solve the first order condition of maximum likelihood for $\boldsymbol{\beta}$,

$$\sum_{i=1}^{m} x_{ih}\left[y_i - \frac{\sum_{k=0}^{r}\frac{ke^{\boldsymbol{x_i^T}\boldsymbol{\beta}k}}{k!}}{\sum_{k=0}^{r}\frac{e^{\boldsymbol{x_i^T}\boldsymbol{\beta}k}}{k!}}\right] = 0,$$

with solutions obtained through nonlinear optimization.

2.4 Double-Truncated Poisson Regression Models

Double-truncated Poisson data are a combination of the left-truncated and right-truncated Poisson data. From (Cohen, 1954), the probability mass function for the double-truncated Poisson random variable y is given by

$$Pr(Y_i = y_i | x_i, l \leq y_i \leq r) = \frac{\lambda_i^{y_i}}{y_i! \sum_{k=l}^{r} \frac{\lambda_i^k}{k!}}, \tag{8}$$

where $\lambda_i = \exp(\boldsymbol{x_i^T \beta})$, and l and r are the lower end and the upper end, respectively, of the interval in which each Y_i can be observed. In the case of the double-truncated Poisson model, the mean and variance of the distribution are readily shown to be

$$E[Y_i = y_i | X_i, l \leq y_i \leq r] = \frac{\sum_{k=l}^{r} \frac{\lambda_i^k}{(k-1)!}}{\sum_{k=l}^{r} \frac{\lambda_i^k}{k!}},$$

and

$$Var(Y_i = y_i | X_i, l \leq y_i \leq r) = \frac{\sum_{k=l}^{r} \frac{k\lambda_i^k}{(k-1)!}}{\sum_{k=l}^{r} \frac{\lambda_i^k}{k!}} - \left(\frac{\sum_{k=l}^{r} \frac{\lambda_i^k}{(k-1)!}}{\sum_{k=l}^{r} \frac{\lambda_i^k}{k!}} \right)^2.$$

While the regular Poisson model typically encounters difficulty due to the assumed equality of mean and variance, the mean and variance of the doubly-truncated Poisson model are characteristic of underdispersion where the variance is less than the mean. Testing for overdispersion must now take this assumption into account (Gurmu and Trivedi, 1992). This assumed inequality also provides an intuitive reason as to why fitting a regular Poisson model to truncated data is a fundamental misspecification. The regular Poisson assumes greater variance than should be expected. The joint likelihood and log-likelihood functions for the double-truncated Poisson model parameters are

$$L(\boldsymbol{\beta}; y_i) = \prod_{i=1}^{m} \frac{\lambda_i^{y_i}}{y_i! \sum_{k=l}^{r} \frac{\lambda_i^k}{k!}},$$

and

$$l(\boldsymbol{\beta}; y_i) = \sum_{i=1}^{m} \left[y_i \boldsymbol{x_i^T \beta} - \ln(y_i!) - \ln\left(\sum_{k=l}^{r} \frac{\lambda_i^k}{k!} \right) \right]. \tag{9}$$

The double-truncated Poisson maximum likelihood estimators (DTMLE's) are obtained by solving the following first order condition of Equation (9),

$$\sum_{i=1}^{m} x_{ih} \left[y_i - \frac{\sum_{k=l}^{r} \frac{k e^{\boldsymbol{x_i^T \beta} k}}{k!}}{\sum_{k=l}^{r} \frac{e^{\boldsymbol{x_i^T \beta} k}}{k!}} \right] = 0,$$

with solutions obtained through nonlinear optimization.

2.5 Model Misspecification

Misspecification of the distribution of truncated data implies that both the first and second conditional moments will also be misspecified. Cameron and Trivedi (1998) verified that misspecification of the response distribution will result in inconsistent estimators of the mean parameters $\boldsymbol{\beta}$. Liu and Pitt (2012) found that ignoring the left-truncation in count data leads to bias in the parameter estimates. Similarly, Kalbfleisch and Lawles (1991) considered issues with modeling right-truncated data. Baud and Frachot (2002) and also Creel and Loomis (1990) considered truncated modeling in general, and showed that ignoring truncation in general leads to biased estimators, with the bias higher than expected. In this study the purpose is to empirically investigate the consequences of misspecifying the response distribution for truncated data of all types.

3. SIMULATIONS

In this section, we present a simulation study to evaluate the impact of incorrectly specifying Poisson regression models in the context of truncated count data. The performances of each of the ordinary Poisson regression model (PRM), the left-truncated Poisson regression model (LTPRM), the right-truncated Poisson regression model (RTPRM), and the double-truncated Poisson regression model (DTPRM) are examined based on bias and standard errors of parameter estimates, and on the power and Type I error rates of Wald hypothesis tests.

The simulation consists of the following process. Independent variables are randomly generated according to a uniform $(0, 3)$ distribution. Sample sizes of 30, 50, 100, 150, and 200 will be considered, and 10000 replicates will be used for each condition. Results presented for all models include average of parameter estimates, empirical parameter estimate standard errors, proportion correctly detected significant as an estimate of power, and proportion incorrectly detected significant as an estimate of type I error rate. All hypothesis tests are performed using Wald statistics for individual parameters.

First, left-truncated responses will be simulated using a single predictor, with responses truncated at $l = 2$. Two models will be fit using these data: an appropriate LTPRM and a PRM ignoring the left-truncation in the response. Second, right-truncated responses will be simulated using a single predictor, with responses truncated at $r = 8$. Two models will be fit using these data: an appropriate RTPRM and a PRM ignoring the right-truncation in the response. Finally, double-truncated responses will be simulated using a single predictor, with responses truncated according to the interval $l = 2$ and $r = 8$. Four models will be fit using these data: an appropriate DTPRM, a LTPRM that ignores the right-truncation in the response, a RTPRM that ignores the left-truncation in the response, and a PRM that ignores all truncation in the response. Each model will include two independent variables, one to evaluate power and the other to evaluate type I error rate.

3.1 Simulation Results for Left-Truncated Responses

In this section, results are presented for models fit to the left-truncated responses, where data were simulated with true values $\beta_1 = 0.4$ and $\beta_2 = 0$. Table 1 shows the results for both models. For the LTPRM with appropriately defined left-truncation, the bias for both parameter estimates decreases with sample size, and power increases with sample size. Type I error rates remain close to the nominal value of 0.05, with some decrease for large samples.

For the PRM ignoring the left-truncation in the data, the power increases with sample size, but at a lower rate than for the correctly specified model. Comparisons of model power are displayed in Figure 1. The bias in the estimate for the significant parameter remains regardless of sample size. The bias in the estimate for the non-significant parameter decreases, but the type I error rate remains consistently very close to 0, suggesting the hypothesis tests may be conservative. Comparisons of model type I error rates are displayed in Figure 2. Overall, misspecifying a PRM such that left-truncation is ignored results in biased parameter estimates and Type I error rates lower than the nominal level.

Table 1: Models for Left-Truncated Responses

Sample Size	Parameter	LTPRM					PRM				
		Estimate	S.E.	Bias	Power	Type I Error	Estimate	S.E.	Bias	Power	Type I Error
30	$\beta_1 = 0.4$	0.4165	0.2132	0.0165	0.5131		0.1677	0.1281	-0.2323	0.171	
	$\beta_2 = 0$	-0.0027	0.2006	-0.0027		0.0417	0.0005	0.1274	0.0005		0.0045
50	$\beta_1 = 0.4$	0.4048	0.1585	0.0048	0.7413		0.1689	0.0974	-0.2311	0.3766	
	$\beta_2 = 0$	-0.0032	0.1488	-0.0032		0.0476	9.4e-05	0.0967	9.4e-05		0.0027
100	$\beta_1 = 0.4$	0.4023	0.1090	0.0023	0.996		0.1698	0.0678	-0.2302	0.7884	
	$\beta_2 = 0$	-0.0016	0.1022	-0.0016		0.0498	0.0002	0.0673	0.0002		0.0032
150	$\beta_1 = 0.4$	0.4028	0.0883	0.0028	0.996		0.1697	0.0551	-0.2303	0.9491	
	$\beta_2 = 0$	-0.0008	0.0826	-0.0008		0.0478	0.0002	0.0547	0.0002		0.0029
200	$\beta_1 = 0.4$	0.4007	0.0762	0.0007	0.9994		0.1693	0.0476	-0.2307	0.9906	
	$\beta_2 = 0$	-0.0005	0.0713	-0.0005		0.0464	7.1e-05	0.0473	7.1e-05		0.0041

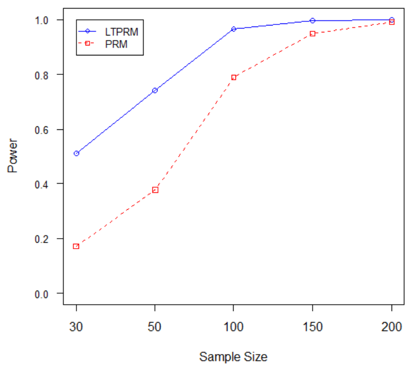

Figure 1: Power Comparisons, Left-Truncated Responses

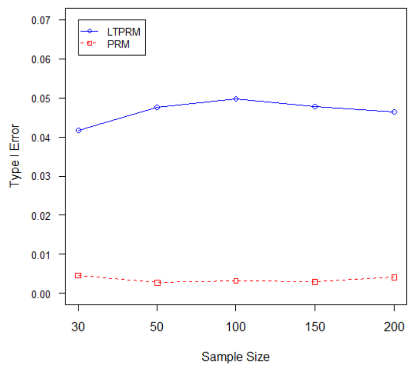

Figure 2: Type I Error Rate Comparisons, Left-Truncated Responses

3.2 Simulation Results for Right-Truncated Responses

In this section, results are presented for models fit to the right-truncated responses, where data were simulated with true values $\beta_1 = -0.18$ and $\beta_2 = 0$. Table 2 shows the results for both models. For the RTPRM with appropriately defined right-truncation, the bias in the estimate for the significant parameter decreases with increased sample size, the power increases, and the type I error rate remains near the nominal level of 0.05. The bias in the estimate for the non-significant parameter does not appear to change with sample size, but remains negligible.

For the PRM ignoring the right-truncation in the data, the bias remains in the estimates for both parameters, regardless of sample size. The power increases at a slightly lower rate as compared to the appropriately defined model, with a comparison presented in Figure 3. The type I error rate remains lower than the nominal level, with estimated values close to .02. Comparisons of type I error rates are displayed in Figure 4. Overall, misspecifying a PRM such that right-truncation is ignored results in biased parameter estimates and type I error rates that are lower than expected. However, the effects do not appear to be as severe as those observed when left-truncation is ignored.

Table 2: Models for Right-Truncated Responses

Sample Size	Parameter	RTPRM Estimate	RTPRM S.E.	RTPRM Bias	RTPRM Power	RTPRM Type I Error	PRM Estimate	PRM S.E.	PRM Bias	PRM Power	PRM Type I Error
30	$\beta_1 = -0.18$	-0.1874	0.1095	-0.0074	0.4014		-0.1259	0.0879	0.0541	0.2692	
	$\beta_2 = 0$	-1.3e-05	0.1077	-1.3e-05		0.0453	-1.9e-05	0.0879	-1.9e-05		0.0195
50	$\beta_1 = -0.18$	-0.1838	0.0824	-0.0038	0.6147		-0.1266	0.0669	0.0534	0.4719	
	$\beta_2 = 0$	-0.0009	0.0813	-0.0009		0.0510	-0.0002	0.0666	-0.0002		0.0185
100	$\beta_1 = -0.18$	-0.1824	0.0571	-0.0024	0.8969		-0.1262	0.0466	0.0538	0.8173	
	$\beta_2 = 0$	-0.0003	0.0563	-0.0003		0.0464	0.0005	0.0465	0.0005		0.0177
150	$\beta_1 = -0.18$	-0.1820	0.0463	-0.0020	0.9804		-0.1259	0.0378	0.0541	0.9478	
	$\beta_2 = 0$	-0.0005	0.0457	-0.0005		0.0477	-1.9e-05	0.0377	-1.9e-05		0.0176
200	$\beta_1 = -0.18$	-0.1811	0.0399	-0.0011	0.9961		-0.1259	0.0327	0.0541	0.9892	
	$\beta_2 = 0$	0.0007	0.0394	0.0007		0.0484	-1.1e-05	0.0326	-1.1e-05		0.0167

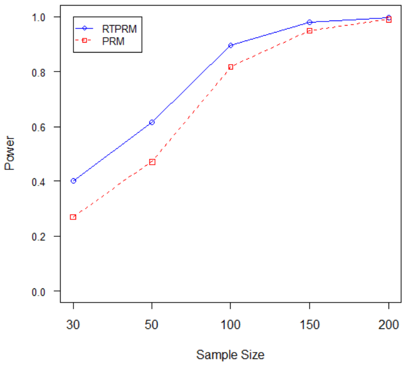

Figure 3: Power Comparisons, Right-Truncated Responses

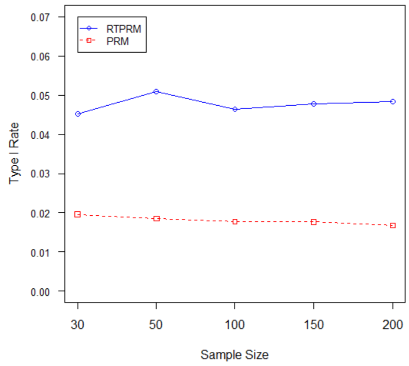

Figure 4: Type I Error Rate Comparisons, Right-Truncated Responses

3.3 Simulation Results for Double-Truncated Responses

In this section, results are presented for models fit to the double-truncated responses, where data were simulated with true values $\beta_1 = 0.4$ and $\beta_2 = 0$. Table 3 shows the results for all four models. For the DTPRM with appropriately defined double-truncation, the bias for both parameter estimates decreases with sample size, and power increases with sample size. Type I error rates remain close to the nominal value of 0.05.

For the LTPRM ignoring the right-truncation but accounting for the left-truncation in the data, the bias appears to increase in magnitude in the estimates for the significant parameter, and remain in the estimates for the non-significant parameter. The power increases with sample size and remains similar to the power for the correctly specified DTPRM. Comparisons of power for all four models are displayed in Figure 5. The type I error rate remains close to the nominal value of .05. Comparisons of type I error rates for all four models are displayed in Figure 6.

For the RTPRM ignoring the left-truncation but accounting for the right-truncation in the data, the bias in both parameter estimates remains, regardless of sample size. Power increases with sample size, but at a lower rate than the DTPRM and the LTPRM. The type I error rate is lower than the nominal level of 0.05, remaining close to 0. In fact, both the power and type I error rate for the RTPRM are similar to those of the PRM.

For the PRM ignoring both the left and right-truncation in the data, the bias in both parameter estimates remains regardless of sample size. Power increases with sample size, but at a lower rate than seen in the correctly specified model. The type I error rate is lower than the nominal level, with values remaining close to 0.

Overall, fitting a LTPRM that ignores the right-truncation of double-truncated responses results in biased parameter estimates, but has little effect on the power and type I error rates of hypothesis tests. On the other hand, fitting either a RTPRM or a PRM that ignored the left-truncation of double-truncated responses results in biased parameter estimates, lower power and type I error rates that remain below the nominal error rate. It appears that failing to account for left-truncation in count data is accompanied by more severe consequences than failing to account for right-truncation in count data.

Table 3: Models for Double-Truncated Responses

Sample Size	Parameter	DTPRM					LTPRM					RTPRM					PRM				
		Estimate	S.E.	Bias	Power	Type I Error	Estimate	S.E.	Bias	Power	Type I Error	Estimate	S.E.	Bias	Power	Type I Error	Estimate	S.E.	Bias	Power	Type I Error
30	$\beta_1 = 0.4$	0.4165	0.2129	0.0165	0.4953		0.4003	0.2131	0.0003	0.4775		0.1746	0.1331	-0.2254	0.1712		0.1677	0.1281	-0.2323	0.1481	
	$\beta_2 = 0$	0.0044	0.2065	0.0044		0.0440	-0.0010	0.2013	-0.0010		0.0397	-0.0008	0.1318	-0.0008		0.0025	0.0005	0.1274	0.0005		0.0029
50	$\beta_1 = 0.4$	0.4048	0.1629	0.0048	0.7397		0.3974	0.1590	-0.0026	0.7275		0.1733	0.1006	-0.2267	0.3697		0.1689	0.0974	-0.2311	0.3557	
	$\beta_2 = 0$	0.0017	0.1535	0.0017		0.0486	0.0015	0.1498	0.0015		0.045	-0.0011	0.0997	-0.0011		0.0037	9.4e-05	0.0967	9.4e-05		0.0038
100	$\beta_1 = 0.4$	0.4028	0.1121	0.0023	0.9644		0.3946	0.1095	-0.0054	0.9597		0.1739	0.0701	-0.2261	0.7857		0.1698	0.0678	-0.2302	0.7539	
	$\beta_2 = 0$	0.0013	0.1051	0.0013		0.0503	-0.0019	0.1027	-0.0019		0.0431	-0.0001	0.06946	-0.0001		0.0044	0.0002	0.0673	0.0002		0.0018
150	$\beta_1 = 0.4$	0.4023	0.0907	0.0028	0.9945		0.3919	0.0886	-0.0081	0.9956		0.1727	0.0569	-0.2273	0.9427		0.1697	0.0551	-0.2303	0.9414	
	$\beta_2 = 0$	0.0007	0.0849	0.00077		0.0480	-0.0003	0.0833	-0.0003		0.0459	-0.0004	0.0564	-0.0004		0.00345	0.0002	0.0547	0.0002		0.0027
200	$\beta_1 = 0.4$	0.4007	0.0781	0.0007	0.9997		0.3927	0.0765	-0.0073	0.9994		0.1735	0.0491	-0.2268	0.9901		0.1693	0.0476	-0.2307	0.9899	
	$\beta_2 = 0$	0.0003	0.0731	0.0003		0.0494	-0.0002	0.0717	-0.0002		0.0476	-0.0001	0.0487	-0.0001		0.0036	7.1e-05	0.0473	7.1e-05		0.0023

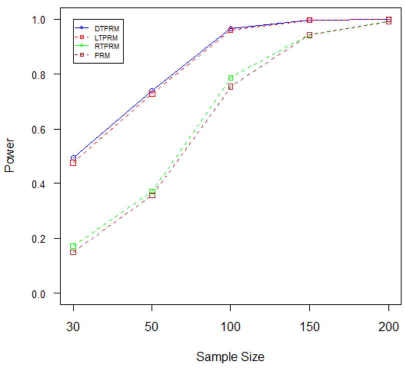

Figure 5: Power Comparisons, Double-Truncated Responses

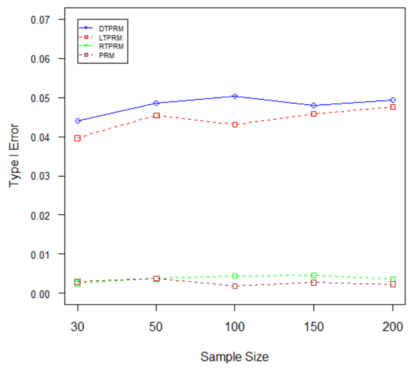

Figure 6: Type I Error Rate Comparisons, Double-Truncated Responses

4. CONCLUSIONS

Count data are common in practice, and often observed counts are truncated on the left, the right, or both. When count data are truncated it is important to properly account for the truncation in any Poisson regression model fitted to the data. While it is common for researchers to account for left-truncation in the specific context of zero-truncated counts, there is a lack of discussion in the literature of general truncation. In this paper we have presented the likelihood functions associated with each type of truncation along with the accompanying maximum likelihood processes. We have discussed the issue of incorrectly specifying a Poisson regression model such that truncation of some kind is ignored, and we have presented a simulation study to examine the empirical results of failing to properly account for truncation in count responses.

Ignoring truncation of any kind in a Poisson regression model results in biased parameter estimates and a reduction in power of hypothesis tests using Wald statistics. Power is more noticeably affected when left-truncation is ignored than when right-truncation is ignored. Similarly, type I error rates are more noticeably affected when left-truncation is ignored than when right-truncation is ignored. Specifically, when left-truncated count responses are fit with an ordinary Poisson regression model the type I error rate tends to be close to 0, far below the nominal level.

For double-truncated count responses, the effects of failing to properly account for left-truncation are more severe than the effects of failing to account for right-truncation. In fact, it appears the performance of a Poisson regression model that accounts for left-truncation but ignores right-truncation in double-truncated data is similar to the performance of an appropriately defined double-truncated Poisson regression model. While some bias in parameter estimates remains, power remains high and the type I error rates are close to the nominal level. Similarly, the performance of a model that accounts for right-truncation but ignores left-truncation is poor, comparable to an ordinary Poisson regression model that fails to account for any type of truncation. Bias remains in parameter estimates, power is reduced and type I error rates remain close to 0. It is evident that left-truncation is the more important data characteristic to account for.

This research addressed the issue of misspecifying the response distribution in Poisson regression models by ignoring types of truncation. However, we did not address the issue of specifying truncation according to an incorrect limit, or of specifying a distribution other than the Poisson such as the Negative Binomial or overdispersed Poisson distributions. The results of this study should be treated as a general warning against ignoring left-truncation in count responses, even if the data re not simply truncated at zero. However, future studies may consider specific effects of misspecifying the truncation limit or extending to generalized linear models other than the traditional Poisson model.

REFERENCES

Baud, N., Frachot, A. and Roncalli, T. (2002), *Internal Data, External Data and Consortium Data for Operational Risk Measurement: How to Pool Data Properly?*.

Brannas, K. (1992), "Limited dependent poisson regression," *The Statistician*, 41,413-423.

Cameron, A. C. and Sohansson, P. (1997), "Count Data Regression Using Series Expansions With Applications," *Journal of Applied Econometrics*, 12, 203-223.

Cameron, A. C. and Trivedi, P. K. (1998), *Regression Analysis Of Count Data*, Cambridge. UK. Cambridge University Press.

Cameron, A. C. and Trivedi, P. K. (2001), "Essentials Of Count Data Regression," *A Companion to Theoretical Econometrics*, Oxford, Blackwell.

Cohen, A. C. (1954), "Estimation Of The Poisson Parameter From Truncated Samples And From Censored Samples," *Journal of the American Statistical Association*, 49, 158-168.

Creel, M. D. and Loomis, J. B. (1990), "Theoretical and Empirical Advantages of Truncated Count Estimators for Analysis of Deer Hunting in California," *American Journal of Agricultural Economics*, 72, 434-441.

Greene, H. W. (2010), *Econometric Analysis 6ed*, Prentice Hall.

Grogger, J. T. and Carson, R. T. (1991), "Models For Truncated Counts," *Journal of Applied Econometrica*, 6, 225-238.

Gurmu, S. (1991), "Tests for Detecting Overdispersion in the Positive Poisson Regression Model," *Journal of Business and Economic Statistics*, 9, 215-222.

Gurmu, S. and Trivedi, P. K. (1992), "Overdispersion Tests for Truncated Poisson Regression Models," *Journal of Econometrics*, 9, 94-113.

Kalbfleisch, J. D. and Lawless, J. F. (1991), "Regression Models for Right Truncated Data with Applications to Aids Incubation Times and Reporting Lags," *Statistica Sinica*, 1, 19-32.

Liu, Q., Pitt, D., Wang, Y. and Wu, X. (2012), "Survival Analysis of Left Truncated Income Protection Insurance Data," *Asia-Pacific Journal of Risk and Insurance*, 7(1).

Moore, P. G. (1954), "A Note On Truncated Poisson Distributions," *Biometrics*, 10, 402-406.

Winkelmann, R. (2008), *Econometric Analysis of Count Data*, Springer-Verlag Berlin Heidelbreg.