Justice Quarterly

ISSN: 0741-8825 (Print) 1745-9109 (Online) Journal homepage: https://www.tandfonline.com/loi/rjqy20

Substantial Bias in the Tobit Estimator: Making a Case for Alternatives

Theodore Wilson, Tom Loughran & Robert Brame

To cite this article: Theodore Wilson, Tom Loughran & Robert Brame (2020) Substantial Bias in the Tobit Estimator: Making a Case for Alternatives, Justice Quarterly, 37:2, 231-257, DOI: 10.1080/07418825.2018.1517220

To link to this article: https://doi.org/10.1080/07418825.2018.1517220

Published online: 28 Dec 2018.

JUSTICE QUARTERLY
2020, VOL. 37, NO. 2, 231–257
https://doi.org/10.1080/07418825.2018.1517220

ACJS Routledge Taylor & Francis Group

Check for updates

Substantial Bias in the Tobit Estimator: Making a Case for Alternatives

Theodore Wilson[a,b], Tom Loughran[c] and Robert Brame[d]

[a]University at Albany, SUNY, Albany, New York, USA; [b]School of Criminal Justice, University at Albany, SUNY, Albany, New York, USA; [c]Department of Sociology and Criminology, Pennsylvania State University, University Park, USA; [d]University of South Carolina, Columbia, South Carolina, USA

ABSTRACT

Censored outcome data are commonly encountered in criminology. Criminologists sometimes use the tobit model to address these censored data. While tobit models make more realistic demands of censored outcome data than ordinary least squares (OLS) regression, they require the researcher to make strong distributional assumptions. When these assumptions are not met, as is often the case in criminological data, tobit models yield biased and inconsistent estimates. We seek to demonstrate this substantial bias in simulation analyses and present easily applied alternative methods. The tobit model and semiparametric alternatives for censored outcome data are applied with simulated data under varying conditions. These simulations are followed with an empirical example using sentencing data. The bias from tobit can be corrected through application of semiparametric alternatives. Criminologists should begin their analyses of censored outcome data with the least restrictive of the available models (CLAD) before progressing to more efficient, but potentially biased, estimators.

ARTICLE HISTORY
Received 27 August 2017
Accepted 22 August 2018

KEYWORDS
Censored data; tobit; CLAD; SCLS; simulations

Introduction

The evolution of knowledge within criminology has been a product of both innovative theoretical ideas and advancements in the available methodological tools to test those theoretical ideas. As criminologists have sought out new methodological tools to overcome the limitations of prior approaches, the current understanding of topics has ebbed and flowed. Nagin, Jones, Lima Passos, and Tremblay (2016) documents this progression well with regard to advances in theories and understanding of developmental processes resulting from applications of group-based trajectory modeling. The group-based trajectory modeling method allowed researchers to unpack heterogeneity in the longitudinal development of criminological phenomena of interest including offending, victimization, drug use, and a host of additional outcomes (Nagin & Odgers, 2010; Nagin & Tremblay, 2005; Nagin et al., 2016). These developments occurred as a

CONTACT Theodore Wilson thwilson@albany.edu
Supplemental data for this article can be accessed here.

result of criminologists and other social scientists applying a new methodological tool to test their theories and appraise their outcomes of interest.

Another exemplar of this developmental progression lay in the race and sentencing literature (Zatz, 1987, 2000). Criminologists have long been interested in racial disparities in the criminal justice system. The critical question in this literature is whether the racial disparities can be explained by legitimate case differences that are correlated with race. As Zatz (2000) notes, late 20th century scholars improved upon simple comparisons of sentencing outcomes among race groups by adjusting for control variables to capture the influence of legitimate case differences (e.g. prior criminal history and legally relevant characteristics of the current offense) and by examining race differences after the introduction of both advisory and mandatory sentencing guidelines. This body of work generally measured weaker race effects but these later analyses themselves were criticized for ignoring interactive, indirect, and cumulative effects of race through different stages of discretionary criminal justice system processing (i.e. arrest, charging, plea and charge negotiations, and sentencing). Another key issue is Kaye's (1982) focus on the distinction between "discrimination in the application of a rule" (which is the focus of much research in racial disparities) and "discrimination in the operation of a rule" (which deals with the question of how neutrally applied rules lead to racial disparities because the rules themselves are discriminatory in nature). Taken as a whole, this literature recognizes that racial disparities are a profoundly important issue but measurement and appropriate attribution of these disparities present great challenges.

More recently, scholars are continuing to advance knowledge within this domain by attempting to apply a number of recent methodological advances (Spohn, 2015). These contemporary efforts are directed at overcoming the limitations or tenuous assumptions in prior work by employing meta-analytic techniques (Stolzenberg, D'Alessio, & Eitle, 2013), structural equation modeling (Wooldredge, Frank, Goulette, & Travis, 2015), or conditional probabilities based upon multiple dispositional outcomes (Kutateladze, Andiloro, Johnson, & Spohn, 2014; Sutton, 2013). The identification and treatment of the limitations of prior approaches within the race and sentencing literature contributed to the formation of the current state of knowledge regarding minority defendants experiencing a potential cumulative disadvantage throughout their treatment by the court and the criminal justice system (Kutateladze et al., 2014; Schlesinger, 2008; Spohn, 2015; Sutton, 2013; Wooldredge et al., 2015).

Another such area that has experienced growth in findings related to modeling advancements concerns censored dependent variables, which commonly arise in criminology. What all censored variables have in common is that their empirical distributions are fully observed for some cases and only observed in an "at least as small" or "at least as large" sense for other cases. A prominent example of a censored dependent variable that is also implicated within the race and sentencing exemplar is sentence severity. Many scholars utilize a sentence length variable measured in terms of months of custodial incarceration as a measure for sentence severity. This measure of months of custodial sentence is then used as an outcome for appraising disparities arising in judicial decision making. The number of months of incarceration is often only measured for those defendants that are convicted at trial and can vary from 0 months of incarceration (those that receive probation) to life sentences.

This measure always retains a large number of individuals who received probation, or a sentence length of 0 months. It is unlikely that all of these people have attained a true zero level of sentence severity and further unlikely that all of these individuals have an equivalent sentence severity to one another in the zero category. Indeed, this is easily confirmed by variation in the length of probation sentences imparted by judges to these defendants that do not receive custodial sentences. A more realistic model of sentence severity assumes that there is variation in sentence severity below this zero months of incarceration level, but that variation is censored. In econometric parlance, we are really interested in modeling a latent concept, y^* (i.e. sentence severity), but we only observe a measure, y, which below some threshold (i.e. sentence severity below zero months of incarceration) we cannot directly observe the true value of y^*.[1] Cases are said to be censored from below or left censored if they sit at this threshold. There is also censoring from above or right censoring when cases with a value at or above some threshold take on the value of that threshold. We can identify such an example within a sentencing context for those that receive a life sentence or the statutorily defined maximum sentence length for an offense. A judge, in some instances, may feel a defendant deserves a further harsh sentence, but that is the limit that he or she is capable of sentencing a defendant. In such circumstances, the maximum value is censoring variation that would otherwise exist if the sentence length measure could be extended upward.

As Sullivan, McGloin, and Piquero (2008) have noted, criminologists have grown familiar with the problem of censored outcomes in their datasets over a wide range of domains including, but not limited to, the working example of sentencing, offense histories, drug use, self-reported offending, and income.[2] Scholars faced with these outcomes, including sentencing scholars, employed ordinary least squares (OLS) regression, as it was the most appropriate measure in the available toolkit at the time. However, these same scholars soon realized OLS was problematic in these contexts due to the clustering of observations at zero biasing the resulting coefficient estimates (Osgood, Finken, & McMorris, 2002). As this understanding grew across the social sciences, Osgood et al. (2002) sought to clarify this point for criminologists in order to advocate for alternative modeling strategies. Models that address the clustering of

[1]Note that in the case of a censored variable we assume the factors $\boldsymbol{x}$ that affect the probability of censoring are the same factors that affect the conditional outcome, $\boldsymbol{y}$, given that it is uncensored. An alternative process is that some factors $\boldsymbol{z}$ explain the probability of observing an outcome y, but a subset of these factors $\boldsymbol{x}$ then explain the truncated outcome y. Amemiya (1985) refers to the former as a type I Tobit and the latter as a type II Tobit. In the latter case, the elements of $\boldsymbol{z}$ that are not contained in $\boldsymbol{x}$ are known as *exclusion restrictions* and this setup corresponds to Heckman's (1976) selection model. If there are terms that explain selection but not the conditional outcome (i.e. exclusion criteria), then a (type I) Tobit model provides an inconsistent representation of the process. For a discussion of how the Heckman estimator has been applied in sentencing research, see Bushway, Johnson, and Slocum (2007).

[2]It is useful to make a distinction between the related problems of censoring and truncation. A censored dataset includes all of the relevant observations. For example, we observe the data up or down to the censoring point but we are able to discern that the remaining cases are at least as small or large as that censoring point. When the data are truncated, cases that fall below or above the truncation point are not included in the sample. This problem raises other difficulties because we generally will know nothing about these cases or even how many there are. This, of course, is an extreme form of sample selection bias (where the rule that determines whether a case appears in a sample is systematically correlated with the outcome of interest) (King, Keohane, & Verba, 1994, pp. 128–132).

cases at the low end of the range are typically used in these situations including the tobit originally advocated by Osgood and colleagues as a superior alternative to OLS. They further relate that the degree of differentiation between tobit and OLS will necessarily depend upon the level of censoring encountered with the outcome data.

Criminologists have now shifted from OLS to the tobit (Tobin, 1958) as the most commonly applied method for assessing left-censored data in criminology (see Felson & Staff, 2006; Pauwels, Weerman, Bruinsma, & Bernasco, 2011; Pogarsky, 2004; Siennick, 2011). This advancement from OLS to tobit is an improvement in mitigating biases arising from OLS. Indeed, the tobit model is frequently used by criminologists due to its advancement over OLS and its availability and ease of use in standard statistical software. The tobit model, much like other statistical models, does produce consistent and asymptotically efficient estimates for effects of regressors on censored outcomes *when the assumptions of the model are met* (Tobin, 1958). However, the tobit model, like other parametric models including OLS, still makes demanding assumptions about the nature of the dependent variable and the respective error term. These assumptions, when violated, produce biases akin to those in OLS that led to departures from OLS to the tobit. These assumptions include normality, proportionality, and homoskedasticity.

The *normality assumption* asserts that there is an underlying latent, normally distributed outcome variable (y^*) that is what would have been observed if there had been no censoring. The *proportionality assumption* requires that the estimated effect of a regressor on the latent outcome is the same as the estimated effect of that regressor on the observed outcome (with proper adjustments for censoring). The *homoskedasticity assumption* requires that the variance of the error term be consistent across all of the observed values for each respective independent variable and is often referred to as the constant variance assumption, again after proper accounting for censoring. Although OLS estimates are still unbiased when the normality and homoskedasticity assumptions are violated, Tobit coefficients will be *biased* and *inconsistent* when these assumptions are violated (Amemiya, 1985; Chay & Powell, 2001).

What makes this issue so important is that these assumptions are rarely tenable in the context of criminological empirical work wherein the criminological data on which the tobit is used is often not normally distributed and heteroskedasticity is omnipresent to varying degrees. This is particularly true in the context of sentence length outcomes, which are heavily impacted by sentencing guidelines in many jurisdictions. Using the Pennsylvania Sentencing Guidelines (Kramer & Ulmer, 2009) as a working example that most sentencing scholars are familiar with, the effect of a one unit increase in the prior record score on the prescribed sentence is dependent upon where that increase takes place from. Moving from a prior record score of 0 to 1 with an offense gravity score of 8 shifts the sentence length from "9–16 months" to "12–18 months." However, the same one-unit increase that moves the prior record score from 4 to 5 with the same offense gravity score of 8 shifts the sentence length from "21–27 months" to "27–33 months." The change in sentence length accompanying a one-unit increase in prior record score at "4" is larger than the change in sentence length accompanying the same increase in prior record score at "0." This is but one example of the ubiquity of heteroskedasticity within sentencing research to ground the econometric assumption in the relevant research.

Our review of the recent literature discussed later indicates that criminologists frequently apply the tobit model to self-report and sentencing data, two instances where the assumptions of homoskedasticity, normality, and symmetry are almost always violated. This stands in stark contrast to the discussed parametric assumptions of the tobit model of normality and homoskedasticity, which seem heroic in certain contexts, and should minimally warrant caution in solely applying the tobit model to censored outcome data within this domain.

We suspect that many criminologists uncritically apply the tobit model when confronted with censored data for two reasons: (1) they assume that the tobit is robust even to strong violations of its underlying assumptions ("tobit can handle it") and thereby a simple "correction" for OLS in the situation with a lot of zeros, and (2) they are unaware of alternative model estimation strategies that may be superior ("what else can I possibly do?"). With respect to the first reason, criminologists must be given credit as the tobit is more appropriate than and an improvement over the previously employed OLS models. However, it seems they may too readily presume the robustness of the tobit model and consider the possible bias from violated assumptions as rather minor. In this article, we explore the fragility of the tobit estimator. Our analysis reveals that *even modest departures create profound bias in the tobit coefficients*. The good news is that some workable alternatives are available. With respect to the second reason, many criminologists may simply be unaware of alternative estimators, or if they are aware, may have a tendency to think that any alternative model requires difficult programming, and stick to the tobit because it is a staple of most statistical software packages.

Our dual purpose in this article is to both disabuse users of the tobit of the robustness belief and to illustrate some alternative models for censored data which are less sensitive to assumption violations that can be estimated from easily accessible statistical software packages that many or most criminologists are already quite familiar with. In fact, there are alternative specifications available for dealing with censored data that make much weaker identifying assumptions than tobit. Further, these models are relatively straightforward to estimate with popular software packages such as STATA and R (see Supporting Information Appendix[3]). The field of economics has moved away from solely applying the tobit model when faced with censored data toward presenting findings from several alternative models for censored outcomes and interpreting the differences across models in light of how those models behave (Chay & Powell, 2001; McDonald & Nguyen, 2015; Wilhelm, 2008), a practice we would encourage criminologists to follow. In this article, we pursue two objectives: (1) to demonstrate that the magnitude of the biases in the tobit model are quite profound when its underlying assumptions are not met, and (2) to provide access to alternative censored models for criminologists to employ where appropriate in place of the tobit. Just as criminologists have moved on from OLS to tobit for censored outcomes, we

[3]Instructions on the installation of and application of the SCLS and CLAD statistical models in STATA and R are presented in the Supporting Information Appendix.

hope they can again move forward from the tobit to the alternative censored models introduced here.

Models for censored dependent variables

Two semiparametric estimators which serve as alternatives to the tobit are the symmetrically censored least squares (SCLS) model and the censored least absolute deviations (CLAD) model.[4] The SCLS model is a variant of ordinary least squares (OLS) regression that is applied to the symmetrically trimmed outcome data in an effort to accommodate the censoring present in the data. The CLAD model is a derivation of quantile regression (Koenker & Bassett, 1978) wherein the error term is assumed to have a median of zero as opposed to a mean of zero.[5]

The SCLS model was first presented by Powell (1986) as a semiparametric alternative to the tobit model for modeling censored outcome data. The SCLS model proceeds first by estimating the proportion of observations that are below censored. It then truncates an equivalent number of observations from the right side of the distribution (above truncation) to mirror the existing censoring process. This results in both tails of the distribution being equivalently, or symmetrically, censored before OLS regression is applied to the remaining "middle" of the data that were not censored initially nor due to the symmetric censoring process.[6] Further, the SCLS model does generally result in comparatively less statistical power as you are necessarily dropping more observations with this approach. However, this method has been shown to be robust to both violations of normality and homoskedasticity (Chay & Powell, 2001; Powell, 1986). Where the tobit assumption of a latent normal curve necessitates a constant variance structure of the error term across all of the values of the independent variables (i.e. homoskedasticity), the trimming procedure employed by SCLS does not require the error term to be homoskedastic, but rather to simply be symmetrically distributed so that the trimming procedure does not produce a bias in the second stage of the procedure. Rather than the tobit assumption of normality, the SCLS model makes an assumption that the error distribution is symmetrically distributed about a mean of zero; this assumption is necessary for the approach of symmetrically trimming the data and then analyzing the middle portion of the data to produce asymptotically consistent, or unbiased, estimates of parameter values. Here, we note that while symmetry is a weaker assumption than normality, it is nonetheless still a strong assumption itself. Thus, SCLS makes weaker assumptions than does tobit.[7] Easily applied and prepackaged solutions are available to estimate the SCLS model in popular software

[4]Fully parametric estimators retain a structure linking the explanatory variables to the outcome measure while also explicating specific distributional assumptions regarding the error term. Semiparametric estimators retain this linkage between the explanatory variables and the outcome measure, but they only place mild requirements upon the error term as compared to more stringent parametric assumptions. The tobit model is fully parametric, that is, there are requisite distributions assumptions (normality) necessary for it estimation. In contrast, semiparametric estimators like SCLS and CLAD do not make such rigid distributional assumptions on the nature of the error term.

[5]See Sullivan et al. (2008) and Chay and Powell (2001) for more thorough discussions of the underlying assumptions of the tobit model and alternative models for handling censored outcome data.

[6]As one would expect, this statistical method cannot be applied if the censoring proportion is greater than 0.5, as all of the observations would be removed in the symmetric censoring step.

[7]To be clear, "weaker" assumptions means "less of a stretch to believe," as opposed to "not as good."

such as STATA and R (see Supporting Information Appendix for instructions on installation and use).

The CLAD model was also first presented by James Powell (1984) as an alternative analytic approach to the tobit model. The CLAD model behaves analogously to the SCLS model in its initial step in that it truncates observations from the uncensored tail of the distribution to match the proportion of observations that are censored in the alternate tail. This truncation procedure, like the trimming procedure of SCLS, is not sensitive to violations of homoskedasticity unlike the tobit's necessity for a continuous latent normality through constant variance of the error term across all values of the independent variables. What differentiates the CLAD model from the SCLS model is the application of quantile regression wherein the model attempts to minimize the sum of the absolute deviations of the points from the median of the data as opposed to applied OLS regression wherein the model attempts to minimize the squared deviations of observations from an estimated line of best fit (Powell, 1984). The CLAD model, like the SCLS model, is robust to violations of normality and homoskedasticity (Chay & Powell, 2001; Powell, 1984). The CLAD model does not require normality or even symmetry in the error distribution, and is thus even more flexible in application than the SCLS model. Thus, while SCLS makes weaker assumptions than tobit, CLAD makes weaker assumptions than both SCLS and tobit. Like the SCLS model, easily applied and prepackaged solutions are available to estimate the CLAD model in popular software such as STATA and R (see Supporting Information Appendix for instructions on installation and use).

Sullivan et al. (2008) discussed the SCLS and CLAD models in a similar context while advocating researchers report and juxtapose the results from multiple models that make different assumptions regarding the nature of the censored data in their analyses. While their recommendation appears justified, we do not yet have a good understanding of just how fragile the results of tobit estimators might be when we violate its distributional assumptions.

The use of tobit in criminological journals

For the reader to get some sense of the magnitude of the "uncritical use of the tobit" problem we examined articles published since 2000 in four top criminology journals, *Criminology*, the *Journal of Research in Crime and Delinquency*, *Justice Quarterly,* and the *Journal of Quantitative Criminology* that mention and/or apply the tobit model (See Table 1). This review reveals that the useful advice offered by Sullivan et al. (2008) was warranted, but is still largely unheeded or unheard. While many authors do articulate violations of the assumptions underlying the tobit model as a justification for using alternative model(s) (Britt, 2009; French, McCollister, Kébreau Alexandre, Chitwood, & McCoy, 2004; Loughran, Paternoster, & Weiss, 2012; McGloin, Sullivan, Piquero, & Bacon, 2008; Steffensmeier & Demuth, 2001; Sweeten, 2012; Vazsonyi, Wittekind, Belliston, & Van Loh, 2004), the findings from this review of the literature suggest that tobit is still the "go-to" model specification when criminologists encounter left-hand censored data. Of the 27 pieces within this review that applied the tobit model directly, most neither mentioned nor tested for violations of the tobit model's

Table 1. Primary applications of the Tobit model in *Criminology* and the *Journal of Quantitative Criminology* since 2000.

Author(s)	Year	Journal	Mentioned Tobit assumptions	Tested at least one assumption	Compared results with an alternative model for censored outcome data	Only compared results with OLS
Broidy	2001	*Criminology*	No	No	No	Yes
Nagin and Pogarsky	2001	*Criminology*	Yes	No	No	No
Wright, Caspi, Moffitt, and Silva	2001	*Criminology*	No	No	No	Yes
Jarjoura, Triplett, and Brinker	2002	*Journal of Quantitative Criminology*	No	No	No	No
Lu and Drass	2002	*Justice Quarterly*	No	No	No	No
Piquero and Pogarsky	2002	*Journal of Research in Crime and Delinquency*	No	No	No	No
Cohen, Gorr, and Singh	2003	*Criminology*	No	No	No	Yes
Pogarsky and Piquero	2003	*Journal of Research in Crime and Delinquency*	No	No	No	No
Pogarsky	2004	*Criminology*	No	No	No	No
Kurlychek and Johnson	2004	*Criminology*	No	No	No	No
Carmichael and Piquero	2004	*Journal of Quantitative Criminology*	No	No	No	No
Messner et al.	2005	*Criminology*	Yes	No	No	No
Felson and Staff	2006	*Criminology*	No	No	No	No
Sullivan et al.	2006	*Criminology*	Yes	Yes	Yes	No
Ousey and Wilcox	2007	*Criminology*	Yes	Yes	Yes	No
Piehl and Bushway	2007	*Journal of Quantitative Criminology*	Yes	Yes	No	No
Tittle, Broidy, and Gertz	2008	*Justice Quarterly*	No	No	No	No
Wright and Younts	2009	*Journal of Research in Crime and Delinquency*	No	No	No	No
Kurlychek and Johnson	2010	*Criminology*	Yes	Yes	No	No
Wiesner et al.	2010	*Journal of Research in Crime and Delinquency*	Yes	Yes	Yes	No
McGloin and Piquero	2010	*Journal of Research in Crime and Delinquency*	No	No	No	No
Roettger and Swisher	2011	*Criminology*	Yes	No	No	No
Siennick	2011	*Criminology*	No	No	No	No
Johnson and Kurlychek	2012	*Criminology*	No	No	No	No
Kamerdze, Loughran, Paternoster, and Sohoni	2014	*Journal of Research in Crime and Delinquency*	No	No	No	No
McGloin and Rowan	2015	*Criminology*	Yes	Yes	No	No
Weerman, Bernasco, Bruinsma, and Pauwels	2016	*Justice Quarterly*	No	No	No	No

parametric assumptions. A few authors discussed all of the parametric assumptions of the tobit model though did not provide any accompanying tests for violations of these assumptions (Messner, Deane, Anselin, & Pearson-Nelson, 2005; Nagin & Pogarsky, 2001; Roettger & Swisher, 2011) with fewer further discussing at least one of the assumptions with an accompanying test for a violation of that assumption (Kurlychek & Johnson, 2010; McGloin & Rowan, 2015; Piehl & Bushway, 2007). Only three articles were found in this review that applied the tobit model, discussed the parametric assumptions of the tobit model, tested for violations of at least one of these assumptions, and compared the results with an alternative model designed for censored outcome data (Ousey & Wilcox, 2007; Sullivan, McGloin, Pratt, & Piquero, 2006; Wiesner, Kim, & Capaldi, 2010).[8]

Alternative models, such as SCLS and CLAD, will be superfluous in a context in which the assumptions of the tobit model have been met, as the tobit results will be unbiased and *more efficient* than the respective SCLS and CLAD results would be in such a context (Chay and Powell, 2001).[9] However, even in such a context, an OLS model is not the appropriate benchmark by which to compare results pertaining to censored outcome data as it has well known biases (Osgood et al., 2002; Sullivan et al., 2008). Sullivan et al. (2008) further highlight this point in presenting CLAD as the appropriate benchmark for statistical model comparisons with censored outcome data due to CLAD being the most general of the discussed statistical models for handling censored data with the weakest assumptions.

The current study

It is clear that the tobit is employed by criminologists toward handling censored outcome data and that the tobit retains strong parametric assumptions. To date, there has not been a clear demonstration of the fragility of the tobit in the face of assumption violations in criminology.[10] Sullivan et al. (2008) provide an empirical example with sentencing data to demonstrate differential behavior across several different statistical models with censored outcome data. While quite useful in displaying the advantages of employing multiple statistical methods simultaneously, this approach *precluded discussion of exactly how much bias* was invoked by misapplying these statistical models as the "true" population coefficient values were unknown in the employed sentencing dataset. Sullivan et al. (2008) could and did discuss the differences across the models, but they could not make definitive statements about which

[8]Several articles that mention the tobit model within this review are not included here as they applied a hierarchical linear modeling framework that precludes discussion in the present context or conducted a different analysis that likewise precluded their discussion here.

[9]In other words, in cases where the assumptions are met the tobit is the preferable estimator, as it will be the most efficient of available statistical models. Indeed, Osgood et al. (2002) present the tobit model as a viable and superior alternative to OLS within the framework of item response theory. While the discussion in Osgood et al. (2002) was limited to the tobit model and OLS, the discussion of the tobit model retaining the most efficient estimates holds even when comparing to the presently discussed semiparametric alternatives so long as the assumptions of the tobit model are tenable.

[10]McDonald and Nguyen's (2015) paper in *Communications in Statistics – Simulation and Computation* presents similar simulation analyses with censored outcome data in constructing further alternative models for censored outcome data. However, their paper is not targeted to a criminological audience, and remains impenetrable under layers of dense econometric formulae.

model was "best" or "least biased" due to this lack of information about the true population values.

An alternative approach to addressing this problem that allows for an explicit discussion and assessment of the bias(es) inherent to specific statistical models when faced with violated parametric assumptions is through statistical simulation (Monte Carlo) analyses. This approach is explicitly called for by Sullivan et al. (2008, p. 417) as the next step toward identifying the most appropriate model(s) for analyzing censored outcome data and identifying the biases accompanying breakdowns of the parametric assumptions of the alternative models:

> Although we used a single set of estimates with empirical data, more may be learned from future studies. Perhaps by relying on simulation models that allow for various data distributions and other conditions, researchers may be able to better understand the properties of these estimators.

Simulation studies create data with known population coefficient values (i.e. the "true" value) that enable not only a comparison across statistical models, but also a discussion of the *exact magnitude of the bias* that arises as a result of violation(s) of the parametric assumptions underlying statistical models. Estimates from the employed statistical models can be compared directly to the known population parameters toward calculating a percentage of bias that is produced with each respective statistical model. This percentage of bias then allows for an analysis that can rank the employed statistical models in the relative efficacy in handling data of varying formats.

It is through this avenue of simulations that this article seeks to build upon the work presented by Sullivan et al. (2008) with simulation analyses in which the population parameters are known in order to illustrate the bias that arises when the tobit model is misapplied. These simulations will produce data that systematically and to varying degrees violates the parametric assumptions of the tobit model pertaining to homoskedasticity and normality. We then apply several estimators including tobit to quantify the bias that can result from inappropriate use of the tobit model in the face of assumption violations.[11] These simulation analyses will depict how the reported biases in misapplying the tobit model can further be avoided through thoughtful application of alternative models (SCLS, CLAD) that can be easily estimated with popular software packages such as STATA and R (see Supporting Information Appendix). To further ground these simulation analyses, we conclude with an empirical demonstration comparing the various statistical models under study with sentencing data from the State Court Processing Statistics (SCPS) to assess for racial disparities in sentence length, a much-appraised research question in sentencing scholarship. These results are designed to continue the discussion regarding the appropriate means of analyzing censored outcome data initiated by Osgood et al. (2002) and continued by Sullivan et al. (2008). We hope that these results will provide an impetus for researchers to exhibit greater caution in applying the tobit model in future analyses and consider the utility afforded by applying multiple methods that make less stringent assumptions regarding the nature of the underlying data.

[11]This article does not treat the proportionality assumption, as Smith and Brame (2003) have already demonstrated the bias that arises in misapplying the tobit model in such a situation.

Simulations

Overview

The data simulations for this work proceed in a systematic manner. Each stage in the subsequent analyses will present results from OLS regression, tobit, SCLS, and CLAD. It is clear that OLS will be biased in all of the following applications involving censored data (Osgood et al., 2002). Nevertheless, OLS is included in these analyses to quantify this bias by demonstrating the problems with using it as a benchmark for comparisons with the tobit model. First, baseline simulations are conducted under conditions in which all of the tobit assumptions are met, to demonstrate that OLS is biased in the presence of censored outcome data and that each of the three models (Tobit, SCLS, CLAD) designed for censored outcome data is unbiased in this case. Second, heteroskedasticity (two forms) is introduced to the baseline model to demonstrate the bias that arises with application of the tobit model, which can be avoided using either SCLS or CLAD. Third, a disturbance term distributed student-*t* is generated at varying degrees of freedom to demonstrate a relative continuum of simulation data ranging from normality to non-normality (while preserving symmetry); this third stage of simulation analyses will present the bias that arises in application of the tobit model to non-normal data, which can likewise be avoided using either SCLS or CLAD. Last, non-symmetric errors are introduced through the incorporation of a disturbance term distributed gamma toward violating the symmetry assumption of SCLS and solidifying CLAD's status as an appropriate benchmark model for cross-model comparative purposes with censored outcome data.

Each stage in the simulation procedures employed 500 simulated datasets with 1000 observations per dataset and varying amounts of censoring.[12] Systematically varying the amount of censoring yields multiple sets of results for each stage of the simulation analyses toward addressing any potential sensitivity in the results to the degree of censoring employed. The percent bias statistics reported represent the average percent bias resulting from application of each respective model across the 500 simulated datasets for each stage in the analysis.

Simulation set 1: Baseline with all tobit assumptions met

The first simulation includes data generated with all of the assumptions of the tobit model having been met; the error term is both normally distributed and homoskedastic. The equation used to generate this simulated data at baseline can be expressed as follows:

$$y^* = \beta_0 + 0.5 * x - 0.5 * z + u \quad (1)$$

where, y^* is the latent outcome of interest that is subsequently censored, β_0 is the given constant to produce the reported level of censoring,[13] x and z are mutually independent and normally distributed variables, and u is a disturbance term

[12]All of the results from these simulation analyses were produced with STATA version 13.0. The code and do-files for these simulation analyses are available upon request.

[13]Each set of data is initially centered about zero and a censoring point is applied at a point less than zero. This censoring point is then introduced as a constant to the model to displace the censoring point toward zero, which is more common within the discipline of criminology.

distributed standard normal ($\mu = 0$, $\sigma = 1$). The manifest variable, y that is analyzed with the previously discussed models is constructed to be censored at zero while retaining values equal to y^* for those observations that are greater than zero as shown in Equation (2). These initial analyses are conducted with censoring proportions of 0.1, 0.2, and 0.3.

$$y = \begin{cases} 0, y^* \le 0 \\ y^*, y^* > 0 \end{cases} \tag{2}$$

Figure 1 depicts the percent bias in the estimates for the population coefficients on x (0.5) and z (-0.5), and the constant (varies by censoring proportion)[14] across the four models for each of the given censoring proportions.[15] The tobit, SCLS, and CLAD models all retain percent biases very close to zero, regardless of the degree of censoring employed. This suggests, in line with *a priori* expectations, that the tobit, SCLS, and CLAD models are all consistent when the homoskedasticity and normality assumptions of the tobit model are met. Further, the results from the tobit model are the most efficient of these three models, as they retain the smallest standard errors for each respective estimate compared with SCLS and CLAD. CLAD retains the largest standard errors of these three models with SCLS coming in between the tobit and CLAD in terms of efficiency. This finding lends credence to applying tobit in cases when the assumptions of the tobit model are met, as it these simulation results confirm that it is, indeed, the most efficient estimator in such a context.

In contrast, the OLS estimates are substantially biased with the magnitude of this bias increasing markedly as the censoring proportion increases. The magnitude of the percent bias for the coefficients on both x and z, respectively, nearly approximate the percentage of observations that are censored. This follows the finding of Greene (2005) that OLS coefficient estimates applied to censored outcome data will be equivalent to the true parameter multiplied by the proportion of cases that are not censored. When 10% of the observations are censored, the OLS output is biased for coefficients on x and z by –9.64% and 10.13%, respectively. Further, when 30% of the observations are censored, the OLS output is biased for the coefficients on x and z by –29.74% and 30.24%, respectively. This substantial bias is in line with the caution expressed within the literature (Osgood et al., 2002; Sullivan et al., 2008) against using OLS regression in the presence of censored outcome data. Despite the clear bias in applying OLS regression to censored outcome data when the assumptions of the tobit model are met, the results from applying OLS are retained and displayed in the subsequent stages of this analysis to further clarify the issues that might arise in attempting to interpret results from OLS as a point of comparison with tobit results.

Simulation set 2: Heteroskedasticity introduced

The next stage of the analysis retains the same structure as the first stage, retaining normality, but we now introduce heteroskedasticity into the model by means of a

[14]As highlighted in Footnote 12, the censoring point to produce the given level of censoring becomes a constant within the model once the data is moved to provide a "new" censoring point of zero.

[15]The tables of results upon which this figure, and all subsequent figures, were created are available upon request.

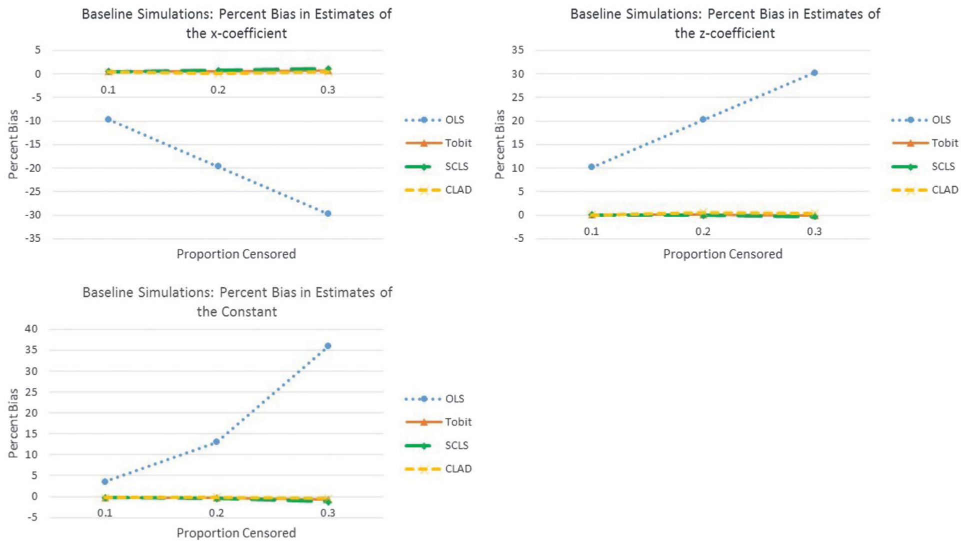

Figure 1. Baseline simulations results.

biskedastic error term. The simulated data for this stage follows Equation (1) in a similar manner to the first stage. The difference between this stage and the first stage lay in the conditional variance of the error term. While in the previous simulation, the error terms were homoskedastic, or more specifically, $\mathrm{var}(u \mid x) = 1$ (i.e. the variance was constant across all values of x), in this set of simulations, the conditional variance of the error term depends on the value of the independent variable x. More specifically, we generate data such that $\mathrm{var}(u \mid x < 0) \neq \mathrm{var}(u \mid x \geq 0)$ Two levels of heteroskedasticity are modeled through this approach—severe and moderate. In each case, $\mathrm{var}(u \mid x \geq 0)$ is distributed standard normal ($\mu = 0$, $\sigma = 1$), while $\mathrm{var}(u \mid x < 0)$ is also distributed normally with mean 0, but now with a standard deviation of $\sigma = 2$ in the moderate heteroskedasticity simulation and a standard deviation of $\sigma = 3$ in the severe heteroskedasticity simulation. All of the other variables and parameters in Equation (3) behave in the same fashion as they did previously in Equation (1). The heteroskedasticity is only incorporated into the model according to the value of x so that estimates of the coefficient on z should remain unaffected in the resultant models due to x and z remaining independent of one another in the simulations.

The left column of Figure 2 presents the percent bias in the coefficients on x, z and, the constant for the moderate heteroskedasticity simulations. The right column of Figure 2 presents the corresponding percent biases for the severe heteroskedasticity simulations. The columns are juxtaposed to one another according to the estimated parameter to demonstrate the similar patterns in the biases in estimates of the same parameter across the statistical models evoked due to the presence of the biskedastic error term. Note that the proportion censored varies between the moderate and severe heteroskedasticity simulations; this was done to insure that the true value of the constant in each simulation would be equivalent across the two sets of simulation results. The “true” value for the constant for the lowest level of censoring in each respective column is equivalent and likewise so for the highest and middle levels of censoring. The only other substantive difference between the two sets of figures is

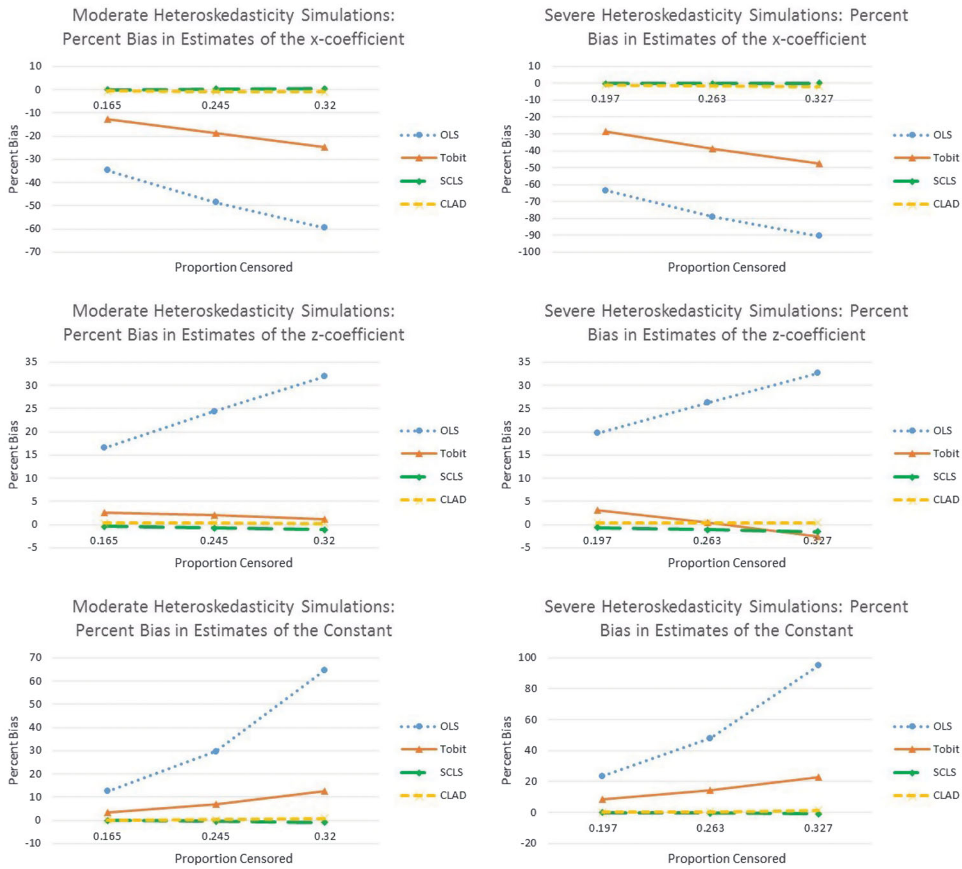

Figure 2. Heteroskedasticity simulations results.

with regard to the scale of the percent bias axis due to the greater magnitude of the biases produced in the severe heteroskedasticity simulations as compared to the moderate heteroskedasticity simulations.

The general pattern of the OLS results in these two sets of figures parallel those presented in the baseline analysis, but with the biases here reaching much higher maximum levels of –90.44% and 32.64%, respectively, for coefficients on x and z, and 94.86% for the constant. These biases are very substantial and point to a compromising misspecification related to their application in this context. The tobit results, while less biased than the OLS results, still yield biased parameter estimates as a result of the heteroskedasticity. The bias on the coefficient estimate for z is relatively small using tobit with the magnitude of the bias being lower than 5% for all cases presented here and each of these estimates falling within one standard deviation from the respective true value. This is expected given that the z variable is explicitly generated so that it does not violate the homoskedasticity assumption of the tobit model. However, the coefficient estimate for x yields a considerable amount of bias, ranging from –12.60% at the lowest censoring level in the moderate heteroskedasticity simulation to –47.55% at the highest censoring level in the severe heteroskedasticity simulation. This is a large under-estimation of the true effect of the independent variable on the given outcome wherein 95% confidence intervals for the coefficient estimates on x do not include the true value of 0.5 in each of the presented cases save for those at the lowest level of censoring with the moderate degree of heteroskedasticity. The bias in estimation, as opposed to the efficiency of the estimates, leads to incorrect decisions for hypothesis testing based upon these findings.

The SCLS and CLAD results, in direct contrast to OLS and tobit, produce very little bias under all specifications of censoring and heteroskedasticity we generated. This is in line with *a priori* expectations, as both of these models are sensitive to neither censoring[16] nor heteroskedasticity in their respective ability to yield unbiased estimates. To quantify this directly in addition to the visual evidence presented to this effect, the maximum bias in the coefficient on x produced from the SCLS results was 0.47% and the maximum bias in the same coefficient produced from the CLAD results was 2.01%. The bias in the comparably "worst" estimate from the tobit model of –47.55% is over 100 times more biased than that of the SCLS model and over 23 times more biased than that of the CLAD model. Further, both the SCLS and CLAD models retain confidence intervals that that include the true population values for each of the parameters estimated in each set of simulation analyses. The chief distinction here between the two semiparametric estimators, both of which are relatively unbiased and yield estimates within one standard deviation of the respective true values, has to do with the magnitude of the standard errors. When the assumption of symmetry is met as it is in the current simulation, then SCLS is more efficient than CLAD.

Simulation set 3: Non-normality introduced

The third stage of these analyses pertained to investigating the impact of departures from normality. Unlike the case in previous stages, a univariate model is employed with a disturbance term based on the student's t distribution as follows:

$$y^* = \beta_0 + 0.5 * x + u \tag{3}$$

where u is distributed as a student's t with degrees of freedom varying from 30 to 2 with subsequent rounds of simulations to generate movement from what is approximately normal according to the Central Limit Theorem (with degrees of freedom equal to 30) toward what is not normally distributed, but still symmetric, with a degree of freedom of 2.[17] Each of the other factors in Equation (3) are specified and modeled in accordance with the methods previously employed, but the model is homoskedastic in line with the results of the baseline simulation toward independently assessing the biases resulting from violations of normality in application of the tobit model. Concurrent with the prior stages, this stage of simulations is conducted at multiple levels of censoring.

[16]While these models are not sensitive to censoring, they do necessitate a censoring proportion less than 0.5 in order for these models to be able to return a statistical estimate.

[17]The student t distribution was chosen for its departure from normality and its symmetrical nature that is a requisite assumption of the competing SCLS model.

The left column of Figure 3 presents the percent bias in the estimates of the coefficient on x and the constant at a censoring proportion of approximately 20%.[18] The right column of Figure 3 presents the respective percent biases in the estimates of the coefficient on x and the constant at a censoring proportion of approximately 33%. The OLS results from this set of simulations echo the prior finding that OLS produces substantial biases in estimation when applied to censored outcome data. The SCLS and CLAD models remain unbiased in application in the face of non-normal errors, as all of their respective points of observation continue to hover about the axis at a percent bias of zero. The tobit results, however, show an increasing bias resulting from the disturbance term's departure from normality. As the degree of freedom for the student *t*-distributed disturbance term decreases (as the magnitude of the non-normality of the distribution increases), the bias in the tobit estimates increases markedly. This is particularly apparent in assessing the behavior of the estimates as the degree of freedom shifts from 5 toward 2. As with the previous stages of these simulations, the pattern of this effect and the respective biases are not sensitive to the degree of censoring in the model as the two columns produce substantively similar results with the primary difference between the outputs being a difference of relative magnitude. Again, the two semiparametric estimators can be differentiated in terms of their relative efficiency—SCLS is more efficient than CLAD given the conditions in which symmetry is met.

Simulation set 4: Nonsymmetry introduced

The last stage of these analyses investigates departures from symmetry through the incorporation of a gamma distributed error term into the univariate model used in stage 3 and represented below in Equation (5):

$$w = f - mean(f) \tag{4}$$
$$y^* = \beta_0 + 0.5 * x + w \tag{5}$$

where f within Equation (4) is a gamma-distributed variable with a shape parameter, $k = 1$, and a scale parameter, $\theta = 0.5$. As a gamma distributed variable cannot take on negative values and necessarily has a mean greater than zero, f is displaced downward by its mean to construct the disturbance term, w that is incorporated into the estimated model in Equation (5). This leads to w being nonsymmetric while retaining a mean that is equal to zero and a median less than zero in this context. Figure 4 presents graphical depictions of the probability density functions of both f and w.

This transformation, while conducted in the interest of tractability with a nod toward SCLS, OLS, and tobit's operation with the error term's mean equaling zero has ramifications with regard to estimation of β_0. The displacement of the gamma distributed variable by its mean introduces a scalar to the model as represented below in Equation (6) which combines Equations (4) and (5):

$$y^* = \left[\beta_0 - \bar{f}\right] + 0.5 * x + f \tag{6}$$

where $[\beta_0 - \bar{f}]$ becomes the new constant in the model. This results in the parameter β_0 being unidentifiable in the current model, as it cannot be parsed from the

[18]As with the prior analyses, the censoring proportion was allowed to vary slightly with different specifications of the disturbance term in order to maintain a consistent known parameter value for the constant in these simulations.

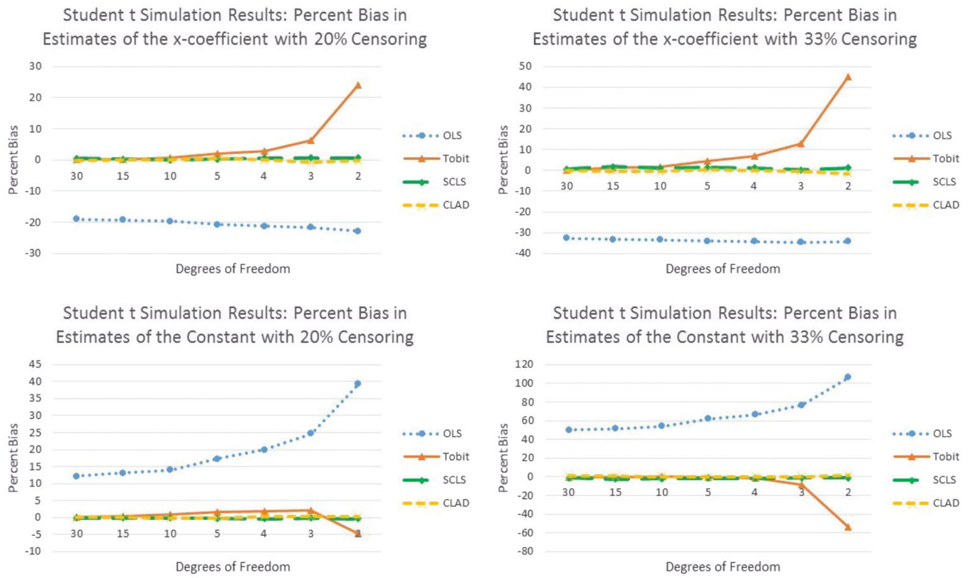

Figure 3. Student *t* simulations results.

incorporated displacement term.[19] This is not an issue that would arise in practice with natural nonsymmetric error terms, as the displacement term would not be "displaced" by a scalar. However, the estimates of the coefficient on x are unaffected by this procedure. As such, we only present and discuss the percent biases in estimation of the coefficient on x in this stage of the simulations.

Figure 5 presents the percent bias in the estimates of the coefficient on x with the gamma distributed disturbance term. The CLAD estimates hover directly around zero percent bias, and demonstrate the CLAD model's robustness to departures from symmetry as compared to the alternative models exhibiting biases ranging from 5% to 10%. While the tobit estimates are less biased than the SCLS estimates in this figure, this is within a framework in which all of the other assumptions of the tobit model have been met. This should not be taken as evidence for a relative superiority of the tobit as compared to SCLS; both models retain violated assumptions resulting in biased estimates that can be corrected for with the CLAD model. This finding couples with our and Sullivan et al. (2008) recommendation to begin an analysis with the CLAD model and then move to models that make "stronger" assumptions.

Findings on efficiency

The previous sets of results have focused on quantifying the bias resulting from systematic violations of assumptions of the tobit model. However, these same simulations can be leveraged toward discussing matters of efficiency in addition to bias.

[19]While we cannot identify specifically in this model due to its combination with the displacement term incorporated through the residual, our primary interest lay with the coefficient(s) on the independent variables and assessing SCLS and CLAD with regard to their respective ability to return consistent estimates in the presence of nonsymmetric errors.

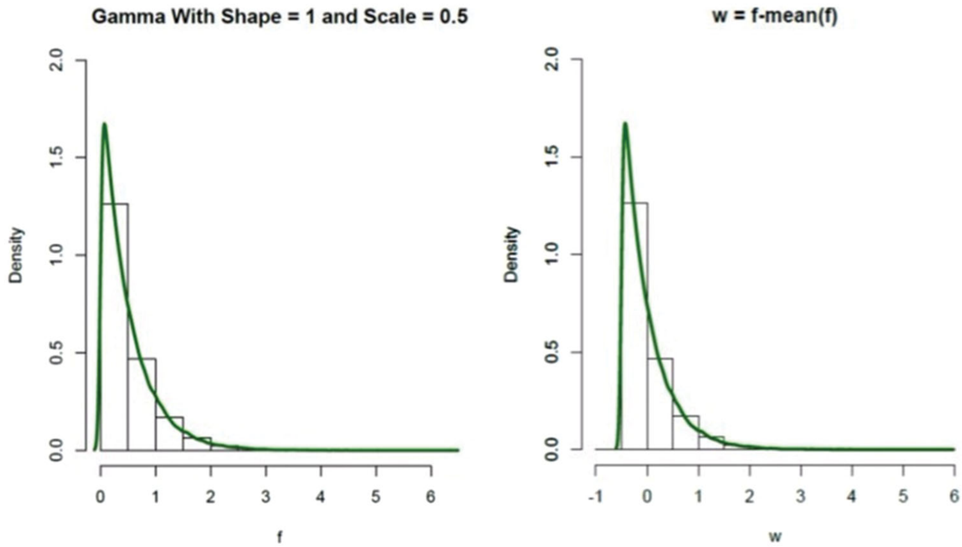

Figure 4. Gamma distributed random variables *f* and we: Probability density functions.

Figure 6 presents 95% confidence intervals for the *x*-coefficient estimates from tobit, SCLS, and CLAD from two sets of the produced simulation results: baseline with 10% censoring and severe heteroskedasticity with 26.3% censoring. The standard deviation for the tobit estimates, and the respective confidence interval, is smaller than those of SCLS and CLAD for all of the simulations produced in this work. In the case of the baseline simulations in Figure 6, this increased efficiency is beneficial given the lack of bias in the estimate. However, in the case of the severe heteroskedasticity simulations in Figure 6, this increased efficiency is detrimental in producing a tighter band around a biased estimate that results in a 95% confidence interval for the *x*-coefficient failing to cross the true value of 0.5.

Empirical demonstration

Overview

As noted in the introduction, sentencing scholars are often concerned with and assessing racial disparities in sentencing decisions such as whether a person will be incarcerated and the length of time imprisoned. We turn, now, to an empirical demonstration evaluating racial disparities using sentencing data from the State Court Processing Statistics Series (SCPS) to provide an application of the simulation analyses presented earlier. The SCPS data are case-level data on felony cases where prosecutors filed charges in 71 large urban counties across 21 states in the United States. This empirical demonstration only uses data pertaining to felony cases that resulted in convictions from the 2006 component of the SCPS data. This leads to a sample of 9636 convicted cases of the 16,211 felony charged cases in the 2006 SCPS data.

The primary independent variables for this demonstration are race (white, black, Hispanic, other race), a count of prior felony convictions, and interaction terms between race and prior felony convictions to account for differential effects of prior

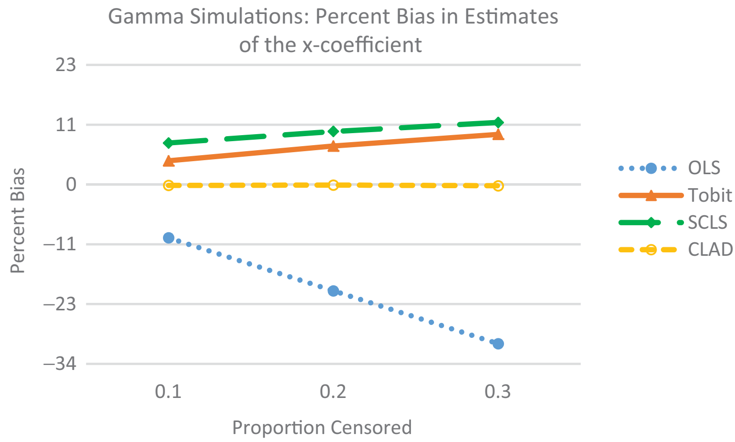

Figure 5. Gamma simulations results.

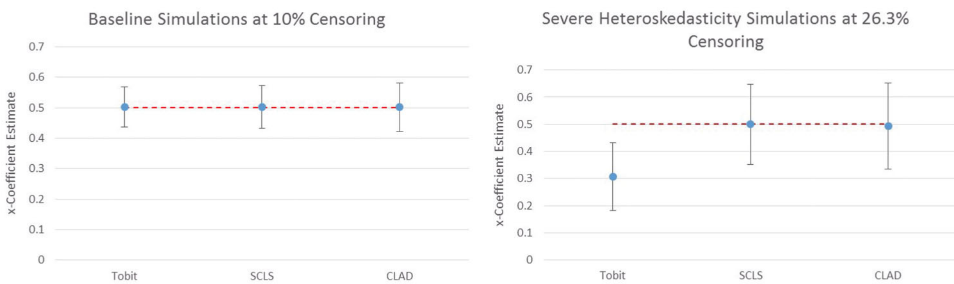

Figure 6. 95% Confidence intervals for the *x*-coefficient estimates across Tobit, SCLS, and CLAD.

convictions on sentence length across racial groups. Prior felony convictions are missing for 82 cases leading to the analytic sample of 9554 cases for the empirical demonstration. The outcome of interest is sentence length, in months, to either jail or prison for these convicted cases. This sample is 30% non-Hispanic white, 43% non-Hispanic black, 23% Hispanic, and 4% other race. The defendants had been convicted of 1.56 felony offenses, on average, prior to the current case and 64.08% of them ended up receiving a sentence of incarceration. This provides a censoring level of roughly 36% (corresponding to the percent of the sample who were not incarcerated). A histogram of the number of months incarcerated top-coded to 360 months (30 years) is provided below in Figure 7.[20]

We can see the clear censoring and uptick at 0 in the histogram, but we also see a heavy positive skew to the data that reveals a dramatic departure from normality in the distribution of the dependent variable. Taken together with the previously presented simulation results, this non-normality leads us to expect biased results from the tobit model. The goal for this empirical demonstration is to see whether we obtain disparate results from tobit, SCLS, and CLAD when using actual sentencing data.

[20]Top-coding of the outcome measure had to be employed in order to account for 183 life sentences. 0.45% of the defendants received a sentence longer than 360 months of incarceration, and 0.79% of the defendants received a sentence longer than 240 months of incarceration.

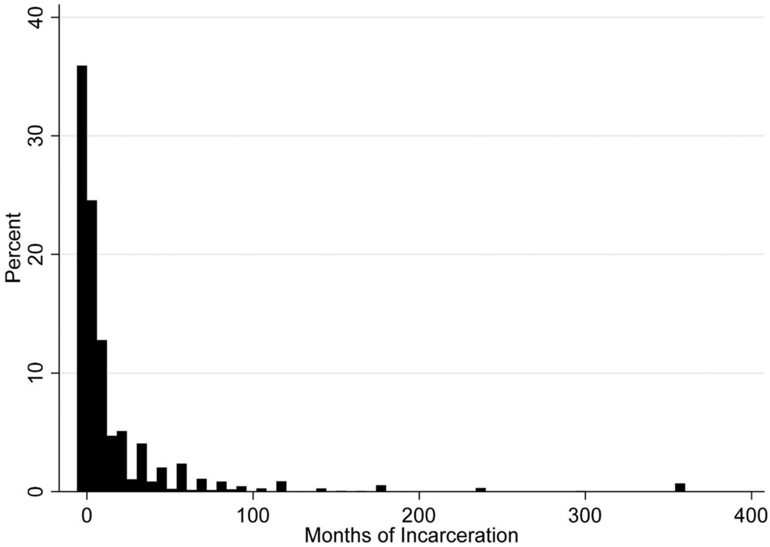

Figure 7. Histogram of months of incarceration top-coded at 360 months ($n = 9554$).

Table 2. Results for empirical demonstration ($n = 9554$).

	Models predicting months of incarceration with different coding of incarceration length							
Incarceration coding	Incarceration months top coded at 360 months				Incarceration months top coded at 240 months			
Model	OLS	Tobit	SCLS	CLAD	OLS	Tobit	SCLS	CLAD
Black	4.97**	7.83**	1.36*	1.43**	4.32**	6.71**	1.36*	1.43**
	(1.24)	(1.82)	(0.54)	(0.48)	(1.05)	(1.55)	(0.54)	(0.54)
Hispanic	1.86	7.00**	2.62**	2.87**	1.83	6.16**	2.62**	2.87**
	(1.41)	(2.06)	(0.54)	(0.54)	(1.20)	(1.75)	(0.54)	(0.58)
Other race	−1.58	−5.73	−0.22	3.37	−1.93	−5.55	−0.22	3.37
	(2.65)	(4.00)	(1.77)	(2.18)	(2.25)	(3.41)	(1.77)	(2.05)
Prior felony convictions	3.72**	6.32**	2.41**	2.37**	3.49**	5.68**	2.41**	2.37**
	(0.35)	(0.48)	(0.25)	(0.30)	(0.29)	(0.41)	(0.25)	(0.33)
Black × prior felony convictions	−2.23**	−2.62**	−1.09**	−0.93**	−2.06**	−2.40**	−1.09**	−0.93**
	(0.43)	(0.60)	(0.27)	(0.33)	(0.37)	(0.51)	(0.27)	(0.33)
Hispanic × prior felony convictions	−2.51**	−3.39**	−1.23**	−0.87*	−2.23**	−2.97**	−1.23**	−0.87*
	(0.56)	(0.78)	(0.32)	(0.39)	(0.47)	(0.66)	(0.32)	(0.41)
Other × prior felony convictions	−0.99	0.12	0.80	0.13	−0.58	0.40	0.80	0.13
	(1.11)	(1.56)	(0.93)	(0.87)	(0.94)	(1.33)	(0.93)	(0.86)
Constant	12.18**	−10.19**	1.05*	0.13	11.71**	−7.07**	1.05*	0.13
	(0.94)	(1.41)	(0.44)	(0.33)	(0.80)	(1.20)	(0.44)	(0.42)

Notes. Each column provides the results from the specified model (OLS, tobit, SCLS, or CLAD) of sentence length in months of incarceration. The first four columns provide results when top-coding months of incarceration to 360 months while the latter four columns provide the same models with the outcome top-coded to 240 months. Each model utilizes the same sample of 9554 respondents and the same independent variables listed in the leftmost column. The reference category for the race coefficients is non-Hispanic white defendants. The only difference across columns is the employed modeling technique.
*$p < .05$; **$p < .01$ (two-tailed).

Results

All four models used in the simulation analyses were also used here to support a comparison of the results. Unlike the simulation analyses, however, the correct parameter estimates for race and prior felony convictions are not available. Those true values are necessarily unknown within an actual dataset. However, given that CLAD provided the least amount of bias within the simulation analyses and Figure 7 revealed a violation of the tobit's normality assumption, we treated the CLAD results as a baseline for comparison for these results. Table 2 displays the results for OLS, tobit, SCLS, and CLAD estimates for the effects of race and prior felony convictions on sentence length with non-Hispanic white defendants serving as the reference category.

The tobit, SCLS, and CLAD estimates all reveal racial disparities in sentence length where non-Hispanic black defendants and Hispanic defendants receive longer carceral sentences than white defendants while controlling for prior felony convictions. However, the magnitude of the observed disparity is heavily dependent upon the statistical model employed. If we only used the tobit model for this analysis, we would conclude that black defendants were sentenced to 7.8 or 6.7 additional months of incarceration over white defendants depending on our top-coding decision. If we had, instead, employed the SCLS or CLAD models, we would find a far smaller disparity of either 1.36 or 1.43 months of incarceration between black and white defendants. This is a substantial difference in the coefficient estimates across these models with the tobit estimate of the black-white disparity being over 5 times greater than the respective CLAD estimate. Based upon the skewed distribution for the outcome measure displayed in Figure 7, we know that this difference in coefficients between tobit and CLAD is being driven at least in part by bias in the tobit estimate arising from violations of the normality assumption for the tobit model. Figure 8 further exemplifies this bias by displaying 95% confidence intervals for the main race differences between blacks and Hispanics as compared with white defendants.

In the case of the black-white difference, the 95% confidence interval around the tobit coefficient does not cross the respective 95% confidence intervals for either the SCLS or CLAD estimates. While the correct value is not knowable in the context of the SCPS data, on the basis of the skewed distribution and nonoverlapping confidence intervals, we can state that the tobit estimate is incorrect and biased due to violated assumptions. The SCLS and CLAD estimates are more credible in this empirical demonstration with SCPS data.

Discussion

The tobit model continues to be used in criminological research, but only occasionally with explicit assessment of whether it is appropriate for the respective context. This is true in spite of the fact that the assumptions of the tobit model are often violated by criminological data most likely to be modeled by the tobit. Many authors leverage the presence of a censored outcome as sufficient justification for estimating a tobit model as opposed to an OLS regression model without mention of alternative models or specifications. The simulation results from this analysis do support these authors in that tobit estimates are generally less biased in application than respective results from OLS regression in the face of censored outcome data. However, the simulation

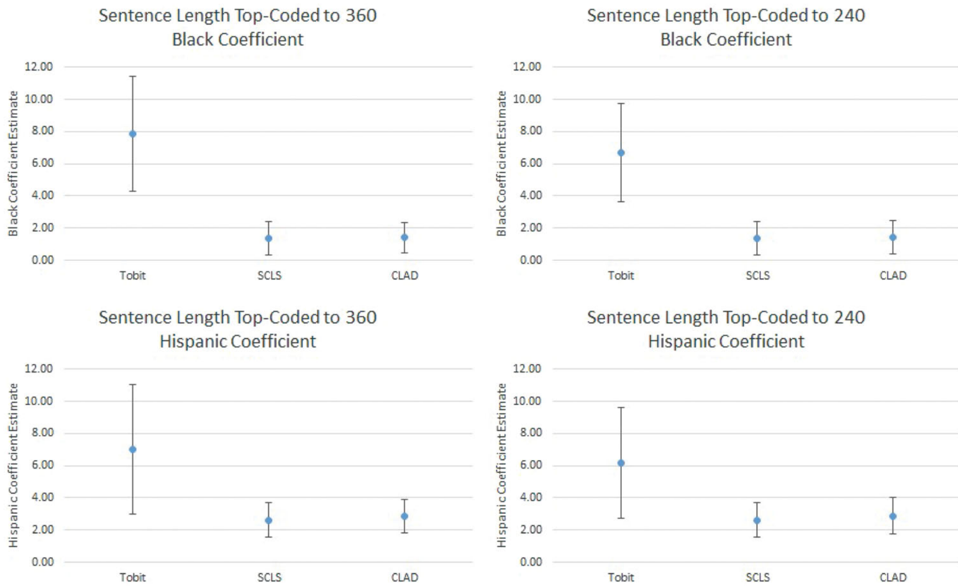

Figure 8. 95% Confidence intervals for main race coefficients from models in Table 2.

results presented here further demonstrate that tobit's relative superiority to OLS does not make it an automatically ideal model for any application involving censored outcome data, as it is still subject to substantial biases when its core assumptions are violated. Both SCLS and CLAD have been shown to be unbiased in application with censored outcome data regardless of whether the homoskedasiticty and normality assumptions of the tobit model have been met. CLAD has been shown to be the most flexible of the models as it is also unbiased in the presence of nonsymmetrical errors.

Given these findings that both the tobit and OLS regression models produce substantially biased estimates that can be corrected with easily applied alternative models in popular statistical software packages (see Supporting Information Appendix), this article strongly advocates the position adopted by Bushway, Robert, and Paternoster (1999) and Sullivan et al. (2008) toward utilizing multiple methods for analyzing data, where researchers should initially estimate models with weaker assumptions then move to models with more restrictive assumptions. Researchers faced with censored outcome data, including sentencing scholars, should consider utilizing a CLAD model as their initial model of estimation due to its more lenient assumptions regarding the nature of the data than those of tobit.[21] This position is strengthened by the results of the simulations with nonsymmetric errors depicting biases in SCLS and tobit that are

[21]Another option not explored in this article but germane nonetheless is the idea of developing bounds on key parameter estimates that respect the uncertainty created by endpoint censoring (Manski, 1995). It is well known that quantiles of a probability distribution can be set or partially identified even though point estimates such as the mean of the distribution cannot be identified without resorting to stronger assumptions. A critical feature of the bounding approach is that high-dimensional problems (i.e. many control variables) are harder to solve and we lose the ability to obtain a single point estimate for a parameter of interest. Conversely, the bounding approaches make much weaker assumptions than conventional statistical models which can enhance the credibility of the estimates that are ultimately obtained.

corrected with CLAD.[22] Much of the censored data criminologists are faced with contains heteroskedasticity and strong positive skews, which violate normality, symmetry, and homoskedasticitiy, and lends further credence to our stance of estimating a CLAD before moving on to other censored outcome models. Further, researchers should estimate a tobit model after obtaining this baseline estimate from CLAD, and then interpret the differences and present both sets of results toward gaining a more nuanced understanding of the underlying relationship(s) within the data. Marked departures of the tobit estimates from the baseline CLAD estimates are likely to be indicative of a violated assumption in application of the tobit model.

The tobit, SCLS, and CLAD models' differential sets of parametric assumptions incorporate different procedures for calculating the standard errors for their respective estimates. The tobit and SCLS packages are able to compute the standard errors analytically while the CLAD model incorporates a bootstrap routine[23] to compute asymptotically consistent estimates for the standard errors. These simulation analyses, in results not displayed here, found that the standard errors for the simulated sampling distributions for the respective parameters were approximated by the bootstrapped standard errors estimated by the CLAD package in STATA. This should provide further confidence for researchers applying the CLAD package appropriately toward analyzing censored outcome data. However, the standard errors for the three models followed the same general relationship wherein the tobit had the smallest standard errors, the CLAD had the largest standard errors, and the SCLS had the middle standard errors of the three models. This is to be expected, as the additional parametric assumptions of the tobit and SCLS models as compared to the CLAD model afford greater efficiency in the resulting estimates.

The primary advantage of applying the tobit model to censored outcome data is its efficiency in terms of producing smaller standard errors than the respective semiparametric approaches discussed here. So, when the underlying assumptions of the tobit model are met, it provides unbiased and comparatively the most efficient parameter estimates. However, this comparative advantage has the potential to backfire when the assumptions of the tobit model are untenable, and these assumptions are unlikely to be tenable for the kinds of data criminologists have typically applied and would like to apply tobit. In such contexts, the tobit model estimates will generally retain smaller standard errors than both SCLS and CLAD estimates, which results in tighter confidence intervals about the coefficient estimates produced by tobit as compared to SCLS and CLAD. However, as the simulation results have clearly shown, tobit estimates will be substantially biased in such cases. This results in the tobit model having a more precise and tighter confidence interval than the comparative models, *but the coefficient estimate of the tobit model is substantially displaced from the true population parameter*. In essence, tobit is more efficient at producing a biased estimate than SCLS and CLAD are at producing unbiased estimates in such contexts, as shown in Figure 6.

[22]SCLS could be argued as an alternative baseline model, but one would not be able to know whether the symmetry assumption of SCLS was met in a practical setting. As such, CLAD is recommended as a starting point as a result of these simulation analyses.

[23]A bootstrap routine involves performing a series of computations (default is 100 for the CLAD package in STATA) for the same value to produce a distribution of estimates from which an average is taken (see Efron & Tibshirani, 1994). The CLAD package takes advantage of this approach to calculate the standard errors for its estimates.

The efficiency gain is more than offset by the bias that accompanies it—a Faustian bargain that we do not think criminologists have to make.

Our empirical demonstration provided an application of the results of the simulation analyses in evaluating racial disparities in sentence lengths for the SCPS data. The sentence length measure was censored and not normally distributed, suggesting that both the tobit and OLS would be inappropriate model specifications. The results revealed large departures in the coefficient estimates between tobit and SCLS or CLAD, leading to greater faith in the CLAD and SCLS estimates given the known violation of an assumption of the tobit model. As performed in this empirical demonstration, researchers should look at the distributional form of their data, evaluate (and report) results from multiple methods that are appropriate for their data, and discuss the results in the context of the inherent statistical assumptions employed by the respective models.

Criminologists likely did not know the full effect of the bias accompanying the fragility of the tobit previously. Our review of published studies in the top three criminology journals clearly illustrated two points: (1) the tobit was the "go to" model in the presence of censored data, and (2) in all but a small number of cases, the tobit was used without explicit regard to its underlying assumptions. Scholars in our field should now be more aware of the substantial bias in tobit coefficients when model assumptions are violated, and should seek to apply CLAD in its stead when faced with censored outcome data. Only when the assumptions of the tobit model are found to be tenable, and the results of tobit and CLAD are substantively similar, should authors seek to take advantage of the efficiency gain that accompanies the tobit. As with the broader race and sentencing literature, it is time for criminologists to employ the new methodological tools of CLAD and SCLS that are better suited for their research questions and available data.

Notes on Contributors

Theodore Wilson, is an assistant professor in the School of Criminal Justice at the University at Albany. His research interests include offender decision-making, courtroom actor decision-making, and quantitative methods.

Tom Loughran, is a professor in the Department of Sociology and Criminology at Pennsylvania State University. His research interests include deterrence and offender decision-making, individuals' responses to sanctioning, and quantitative methods.

Robert Brame, is a professor in the Department of Criminology and Criminal Justice at the University of South Carolina. His research interests include criminal recidi-vism, the prevalence of criminal justice system involvement, and crime measurement.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Amemiya, T. (1985). *Advanced econometrics*. Cambridge: Harvard University Press.

Britt, C. L. (2009). Modeling the distribution of sentence length decisions under a guidelines system: An application of quantile regression models. *Journal of Quantitative Criminology*, *25*(4), 341–370.
Broidy, L. M. (2001). A test of general strain theory. *Criminology*, *39*(1), 9–36.
Bushway, S., Robert, B., & Paternoster, R. (1999). Assessing stability and change in criminal offending: A comparison of random effects, semiparametric, and fixed effects modeling strategies. *Journal of Quantitative Criminology*, *15*(1), 23–61.
Bushway, S., Johnson, B. D., & Slocum, L. A. (2007). Is the magic still there? The use of the Heckman two-step correction for selection bias in criminology. *Journal of Quantitative Criminology*, *23*(2), 151–178.
Carmichael, S., & Piquero, A. R. (2004). Sanctions, perceived anger, and criminal offending. *Journal of Quantitative Criminology*, *20*(4), 371–393.
Chay, K. Y., & Powell, J. L. (2001). Semiparametric censored regression models. *The Journal of Economic Perspectives*, *15*(4), 29–42.
Cohen, J., Gorr, W., & Singh, P. (2003). Estimating intervention effects in varying risk settings: Do police raids reduce illegal drug dealing at nuisance bars? *Criminology*, *41*(2), 257–292.
Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. London: Chapman and Hall.
Felson, R. B., & Staff, J. (2006). Explaining the academic performance-delinquency relationship. *Criminology*, *44*(2), 299–319.
French, M. T., McCollister, K. E., Kébreau Alexandre, P., Chitwood, D. D., & McCoy, C. B. (2004). Revolving roles in drug-related crime: The cost of chronic drug users as victims and perpetrators. *Journal of Quantitative Criminology*, *20*(3), 217–241.
Greene, W. (2005). Censored data and truncated distributions. *Palgrave Handbook of Econometrics*, *1*, 695–736.
Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In *Annals of Economic and Social Measurement*, (pp. 475–492). US: NBER.
Jarjoura, G. R., Triplett, R. A., & Brinker, G. P. (2002). Growing up poor: Examining the link between persistent childhood poverty and delinquency. *Journal of Quantitative Criminology*, *18*, 159–187.
Johnson, B. D., & Kurlychek, M. C. (2012). Transferred juveniles in the era of sentencing guidelines: Examining judicial departures for juvenile offenders in adult criminal court. *Criminology*, *50*(2), 525–564.
Kamerdze, A. S., Loughran, T., Paternoster, R., & Sohoni, T. (2014). The role of affect in intended rule breaking: Extending the rational choice perspective. *Journal of Research in Crime and Delinquency*, *51*(5), 620–654.
Kaye, D. H. (1982). Statistical evidence of discrimination. *Journal of the American Statistical Association*, *77*(380), 773–783.
King, G., Keohane, R. O., & Verba, S. (1994). *Designing social inquiry: Scientific inference in qualitative research*. New Jersey: Princeton University Press.
Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, *46*(1), 33–50.
Kramer, J. H., & Ulmer, J. T. (2009). *Sentencing guidelines: lessons from Pennsylvania*. Boulder, CO: Lynne Rienner Publishers.
Kurlychek, M. C., & Johnson, B. D. (2004). The juvenile penalty: A comparison of juvenile and young adult sentencing outcomes in criminal court. *Criminology*, *42*(2), 485–517.
Kurlychek, M. C., & Johnson, B. D. (2010). Juvenility and punishment: Sentencing juveniles in adult criminal court. *Criminology*, *48*(3), 725–757.
Kutateladze, B. L., Andiloro, N. R., Johnson, B. D., & Spohn, C. C. (2014). Cumulative disadvantage: Examining racial and ethnic disparity in prosecution and sentencing. *Criminology*, *52*(3), 514–551.
Loughran, T. A., Paternoster, R., & Weiss, D. (2012). Hyperbolic time discounting, offender time preferences and deterrence. *Journal of Quantitative Criminology*, *28*(4), 607–628.
Lu, H., & Drass, K. A. (2002). Transcience and the disposition of theft cases in China. *Justice Quarterly*, *19*(1), 69–96.

Manski, C. F. (1995). *Identification problems in the social sciences*. Cambridge, MA: Harvard University Press.
McDonald, J. B., & Nguyen, H. (2015). Heteroskedasticity and distributional assumptions in the censored regression model. *Communications in Statistics – Simulation and Computation*, *44*(8), 2151–2168.
McGloin, J. M., & Piquero, A. R. (2010). On the relationship between co-offending network redundancy and offending versatility. *Journal of Research in Crime and Delinquency*, *47*(1), 63–90.
McGloin, J. M., & Rowan, Z. R. (2015). A threshold model of collective crime. *Criminology*, *53*(3), 484–512.
McGloin, J. M., Sullivan, C. P., Piquero, A. R., & Bacon, S. (2008). Investigating the stability of co-offending and co-offenders among a sample of youthful offenders. *Criminology*, *46*(1), 155–188.
Messner, S. F., Deane, G. D., Anselin, L., & Pearson-Nelson, B. (2005). Locating the vanguard in rising and falling homicide rates across U.S. cities. *Criminology*, *43*(3), 661–696.
Nagin, D. S., Jones, B. L., Lima Passos, V., & Tremblay, R. E. (2016). Group-based multi-trajectory modeling. *Statistical Methods in Medical Research*, *27*(7), 2015–2023.
Nagin, D. S., & Odgers, C. L. (2010). Group-based trajectory modeling in clinical research. *Annual Review of Clinical Psychology*, *6*(1), 109–138.
Nagin, D. S., & Pogarsky, G. (2001). Integrating celerity, impulsivity, and extralegal sanction threats into a model of general deterrence: Theory and evidence. *Criminology*, *39*(4), 865–892.
Nagin, D. S., & Tremblay, R. E. (2005). What has been learned from group-based trajectory modeling? Examples from physical aggression and other problem behaviors. *The Annals of the American Academy of Political and Social Science*, *602*(1), 82–117.
Osgood, D. W., Finken, L. L., & McMorris, B. J. (2002). Analyzing multiple-item measures of crime and deviance II: Tobit regression analysis of transformed scores. *Journal of Quantitative Criminology*, *18*(4), 319–347.
Ousey, G. C., & Wilcox, P. (2007). The interaction of antisocial propensity and life-course varying predictors of delinquent behavior: Differences by method of estimation and implications for theory. *Criminology*, *45*(2), 313–354.
Pauwels, L., Weerman, F., Bruinsma, G., & Bernasco, W. (2011). Perceived sanction risk, individual propensity and adolescent offending: Assessing key findings from the deterrence literature in a Dutch sample. *European Journal of Criminology*, *8*(5), 386–400.
Piehl, A. M., & Bushway, S. D. (2007). Measuring and explaining charge bargaining. *Journal of Quantitative Criminology*, *23*(2), 105–125.
Piquero, A. R., & Pogarsky, G. (2002). Beyond Stafford and Warr's reconceptualization of deterrence: Personal and vicarious experiences, impulsivity, and offending behavior. *Journal of Research in Crime and Delinquency*, *39*(2), 153–186.
Pogarsky, G. (2004). Projected offending and contemporaneous rule-violation: Implications for heterotypic continuity. *Criminology*, *42*(1), 111–135.
Pogarsky, G., & Piquero, A. R. (2003). Can punishment encourage offending? Investigating the "resetting" effect. *Journal of Research in Crime and Delinquency*, *40*(1), 95–120.
Powell, J. L. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics*, *25*(3), 303–325.
Powell, J. L. (1986). Symmetrically trimmed least squares estimation for Tobit models. *Econometrica*, *54*(6), 1435–1460.
Roettger, M. E., & Swisher, R. R. (2011). Associations of fathers' history of incarceration with sons' delinquency and arrest among black, white, and Hispanic males in the United States. *Criminology*, *49*(4), 1109–1147.
Schlesinger, T. (2008). The cumulative effects of racial disparities in criminal processing. *The Advocate*, *13*(3), 22–34.
Siennick, S. E. (2011). Tough love? Crime and parental assistance in young adulthood. *Criminology*, *49*(1), 163–195.
Smith, D. A., & Brame, R. (2003). Tobit models in social science research: Some limitations and a more general alternative. *Sociological Methods and Research*, *31*(3), 364–387.

Spohn, C. (2015). Race, crime, and punishment in the twentieth and twenty-first centuries. *Crime and Justice*, *44*(1), 49–97.
Steffensmeier, D., & Demuth, S. (2001). Ethnicity and judges' sentencing decisions: Hispanic-black-white comparisons. *Criminology*, *39*(1), 145–178.
Stolzenberg, L., D'Alessio, S. J., & Eitle, D. (2013). Race and cumulative discrimination in the prosecution of criminal defendants. *Race and Justice*, *3*(4), 275–299.
Sullivan, C. J., McGloin, J. M., & Piquero, A. R. (2008). Modeling the deviant y in criminology: An examination of the assumptions of censored normal regression and potential alternatives. *Journal of Quantitative Criminology*, *24*(4), 399–421.
Sullivan, C. J., McGloin, J. M., Pratt, T. C., & Piquero, A. R. (2006). Rethinking the "norm" of offender generality: Investigating specialization in the short-term. *Criminology*, *44*(1), 199–233.
Sutton, J. R. (2013). Structural bias in the sentencing of felony defendants. *Social Science Research*, *42*(5), 1207–1221.
Sweeten, G. (2012). Scaling criminal offending. *Journal of Quantitative Criminology*, *28*(3), 533–557.
Tittle, C. R., Broidy, L. M., & Gertz, M. G. (2008). Strain, crime, and contingencies. *Justice Quarterly*, *25*(2), 283–312.
Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica*, *26*(1), 24–36.
Vazsonyi, A. T., Wittekind, J. E. C., Belliston, L. M., & Van Loh, T. D. (2004). Extending the general theory of crime to "the East:" Low self-control in Japanese late adolescents. *Journal of Quantitative Criminology*, *20*(3), 189–216.
Weerman, F. M., Bernasco, W., Bruinsma, G. J. N., & Pauwels, L. J. R. (2016). Gender differences in delinquency and situational action theory: A partial test. *Justice Quarterly*, *33*(7), 1182–1209.
Wiesner, M., Kim, H. K., & Capaldi, D. M. (2010). History of juvenile arrests and vocational career outcomes for at-risk young men. *The Journal of Research in Crime and Delinquency*, *47*(1), 91–117.
Wilhelm, M. O. (2008). Practical considerations for choosing between Tobit and SCLS or CLAD estimators for censored regression models with an application to charitable giving. *Oxford Bulletin of Economics and Statistics*, *70*(4), 559–582.
Wooldredge, J., Frank, J., Goulette, N., & Travis, L. (2015). Is the impact of cumulative disadvantage on sentencing greater for black defendants? *Criminology & Public Policy*, *14*(2), 187–223.
Wright, B. R. E., Caspi, A., Moffitt, T. E., & Silva, P. A. (2001). The effects of social ties on crime vary by criminal propensity: A life-course model of interdependence. *Criminology*, *39*(2), 321–352.
Wright, B. R. E., & Younts, C. W. (2009). Reconsidering the relationship between race and crime. *Journal of Research in Crime and Delinquency*, *46*(3), 327–352.
Zatz, M. S. (1987). The changing forms of racial/ethnic biases in sentencing. *Journal of Research in Crime and Delinquency*, *24*(1), 69–92.
Zatz, M. S. (2000). The convergence of race, ethnicity, gender, and class in court decisionmaking: Looking toward the 21st century. *Criminal Justice*, *3*, 503–552.