J Quant Criminol (2009) 25:341–370
DOI 10.1007/s10940-009-9066-x

ORIGINAL PAPER

Modeling the Distribution of Sentence Length Decisions Under a Guidelines System: An Application of Quantile Regression Models

Chester L. Britt

Published online: 13 June 2009

Abstract How should sentencing disparity be assessed when decisions are constrained under a sentencing guidelines system? Much of the debate over the measurement of sentence disparity under a guidelines system has focused primarily on using specific values from within the sentencing grid (e.g., minimum recommended sentence) or on using interaction terms in regression models to capture the non-additive effects of offense severity and prior record on length of sentence. In this paper, I propose an alternative method for assessing sentencing disparity that uses quantile regression models. These models offer several advantages over traditional OLS analyses (and related linear models) of sentence length, by allowing for an examination of the effects of case and offender characteristics across the full distribution of sentence lengths for a given sample of offenders. The analysis of the distribution of sentence lengths with quantile regression models allows for an examination of questions such as: Do offender characteristics, such as race or offense severity, have the same effect on sentence length for the 10% of offenders who receive the shortest sentences as they do for the 10% of offenders who receive the longest sentences? I illustrate the application and interpretation of these models using 1998 sentencing data from Pennsylvania. Key findings show that the effects of case and offender characteristics are variable across the distribution of sentence lengths, meaning that traditional linear models assuming a constant effect fail to capture important differences in how case and offender characteristics affect punishment decisions. I discuss the implications of these findings for understanding sentencing disparitites, as well as other possible applications of quantile regression models in the study of crime and the criminal justice system.

Keywords Quantile regression · Sentence length decisions · Distributional analysis · Linear models

C. L. Britt (✉)
College of Criminal Justice, Northeastern University, 360 Huntington Avenue,
Boston, MA 02115, USA
e-mail: c.britt@neu.edu

Introduction

When offenders are sentenced under a guidelines system, how should the researcher assess the degree of disparity in sentencing decisions, given that the judge's choices for sentences are constrained? A substantial body of research on sentencing decisions within a guidelines context has focused primarily on assessing unwarranted disparity in sentencing decisions using characteristcs of the offender, the case, and the judge. Within the larger body of research, a smaller number of studies have examined the effect of the guidelines grid on sentencing decisions (see, e.g., Bushway and Piehl 2001; Engen and Gainey 2000a, b; Miethe and Moore 1986; Mustard 2001; Ulmer 2000). Common to these approaches is the expectation that information about the structure of the sentencing guidelines grid can be used in the form of statistical controls that will reduce or eliminate spurious effects in sentencing research and provide for more accurate tests for sentencing disparities.

Although the research focusing on the structure of the sentencing grid offers substantial improvements over prior tests for disparity, missing from this work is any attention to the *distribution* of sentence length decisions that may be crucial to understanding how a guidelines system may affect sentence length decisions. No prior research has focused directly on analyzing the full distribution of sentence length decisions that would allow for a discussion of how legal and extra-legal characteristics affect sentence length decisions at various points in the distribution. For example, to the extent that a characteristic such as the race of the offender has an effect on sentence length, once relevant legal background characteristics have been controlled statistically, does the effect of race operate in the same way among less serious sentences as it does among more serious sentences? In considering the distribution of sentence length decisions, we might begin by looking at the effect of race for the shortest 10% of sentence length decisions, then move to the median sentence length, followed by the longest 10% of sentence length decisions to see whether the effect of race on the length decision changes. As I explain below, the typical analysis of sentencing decisions assumes that race, as well as every other variable included in the typical multivariate statistical model, has a constant effect across the distribution of sentence length decisions. The constant effect of a characteristic, such as race, means that it does not matter whether we are looking at relatively short sentences (e.g., less than 12 months) or relatively long sentences (e.g., 10 years or more), the effect of race is assumed to be constant across the full range of sentence lengths.[1]

The inclusion of an interaction effect between any two case or offender characteristics does not solve this problem, either. All that an interaction effect will do is allow the effect of one case or offender characteristic to vary by the level of another case or offender characteristic. The effects are still fixed and ignore the the distribution of sentence lengths. For example, Spohn and Cederblom (1991) and Ulmer (1997) include interaction effects between race and offense severity. These interaction terms allowed the researchers to describe the (fixed) effect of offense severity for each racial group. The interaction effects tell us nothing about whether the effect of race or the effect of offense severity differentially affects sentence length at different segments of the distribution of sentence lengths.

[1] Similarly, a small but growing body of research has shown the effects of case and offender characteristics to differentially affect the jail v. prison decision (see, e.g., Holleran and Spohn 2004).

Distributional methods offer an alternative framework for analysis that allows the researcher to assess whether the effects of the independent variables on the dependent variable are constant across the distribution of the dependent variable. In roughly the last 10 years, there has been rapid growth in the availability and use of various distributional methods, particularly in research on the differential effects of background characteristics, such as education, on income and wealth (see, e.g., Buchinsky 1998; Handcock and Morris 1999). Included in the distributional methods framework are parametric quantile regression models that permit estimation of unique effects of the independent variables at any specified point in the distribution of the dependent variable. The primary purpose of this paper is to illustrate the application and interpretation of quantile regression techniques to the study of sentence length decisions. As I show below, the use of quantile regression techniques presents new findings regarding the effects of offender and case characteristics on sentence length decisions that cannot be easily discerned in more traditional analyses of sentence length decisions.

The Research Problem

The literature on sentencing guidelines has often noted how the creation and implementation of sentencing guidelines reduced judicial discretion, but failed to eliminate unwarranted disparity (see, e.g., Tonry 1996). More recent research on sentencing disparities within sentencing guidelines systems—both state and federal—has attempted to push the discussion of disparities further by asking whether previous findings are artifacts of the sentencing grids used by the states or the federal system (Bushway and Piehl 2001; Engen and Gainey 2000a, b; Miethe and Moore 1986; Mustard 2001; Ulmer 2000). The concern with the sentencing guidelines grid, regardless of whether it is a state or a federal grid, has focused primarily on whether evidence for or against unwarranted disparities is simply a consequence of the way the sentencing grid was constructed.

To illustrate the problem confronting sentencing research on offenders sentenced under a guidelines system, consider Fig. 1, which presents the sentencing grid for Pennsylvania (effective 13 June 1997). Within any level of offense severity (indicated by the rows of the grid and has column label OGS, for Offense Gravity Score—a measure of offense severity), as one moves from the lowest prior record score to the highest prior record score, the minimum and maximum recommended sentences increase, but the increases are not always linear. For example, when the prior record score has a value of 0 and the offense severity level has a value of 6, the minimum and maximum incarceration sentences are 3 and 12 months, respectively.[2] Moving across columns of increasing prior record score, the minimum sentence increases 3 months for each increase in prior record score until the prior record score takes on a value of 5, at which point the minimum increases by 6 months. The maximum sentence length increases by two months for each increase in prior record score until the prior record score takes on a value of 4 when it increases by 3 months. At a prior record score of 5, the maximum sentence increases by 6 months. Further complicating the measurement of disparity, the range of months in the recommended sentence length varies.[3] Using the same row of the sentencing grid: when the prior

[2] Boot camp incarceration is also an option for the sentencing judge, but since a boot camp incarceration would not typically be included in an analysis of sentence length decisions, it is excluded from the range of sentence lengths noted here.

[3] Bushway and Piehl (2001) noted a similar issue in their analysis of data from Maryland.

§303.16. Basic Sentencing Matrix PRIOR RECORD SCORE 5th Edition (6/13/97)

Level	OGS	Example Offenses	0	1	2	3	4	5	RFEL	REVOC	AGG/MIT
LEVEL 5 State Incar	14	*Murder 3* *Inchoate Murder/SBI*	72-240	84-240	96-240	120-240	168-240	192-240	204-240	240	+/- 12
	13	*Inchoate Murder/no SBI* Drug Del. Result in Death PWID Cocaine, etc. (>1,000 gms)	60-78	66-84	72-90	78-96	84-102	96-114	108-126	240	+/- 12
	12	*Rape* *IDSI* *Robbery (SBI)* *Robbery/car (SBI)*	48-66	54-72	60-78	66-84	72-90	84-102	96-114	120	+/- 12
	11	*Agg Asslt (SBI)* *Voluntary Manslaughter* Sexual Assault PWID Cocaine,etc.(100-1,000 gms)	36-54 BC	42-60	48-66	54-72	60-78	72-90	84-102	120	+/- 12
	10	*Kidnapping* *Arson (person inside)* *Agg Asslt (att. SBI)* *Robbery (threat. SBI)* Agg. Indecent. Asslt Causing Catastrophe(F1) PWIDCocaine,etc.(50-<100 gms)	22-36 BC	30-42 BC	36-48 BC	42-54	48-60	60-72	72-84	120	+/- 12
	9	*Robbery/car (no SBI)* *Robbery (F1/F2)* *Burglary (home/person)* *Arson (no person)*	12-24 BC	18-30 BC	24-36 BC	30-42 BC	36-48 BC	48-60	60-72	120	+/- 12
LEVEL 4 State Incar/ RIP trade	8 [F1]	Agg Asslt (BI w/DW) Agg Asslt (att. BI w/DW) Invol. Mansl. (when DUI) Hom. by Vehicle (when DUI) Theft (>$100,000) PWID Cocaine,etc. (10-<50 gms)	9-16 BC	12-18 BC	15-21 BC	18-24 BC	21-27 BC	27-33 BC	40-52	NA	+/- 9
LEVEL 3 State/ County Incar RIP trade	7 [F2]	Robbery (inflicts/threatens BI) Burglary (home/ no person) Statutory Sexual Assault Theft (>$50,000-$100,000) Sexual Abuse/Child (take photo) PWID Cocaine,etc.(2.5-<10 gms)	6-14 BC	9-16 BC	12-18 BC	15-21 BC	18-24 BC	24-30 BC	35-45 BC	NA	+/- 6
	6	Invol. Mansl.(when no DUI) Hom. by Vehicle (when no DUI) Burglary (not home/person) Theft (>$25,000-$50,000) Arson (property) PWID Cocaine,etc.(<2.5 gms)	3-12 BC	6-14 BC	9-16 BC	12-18 BC	15-21 BC	21-27 BC	27-40 BC	NA	+/- 6
LEVEL 2 County Incar RIP RS	5 [F3]	Burglary (not home/no person) Corruption of Minors Robbery (prop by force) Firearms (loaded) Theft (>$2000-$25,000) PWID (1-<10 lb of marij)	RS-9	1-12 BC	3-14 BC	6-16 BC	9-16 BC	12-18 BC	24-36 BC	NA	+/- 3
	4	Indecent assault Forgery (money, stock, etc.) Firearms (unloaded) Crim Trespass (breaks in)	RS-3	RS-9	RS-<12	3-14 BC	6-16 BC	9-16 BC	21-30 BC	NA	+/- 3
	3 [M1]	Simple Assault Terr. Threats Theft ($200-$2000) Retail Theft (3rd) DUI (M1) Drug Poss.	RS-1	RS-6	RS-9	RS-<12	3-14 BC	6-16 BC	12-18 BC	NA	+/- 3
LEVEL 1 RS	2 [M2]	Theft ($50-<$200) Retail Theft (1st ,2nd) DUI (M2) Bad Checks	RS	RS-2	RS-3	RS-4	RS-6	1-9	6- <12	NA	+/- 3
	1 [M3]	Most Misd. 3's;Theft (<$50) Drug Paraph. Poss. Small Amount Marij.	RS	RS-1	RS-2	RS-3	RS-4	RS-6	3-6	NA	+/- 3

Key: Level 1 = Purple, Level 2 = White, Level 3 = Blue, Level 4 = Yellow, Level 5 = Green, AGG/MIT = Tan

1. Yellow (Level 4) and Blue (level 3) shaded areas of the matrix indicate restrictive intermediate punishments may be imposed as a substitute for incarceration.
2. When restrictive intermediate punishments are appropriate, the duration of the restrictive intermediate punishment program shall not exceed the guideline ranges.
3. When the range is RS through a number of months (e.g. RS-6), RIP may be appropriate.
4. All numbers in sentence recommendations suggest months of minimum confinement pursuant to 42 PA.C.S. §9755(b) and §9756(b).
5. Statutory grades in brackets correspond with the omnibus OGS for the grade.

BC	=	boot camp	PWID	=	possession with intent to deliver
CNTY	=	county	RIP	=	restrictive intermediate punishments
INCAR	=	incarceration	RS	=	restorative sanctions
Italics	=	Three Strikes Offense	RFEL	=	repeat felony 1 and felony 2 offender category
< ; >	=	less than;greater than	REVOC	=	repeat violent offender category

Fig. 1 Pennsylvania Sentencing Guidelines Grid, 1997

record score is 0, the range is 9 months, it then shrinks by 1 month for each increase in prior record score up to a prior record score of 3, where the range has been reduced to 6 months and remains that way through a prior record score of 5. Similar kinds of variations in the magnitude of the increase of minimum and maximum sentence lengths and the range of recommended sentences appear in other cells of the sentencing grid.

The variation in recommended minimum and maximum sentence lengths points to the importance of considering every possible cell in a sentencing grid to best test for unwarranted disparities (Engen and Gainey 2000a). For example, suppose that we have three offenders with an offense gravity score (OGS) of 7 and who have each received a minimum sentence of 18 months. The first offender, whose prior record score is 2, has received the maximum recommended sentence for this cell in the grid. The second offender, whose prior record score is 3, has received a sentence at the midpoint for this cell in the grid. The third offender, whose prior record score is 4, has received the minimum recommended sentence for this cell in the grid. All three of the sentences fall within the recommended range of sentence lengths, and would reflect an appropriate degree of discretion exercised by the sentencing judge. Yet, at the same time, they reflect a disparity of sentences relative to the range of sentences that each offender *could* have received and raise questions about how to test for this type of disparity in a multivariate analysis of sentence length decisions. For example, we might want to know how these 18 month sentences compare to other sentences within each cell. Is a sentence of 18 months relatively harsh or relatively lenient, given the other offenders who had identical offense severity and prior record scores?

A number of papers have proposed a variety of ways to better account for these kinds of patterns in various state and federal sentencing guidelines grids. All of the suggestions proposed thus far rely on adding one or more variables to the regression model used to estimate sentence length. Three major variants of the proposals include: (1) the addition of an interaction effect between offense severity and prior record measures (Engen and Gainey 2000a), (2) the addition of multiple dummy variables to represent *every* combination of offense severity and prior record (Engen and Gainey 2000a; Mustard 2001), and (3) the presumptive sentence indicated in the guidelines grid. Much of the debate over accounting for the structure of the guidelines grid has focused on how best to measure the presumptive sentence. Some research suggests that the minimum expected sentence within each cell of a guidelines grid works best with sentencing data from Pennsylvania (Ulmer 2000), while other research indicates that the midpoint (Engen and Gainey 2000a; Miethe and Moore 1986) or the logarithm of the midpoint (Bushway and Piehl 2001) of sentence lengths within each cell of a guidelines grid provide a better measure of presumptive sentence length.

The logic to each of these approaches is based on the idea that by controlling statistically for unique features within a sentencing grid, a better test for unwarranted disparity in sentence length decisions can be accomplished. The expectation is that by introducing statistical controls into a multivariate model of sentence length, the results will provide more accurate measures of the impact of characteristics such as age, race, and gender. For example, as the midpoint for cells within a sentencing grid increases, sentence lengths should also increase, controlling for other offender and case characteristics. Evidence presented in each of the papers using data from several different states and the federal system offers at least modest support for each position. For example, Engen and Gainey's (2000a, b) analyses of data from Washington state and Bushway and Piehl's (2001) analysis of data from Maryland confirm that their choice of the midpoint and log-midpoint of the cell, respectively, in the guidelines grid is a more appropriate choice for modeling sentence length disparity. Mustard's (2001) use of multiple dummy variables appears to offer an effective way of accounting for patterns of increasing sentence lengths in the federal sentencing guidelines, while Ulmer's (2000) analysis of data from Pennsylvania indicates that the minimum recommended sentence length is the best measure for that state's data. More recently, Ulmer et al. (2007) used the difference between the guideline recommended minimum and the mandatory minimum. Each of these approaches has something to offer and have been incorporated in more recent studies of sentencing decisions (see, e.g., Johnson 2005, 2006; Steen et al. 2005; Ulmer and Bradley 2006; Ulmer and Johnson 2004). In addition to the apparent simplicity of adopting one or more of the approaches suggested in this work, each approach also appears to improve the explained variance of the dependent variable.

Interestingly, none of this work has used information about the midpoint or the minimum expected sentence in such a way that it provides a better answer to the questions related to disparity. Rather, this information is used strictly as a form of statistical control in whatever multivariate analysis is performed. Put simply, once the midpoint (or minimum) of a cell is accounted for statistically, the focus turns to whether other case and offender characteristics have statistically significant effects on sentence length. The inability to test for disparity more directly using this information can again be illustrated with the three hypothetical offenders who each received an 18-month sentence. If we use the midpoint as the presumptive sentence length, we gain a slightly different view of the severity of the sentence. The first offender would have received a sentence 3 months longer than the midpoint, the second offender a sentence equal to the midpoint, and the third offender a sentence 3 months shorter than the midpoint. If the offenders are identical on every other case characteristic, then the use of the midpoint implies some kind of disparity in the length of sentence received due to some other characteristic of the offenders. As I explain below, however, there are important conceptual questions that these approaches ultimately leave unanswered, and which cannot be answered with the kinds of statistical models that have been used thus far in the analysis of sentence length decisions.

Methodological Issues

Although each of the suggestions proposed in prior research offers a way of improving the explained variance in a multivariate statistical model, and presumably reduces the chances of spurious effects of other variables in a statistical model, they are limited in the same way. Regardless of the proposed change, each continues to use a statistical model that (1) estimates constant effects of the independent variables on the conditional mean of the dependent variable and (2) assumes a normal distribution of errors with constant variance. As an illustration, consider the exchange between Engen and Gainey (2000a) and Ulmer (2000). They do not debate the appropriateness of OLS to estimate the expected sentence length, conditional on the independent variables, but focus instead on the best measure for recommended sentence length. Bushway and Piehl (2001) opt for the log-midpoint and eschew the OLS technique in favor of a tobit model that controls statistically for censoring, but their use of a tobit model in this case still estimates a conditional mean of sentence length based on values of the independent variables.

The problem with using a statistical model that predicts the conditional mean of the dependent variable is that it assumes the effect of every independent variable is *constant* across the full range of the dependent variable. Why would this be problematic in the context of sentencing decisions? Consider again the Pennsylvania sentencing grid in Fig. 1. The range of recommended sentence lengths within each cell increases at the same time that the recommended minimum and maximum sentence lengths increase. This means that a *differential* effect of offense severity and prior record is built into the guidelines system that is contingent on the level of each of these characteristics. What this implies is that these two characteristics should have widely variable effects on sentence length decisions. Offenders who receive relatively short sentences should show a modest effect of offense severity and prior record, while among offenders who receive relatively long sentences, the effects of offense severity and prior record should be substantially larger, but the magnitude of this effect is not the same across all levels of offense severity and prior record.

As a further illustration of this problem, suppose that we estimate a very simple model of sentence length using only offense severity and prior record as the independent variables. The OLS model predicts the mean of sentence length conditional on the level of offense severity and prior record score. As shown in the Pennsylvania sentencing grid (Fig. 1), as well as in Engen and Gainey's (2000a) analysis of Washington data, there is not a linear trend in the minimum, the maximum, the midpoint, or the range of sentences from one cell to another in a sentencing grid. The variation that we observe in the structure of sentencing grids raises doubts about the appropriateness of coefficients in an analysis where we assume constant effects of the independent variables on the dependent variable. OLS regression coefficients indicate the level of change in the conditional mean of the dependent variable, given a one-unit change in the independent variable across the full range of the dependent variable. In an analysis of sentence length decisions that relies on offense severity and prior record, the assumption of a constant effect seems questionable in light of the structure of the sentencing grid. Similarly, the inclusion of a measure of presumptive sentence length will improve the prediction of the dependent variable, but the statistical model still implies that there are constant effects of the independent variables that are not accurate (e.g., that for every one-unit increase in the presumptive sentence, the expected sentence length changes by some fixed amount). Similar concerns arise even when an interaction term for offense severity and prior record and/or quadratic terms are included in the model (Engen and Gainey, 2000a, b).

The distribution of cases within each cell of a guidelines grid is also problematic. Since OLS, as well as the tobit model, assume that the errors around the regression line or

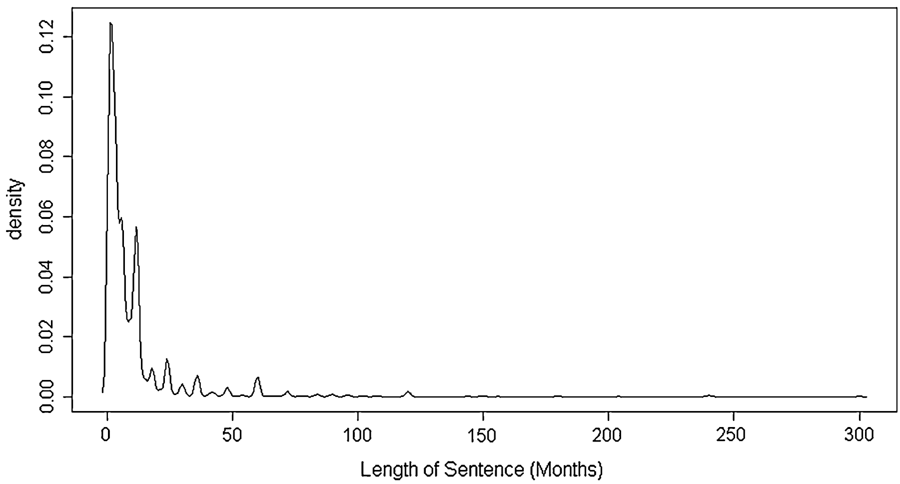

Fig. 2 Density plot of sentence lengths in Pennsylvania, 1998

regression plane are normally distributed with constant variance. Engen and Gainey's (2000a) analysis of Washington sentencing data illustrates very well the non-normality of the errors from a regression model (see their Figs. 1, 2). A more detailed look at sentencing outcomes in Pennsylvania suggests a related problem. Figure 2 displays a density plot for sentence lengths of at least 1 month for cases sentenced in 1998. Note that there are peaks at 6-month intervals, indicating a preference of judges for a period of time that has an easy interpretation in either whole years or half-years. That the sentence lengths cluster at obvious values that are easily converted to years or half-years indicates that judges are likely not using the full range of sentence lengths within each cell, which in turn produces a distribution of cases within each cell of the sentencing grid that is likely far from normal with constant variance. Put differently, there are likely highly skewed distributions of cases within each cell, clustering around a sentence length indicating a whole year or a half-year. A highly skewed distribution of cases within each cell implies that the use of OLS and the estimation of the conditional mean of the dependent variable is likely not the best measure of central tendency in a multivariate analysis. Transforming sentence length by taking the logarithm will help to reduce the positive skew in the sentence length distribution (see, e.g., Bushway and Piehl 2001; Wheeler et al. 1982). However, the logged value of sentence length will not address other idiosyncratic features of the distribution of sentence lengths (e.g., modes at six-month intervals).

All of this points to a substantively meaningful, and practically important, question in research on sentence length decisions: What accounts for the relative severity or leniency that some offenders experience across different segments of the distribution of sentence lengths? If we pick a particular point in the distribution of sentence lengths, then how do case and offender characteristics help us to understand the length of sentence received by each offender? One of the key goals in any analysis of sentencing decisions is to try to gain a better understanding of the dynamics of decision-making in the criminal courts. Clearly, any sentencing study can only hope to approximate the effects of case and offender characteristics on the overall punishment process. Many other parties—prosecutors, defense attorneys, victims—may have some kind of input, either directly or indirectly, to the judge's final sentencing decision. If we analyze only the outcomes—sentence length decisions—we inherently miss the complexity of the processes at work in criminal trial courts (Ostrom et al. 2007). Bearing these limits in mind, an approach that moves us beyond a single point estimate for the effect of case and offender chartacteristics should point us in the direction of a more nuanced understanding of how case and offender characteristics affect sentence length decisions.

Quantile Regression Models

Quantile regression models were initially proposed by Koenker and Bassett (1978) as a method of robust regression that would account for a non-normal distribution of error terms and as a test for heteroskedastic error terms (see also, Gilchrist 2000; Hao and Naiman 2007; Koenker and Bassett 1982; Koenker 2005).[4] More recently, quantile regression models have been used to study income inequality (Buchinsky 1994, 1998; Handcock and Morris 1999; Mello and Perrelli 2003; Miller 2005), earnings (Eide and Showalter 1999; Eide et al. 2002; Mora and Davila 2006; Walker 2000), and wealth (Conley and Galenson

[4] A quantile refers to a percentile in the distribution of a variable. For example, a quantile of 0.70 would refer to the same point in a distribution as the 70th percentile.

1998; Walker 2000) distributions. In cases where there may be a substantial number of zeroes on the dependent variable—representing censoring of the dependent variable—quantile regression models also allow for unbiased estimates of the effects of the independent variables on the dependent variable at higher quantiles, where the influence of the zeroes will be minimal (Powell 1984, 1986).[5]

The Statistical Model

Koenker and Hallock (2001) described the estimation of the coefficients of a quantile regression model by explaining how it could be reduced to an optimization problem (for detailed descriptions, see Koenker 2005; Koenker and Bassett 1978). At the heart of the optimization problem is the minimization of asymmetrically weighted absolute residuals. By asymmetrically weighted residuals, they mean assigning different weights to positive and negative residuals. This results in the following minimization equation for unconditional quantiles:

$$\min_{\xi \in \Re} \sum_{i=1}^{n} \rho_\tau(y_i - \xi),$$

where $\rho_\tau(\cdot)$ is the absolute value function that gives the τth sample quantile, y_i is the observed value of the dependent variable, and ξ is the predicted value.

As a point of reference, least squares regression minimizes the following equation:

$$\min_{\mu \in \Re} \sum_{i=1}^{n} (y_i - \mu)^2$$

to obtain an unconditional population mean, $\mu = \mathrm{E(Y)}$. By replacing μ with a parametric function $\mu(x,\beta)$ (McCullagh and Nelder 1989), we obtain the conditional mean of Y (i.e., E(Y | x)) by minimizing

$$\min_{\mu \in \Re} \sum_{i=1}^{n} (y_i - \mu(x, \beta))^2.$$

For quantile regression, the same type of progression is used to estimate the conditional quantile by replacing ξ with a parametric function $\xi(x, \beta)$ and specifying τ (the quantile). Thus, if we wanted to estimate the conditional median function, we would set $\tau = 0.5$ (i.e., the 50th percentile) and minimize the function to obtain parameter estimates for the effects of x on y at the median. To obtain other conditional quantiles, we simply change the value of τ and minimize

$$\min_{\xi \in \Re} \sum_{i=1}^{n} \rho_\tau(y_i - \xi(x, \beta)).$$

[5] This is a situation most sentencing researchers are also confronted with: Those offenders who are not incarcerated are effectively a "0" on the dependent variable and then excluded from the analysis. Quantile regression techniques could be used to address this concern without relying on various sample selection models that may be more or less appropriate for the analysis of sentencing decisions (Bushway et al. 2007).

Interpretation of the Coefficients

In general, the coefficients from an OLS regression model may be interpreted as the change in the conditional mean of the dependent variable based on a unit change in the independent variable. The coefficients from a quantile regression model may be interpreted in a similar manner, with the primary difference being that the coefficients from a quantile regression model represent the effect of a unit change in the independent variable at some defined point of the conditional distribution of the dependent variable (i.e., the quantile). For example, if we had set $\tau = 0.5$—the median of the distribution of the dependent variable—then we would interpret the coefficient as the change in the conditional median of the dependent variable based on a unit change in the independent variable. Similarly, if $\tau = 0.75$, then we would interpret the coefficient as the change in the conditional 0.75 quantile (75th percentile) given a unit change in the independent variable.

Part of what quantile regression coefficients permit is an examination of how or whether the effects of the independent variables change across different segments of the distribution of the dependent variable. When the dependent variable has a distribution that is highly skewed, such as income, earnings, and sentence lengths for convicted offenders, it is possible for the independent variables to have differential effects on the dependent variable that are conditional on some designated point of the distribution of the dependent variable.

We can return to the three hypothetical offenders who received 18 month sentences to illustrate the utility of this approach. Suppose we use only offense severity and prior record score to predict sentence length, reflecting the structure of the sentencing guidelines grid. Within each cell of the grid—each combination of offense severity and prior record score—we can compute the distribution of cases within that cell. In turn, this allows us to compute percentiles for each offender's sentence within their respective cells. The percentile values for each sentence length decision reflect the relative severity or leniency of that particular sentence compared to other sentences in that cell. For example, if we discover that within a single cell an 18 month sentence is at the 25th percentile, we know that this sentence represents leniency relative to other offenders whose offense severity and prior record scores placed them in the same cell. Quantile regression coefficients allow us to take the analysis one step further by then assessing the relative distributions of cases across different combinations of the independent variables. Consider the cell where the 18-month sentence represents the 25th percentile. If we increase prior record score by one, what is the expected 25th percentile in this new cell? How much did the expected sentence length change? Or did it remain the same (i.e., is an 18-month sentence still at the 25th percentile)?

Figure 3 provides a more concrete example for interpreting the results from a quantile regression model. Figure 3 presents the boxplots of the distribution of sentence lengths for white, black, and Hispanic offenders in Pennsylvania in 1998 who received a sentence of at least 1 month. Boxplots are a simple, but powerful, way of conveying a great deal of information about the distribution of the variable being plotted. The gray box represents the inter-quartile range—the range from the 25th percentile (lower end of the gray box) to the 75th percentile (upper end of the gray box). The median is represented by the horizontal line inside the gray box. The mean for each race is represented by the + symbol. The lines extending from the gray box represent what are variably referred to as the "fence" or "whiskers" of the plot. The limits—the fence—represent 1.5 times the inter-

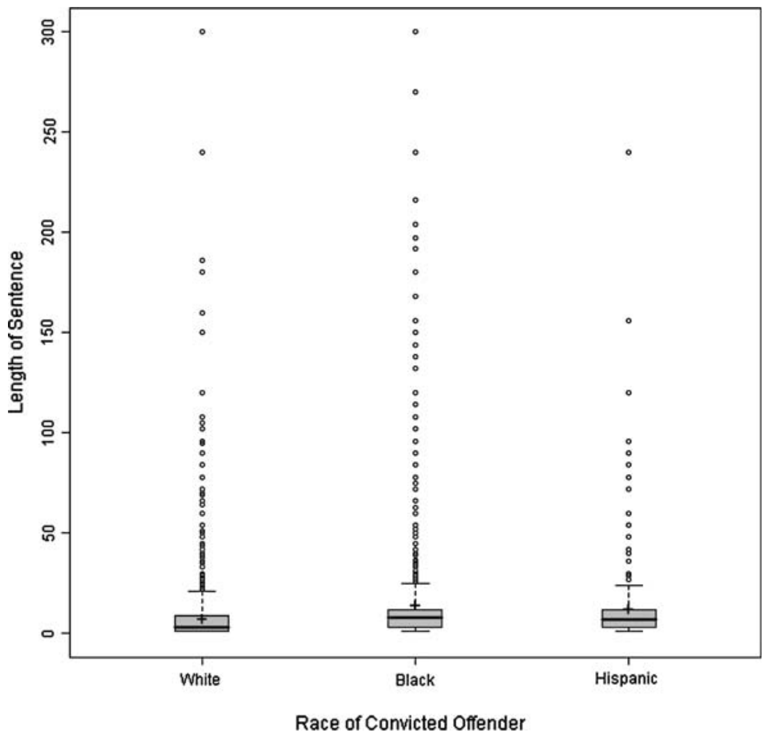

Fig. 3 Box plots of sentence lengths by race-ethnicity

quartile range. The circles appearing beyond the fence represent relatively extreme values and in some contexts, may be treated as potential outliers.[6]

The inter-quartile range highlights some of the potential differences and similarities in sentence lengths. At the 0.25 quantile, the sentence length is 3 months for white offenders, and 4 months for both black and Hispanic offenders. The median sentence length (i.e., the 0.50 quantile) is 6 months for white offenders, and 9 months for both black and Hispanic offenders. By the 0.75 quantile, there appears to be near parity in the sentences by race, where the sentence length is 11.5 months for white offenders and 12 months for both black and Hispanic offenders. Interestingly, if we extend our focus to the 0.90 quantile, the differences across race-ethnicity become more pronounced, where the sentence length is 21 months for white offenders, 36 months for black offenders, and 30 months for Hispanic offenders.

Alternatively, we could interpret the pattern of results in Fig. 3 to show that at the 0.25 quantile, black and Hispanic offenders receive sentences about 1 month longer than white offenders. At the median, black and Hispanic offenders receive sentences that are 3 months longer than those for white offenders. At the 0.75 quantile, black and Hispanic offenders receive sentences that are only about 0.5 months longer than white offenders. Yet, at the 0.90 quantile, black offenders receive sentences that are about 15 months longer than those received by white offenders and and 6 months longer than those received by Hispanic

[6] For a comprehensive discussion of the construction and interpretation of boxplots, see Tukey (1977).

offenders. Regardless of the approach used to interpret the distributional patterns in Fig. 3, the boxplots reveal substantial differences in the distribution of sentence lengths by race-ethnicity. It is also important to keep in mind that the distribution of sentence lengths presented in Fig. 3 does not account for the severity of the conviction offense, the severity of prior record of the offender, or any other legally relevant case or offender background characteristic. It is possible that once a variety of legal factors have been controlled in a statistical model, the differences would disappear. This is another way that quantile regression techniques provide such a useful form of analysis, and allows for testing whether these differences in distributional patterns will disappear, once other relevant legal characteristics have been taken into account.

Location Shift and Location-Scale Hypotheses

The location shift hypothesis is a test for whether a single coefficient could be used to capture the effect of an independent variable on the dependent variable across the distribution of the dependent variable. The location shift model is equivalent to an OLS regression model where the coefficient is constant, meaning that for a fixed change in the independent variable, the conditional mean of the dependent variable is expected to change by a fixed amount. In the context of quantile regression results, if the effect of an independent variable varies around some constant value across the distribution of the dependent variable, then the effect of the variable is limited to one of shifting the location—the level—of the dependent variable.

The location-scale hypothesis is akin to testing for heteroskedasticity in the OLS model. The location-scale hypothesis reflects the possibility that as the value of the dependent variable changes in magnitude, the variance of the residuals—the spread of cases around the regression plane—will also change, sometimes increasing, sometimes decreasing. Consequently, the location-scale hypothesis tests for the possibility of a constant effect of the independent variable on the dependent variable, accounting for the increasing (decreasing) spread of cases across the distribution of the dependent variable. Koenker and Xiao (2002a) note that the model intercept in quantile regression represents a normalized version of the residuals and to the extent the effects of the independent variable follow the same pattern as the model intercept, they represent simple shifts in location and scale of the dependent variable and residuals.

Koenker and Xiao (2002a, b) present a test and critical values for the location shift and the location-scale hypotheses for the full model and the individual coefficient estimates. The critical values for the tests depend on the number of coefficients estimated, the range of quantiles, and the desired level of statistical significance; sample size has no bearing on the test statistic or the critical value. Practically, the calculated test statistic is compared to the critical value, and if the test statistic is greater than the critical value the null hypothesis of location shift or location-scale is rejected. The rejection of the location shift hypothesis implies that the OLS model, based on conditional means, does not provide a good representation of the effects of the independent variables on the dependent variable. Rejection of the location-scale hypothesis implies not only that the OLS model does not represent the effects of the independent variables well, but the shape of the conditional distribution of the dependent variable changes in ways not modeled by OLS, even after accounting for heteroskedasticity. In other words, there is no simple, straightforward transformation of the residuals. Koenker and Xiao (2002b) includes critical values for varying levels of statistical significance, number of coefficients, and range of quantiles covered for the test and provide the basis for such tests below.

Data and Methods

Data

I use data collected under the Pennsylvania sentencing guidelines for the year 1998 and made available through the National Archive of Criminal Justice Data.[7] All of the offenders were sentenced based on the 1997 change to the PA guidelines. In the following analyses, I use data on 14,977 sentence length decisions of at least 1 month that represent all 67 of Pennsylvania's counties. In 1998, there were 48,883 offenders sentenced under the 1997 guidelines grid. Of these offenders, 25,626 received an incarceration sentence of at least 1 day. Of all offenders receiving an incarceration sentence, 6,988 received a sentence of less than 1 month (of these, 4,833 received a sentence of 0.07 months—2 days in a 31-day month). Of the remaining 18,638 offenders who received a sentence of at least 1 month, 3,661 cases were dropped due to (1) method of case disposition and (2) race. In order to make the interpretation of the results more straightforward, I eliminated cases that were not disposed of through a guilty plea or a trial (bench or jury). I also restricted the race-ethnicity of cases to those who were identified as white, black, or Hispanic.

Measures

Dependent Variables

Sentence Length I use two measures of sentence length reflective of suggestions to account for concerns that have periodically been expressed about the positive skew to sentence length decisions (see, e.g., Bushway and Piehl 2001; Wheeler et al. 1982). *Length of sentence* is coded as the minimum number of months the offender was sentenced to confinement in either a county jail or a state prison. *Log-sentence length* is the logarithm of length of sentence that is expected to reduce the positive skew of sentence length. In addition, the use of the log of sentence length will permit a straightforward interpretation of the effect of a one-unit change in the independent variables as a percentage change in sentence length.

Independent Variables

Demographic Characteristics Race is measured with two dummy variables. *Black* is coded as 1 if the offender was black; 0 otherwise. *Hispanic* is coded as 1 if the offender was Hispanic (any race); 0 otherwise. *Female* is coded as 1 for offenders who were women; 0 for men. The offender's *age* is coded as the number of years.

Criminal History The offender's criminal history is measured with the *prior record score*. This is an 8-category scale with a range from 0 to 7 that was developed by the Pennsylvania Commission on Sentencing in conjunction with judges and prosecutors in Pennsylvania. Prior record score is a weighted scale that measures both the number and the severity of prior felony and misdemeanor convictions; higher values indicate a greater number of prior convictions and more serious prior criminal acts.

[7] The choice of data set was somewhat arbitrary. My goal was to illustrate the application and the interpretation of quantile regression models with data that had been analyzed in other published studies. Hopefully, interested readers would then make comparisons between the results reported here and those in other published works using the same or similar data.

Case Characteristics Offense severity is measured as a 14 category ordered scale, with a range from 1 to 14; higher values indicate more severe conviction offenses.

Case disposition is measured with two dummy variables coded as 1 if the offender was convicted through a *jury trial* or a *bench trial*; 0 if the offender was convicted through a plea (either negotiated or non-negotiated), as reported to the Pennsylvania Commission on Sentencing.

Type of offense is measured with three dummy variables coded as 1 if the offender was convicted of a *violent offense*, a *property theft offense*, or a *drug offense*, and coded as 0 if the offender was convicted of any other offense.

The offender's *number of current convictions* has values ranging from 1 to 8 (or more).

An offender convicted of an offense that carried a *mandatory minimum* sentence was coded as 1 if there was a mandatory minimum; 0 otherwise.

Findings

Distribution of Sentence Lengths

In order to gain some perspective on the differential distribution of sentence lengths in the 1998 Pennsylvania data, two sets of running boxplots are used to display the range of sentence lengths by offense severity score and prior record score, since these are the two legally designated characteristics that define the dimensions of the sentencing guidelines grid. Figures 4 and 5 present running boxplots of sentence length and log-sentence length by offense severity score. Figures 6 and 7 display running boxplots for sentence length and log-sentence length by prior record score.

Figures 4 and 5 both show that as offense severity score increases, the median sentence length increases, especially after an offense severity score of 6. More important to the use of OLS methods of analysis, which assumes a constant distribution of the dependent variable across levels of the independent variable, the distribution of sentence lengths increases dramatically as offense severity score increases. The magnitude of the increased

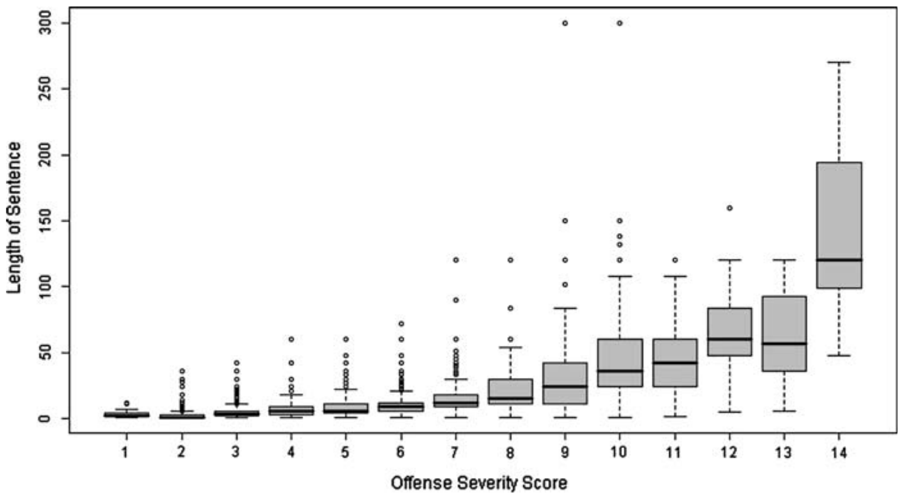

Fig. 4 Box plots of sentence lengths by offense severity

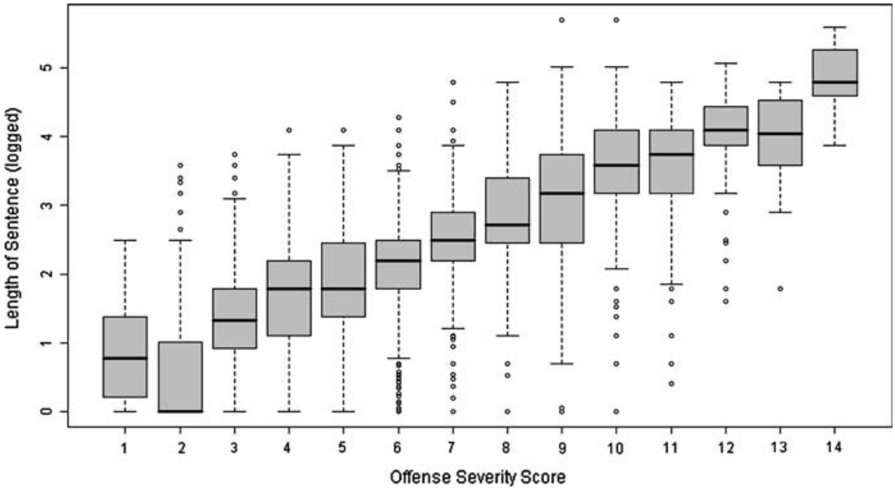

Fig. 5 Box plots of sentence lengths (logged) by offense severity

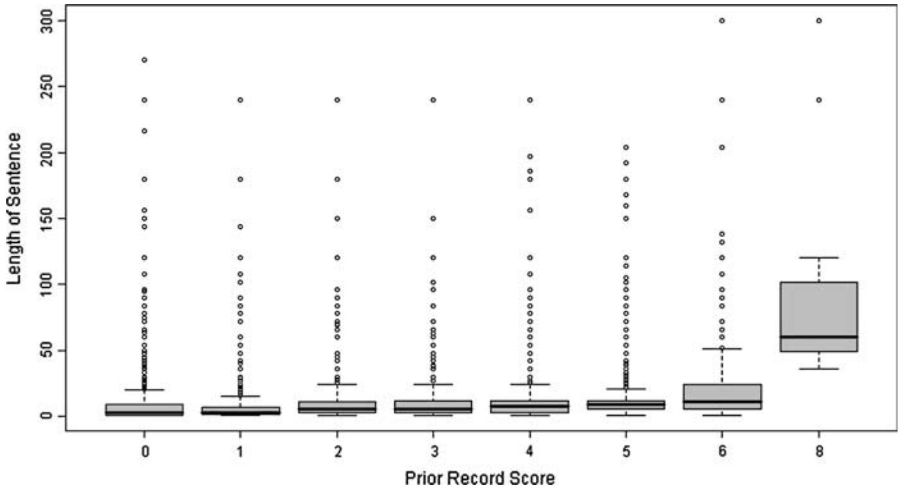

Fig. 6 Box plots of sentence lengths by prior record score

spread is perhaps best conveyed by looking at the inter-quartile range (i.e., the difference between the 25th to the 75th percentiles) represented by the lower and the upper ends of the gray box. For less serious offenses, the inter-quartile range is relatively small. Then, starting at about offense severity level 4, the range begins to increase rapidly across the remaining levels of offense severity. It is important to note while the logged value of sentence length has reduced the positive skew represented by the individual boxplots, the plots in Fig. 5 clearly show the distribution of log-sentence lengths to vary in ways that call into question the use of linear models that rely on conditional mean functions.

Figures 6 and 7 show a pattern of varying distributions of sentence lengths across prior record score, although not quite as dramatic as we observed in Figs. 4 and 5. In Fig. 6, we see that as prior record score increases, the median sentence length increases slightly from

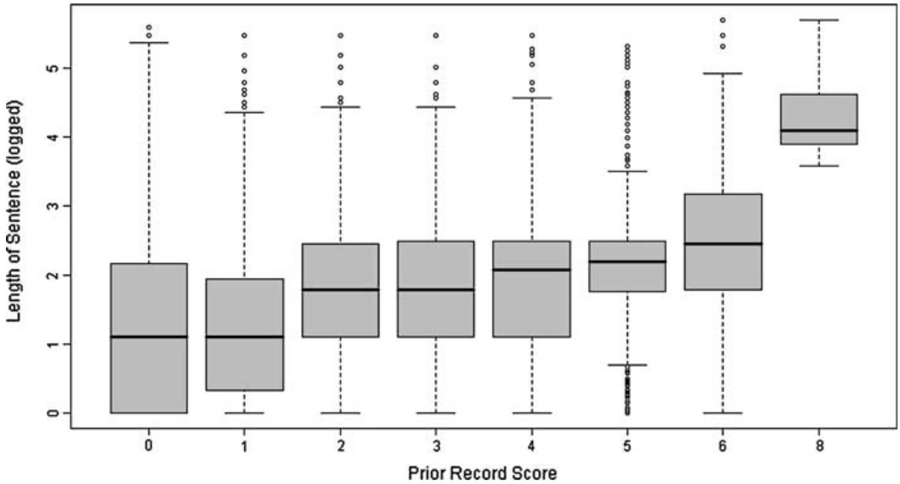

Fig. 7 Box plots of sentence lengths (logged) by prior record score

a prior record score of 0 to a prior record score of 6, but more importantly shows a variable spread of cases in the inter-quartile range. Figure 7 also shows a pattern of variability in the distribution of log-sentence length. Again, the logged value of sentence length has reduced the magnitude of the positive skew of the sentence length decisions by prior record score, but the pattern of variability in the spread of sentence length decisions is clear and continues to cast doubt on the use of typical linear models.

Taken together, Figs. 4–7 show that regardless of how sentence length is measured—number of months or logged number of months—we find a pattern where the distribution of lengths is not equal across the different levels of offense severity and prior record scores. In part, this reflects the non-linearities in sentence length decisions that have been factored into sentencing guidelines systems and noted in prior guidelines-based research (Bushway and Piehl 2001; Engen and Gainey 2000a; Ulmer 2000). At the same time that these figures confirm the non-linearity of sentence lengths within a guidelines system, the variable dispersion of sentence lengths in these two figures indicate the difficulty in assuming that the coefficients from standard linear statistical models based on the conditional mean of the dependent variable (i.e., OLS and Tobit) are constant across the distribution of sentence lengths.

Quantile Regression Results

Length of Sentence

Table 1 presents the OLS and quantile regression results for sentence length regressed on the full set of independent variables.[8] The OLS results for sentence length show all legally

[8] The quantreg library (Koenker 2008), a component of the R system (R Development Core Team 2003), was used to estimate all quantile regression models. Readers interested in testing quantile regresison models for their research questions may also estimate these models in Stata and SAS. Stata's version 10.0 (Stata Corporation 2007) includes quantile regression procedures that allow for the estimation of individual quantiles or a range of quantiles. SAS's version 9.1 includes an experimental quantile regression procedure (PROC QUANTREG) that can be downloaded from the SAS web site (http://www.sas.com). PROC

relevant case and offender variables to have a statistically significant effect ($p < .001$).[9] As expected, offense severity, prior record, number of current convictions, conviction through a bench or a jury trial, conviction of a violent or a property offense, and conviction of an offense with a mandatory minimum increase sentence length, while conviction of a drug offense (compared to the miscellaneous offense category) decreases sentence length. In regard to demographic characteristics of offenders, the only significant effect shows black offenders receiving longer sentences than white offenders.[10] The OLS results serve two purposes for this paper: (1) to validate the data analyzed here, since the pattern of OLS results presented in Table 1 is generally consistent with other studies using late-1990s Pennsylvania sentencing data but alternative modeling strategies (e.g., Ulmer and Johnson 2004) and (2) to provide a comparison for the quantile regression results below.[11]

The quantile regression coefficients represent the effect of a one-unit change in the independent variable on the dependent variable at the τ quantile.[12] For example, when $\tau = .10$, the effect of offense severity is 0.92, meaning that a one-unit increase in the offense severity score is expected to increase sentence length by about 0.9 months at the 10th percentile of sentence length. Viewed in a slightly different way, at the lower end of the sentence length distribution—that end of the distribution where the sentences are relatively short—the effect of offense severity is modest and results in the expected sentence length increasing by a small amount. At the higher end of the distribution, where the punishments are substantially longer, the magnitude of the effect of offense severity is much larger. For example, at $\tau = .90$, the effect of offense severity is 5.73, meaning that a one-unit change in offense severity at the 90th percentile of sentence length results in the expected sentence length increasing by almost 6 months.

The magnitude of each of the quantile regression coefficients presented in Table 1 appears to vary across quantile. In addition to the effect of offense severity being more than six times larger at the 90th percentile than at the 10th percentile, the effects of many of the remaining variables also show what appears to be substantial variation across the different quantiles. Are these apparent differences meaningful? As noted above, there are two tests for the stability of the coefficients: location shift and location-scale. A test for location shift is a test of whether the quantile regression coefficients can be constrained to a single value—equivalent to an OLS coefficient—that tracks changes in the level (location) of the dependent variable across levels of the independent variables, but does not change in magnitude. The location-scale test assesses the degree to which nonconstant error variance affects the coefficient estimates. To the extent that the increased (decreased) spread of

Footnote 8 continued
QUANTREG (updated in January 2009) also estimates individual quantiles as well as a range of quantiles. By using the ODS feature in SAS, it is possible to plot the quantile regression process in a manner consistent with the results displayed in Figs. 8 and 9 in this paper. However, at the time of this writing, neither Stata's nor SAS's quantile regression procedures allow for testing the location and location-scale hypotheses.

9 All of the coefficients noted as not statistically significant have $p > 0.10$.

10 Basic collinearity diagnostics did not indicate any problems with the inclusion of this set of case and offender characteristics.

11 To the extent that there are minor differences between the results reported here and those published in prior research using the Pennsylvania data, the differences are due to the large number of cases with very short sentences (i.e., less than one month) that were excluded from the analyses reported here.

12 A reviewer raised a concern about collinearity diagnostics for the quantile regression results. At this point, there are no unique ways of testing for collinearity in a multivariate quantile regression model. An indicator that collinearity is likely not a problem is reflected by the pattern of most case and offender characteristics remaining statistically significant across the range of quantiles.

Table 1 OLS and quantile regression results

Variable	Percentile					
	OLS	10th	25th	50th	75th	90th
Intercept	−19.83	−2.95	−4.65	−6.81	−9.68	−9.83
Prior record score	1.88	0.38	0.63	1.01	1.39	1.42
Offense severity	5.01	1.01	1.62	2.52	4.09	6.19
Black	0.94	0.36	0.50	*0.11*	0.31	0.70
Hispanic	*−0.02*	0.17	0.37	*0.06*	*0.07*	*−0.46*
Female	*−0.30*	−0.25	−0.22	*−0.04*	*0.00*	*0.46*
Age ($\times 10^{-4}$)	*−8.00*	*0.00*	*0.00*	*0.00*	*0.00*	*−10.00*
Number of conviction offenses	0.28	0.05	0.13	0.20	0.32	0.48
Jury trial	16.10	0.99	3.00	8.15	23.32	31.37
Bench trial	3.56	0.99	1.06	1.94	3.41	6.68
Violent offense	3.72	0.70	0.65	1.06	1.07	*−1.61*
Property theft offense	1.58	0.48	0.47	0.48	0.39	*−1.42*
Drug offense	−1.95	*−0.13*	*0.15*	*−0.04*	−1.86	−5.72
Mandatory minimum	9.55	0.88	1.64	2.58	2.18	*−0.53*

Note: Italicized coefficients are *not* statistically significant at $p < .001$

Table 2 Location shift and location scale hypothesis test results

Variable	Location shift	Location-scale
Prior record score	18.59**	1.87
Offense severity	13.39**	4.82**
Black	3.60**	3.44**
Hispanic	1.65	1.28
Female	1.91	3.63**
Age ($\times 10^{-4}$)	0.99	1.18
Number of conviction offenses	3.43**	0.63
Jury trial	22.28**	1.08
Bench trial	3.30**	2.26*
Violent offense	1.81	2.10*
Property theft offense	1.18	2.22*
Drug offense	2.18*	2.73**
Mandatory minimum	5.95**	6.88**
Full model	73.85**	50.90**

* $p < 0.05$

** $p < 0.01$

observations around the regression plane affects the values of the coefficients, the quantile regression estimates for each of the independent variables should change in ways very similar to the intercept of the model. The results for these tests are presented in Table 2.

Koenker and Xiao (2002a, b) have produced tables that present critical values for the test statistic for the location shift and the location-scale models. For the analysis here, with thirteen independent variables and spanning quantiles 0.10–0.90, the critical values for the test statistic for the full model are 13.90 at the 5% significance level and 15.59 at the 1% significance level. For each individual coefficient, the critical value at a 5% level of significance is 2.102 while at a 1% level of significance it is 2.640. The calculated model test statistic is 73.849 for the location shift model and 50.901 for the location-scale model, indicating that neither the typical OLS model (location shift) nor the OLS model accounting for heteroskedasticity (location-scale) adequately represent the effects of the independent variables on sentence length across the full distribution of sentence lengths.

Table 2 also reports the test statistics for each individual coefficient. Koenker and Xiao (2002a) warn that the tests for individual coefficients are not independent and need to be interpreted cautiously. However, the test statistics for individual coefficients provide a rough guide in regard to the coefficients that have the greatest impact on the model test statistic. From the results in Table 2, we see that the coefficients for jury trial, prior record score, and offense severity respectively have the greatest effects on the location shift results, while the coefficients for mandatory minimum, offense severity, female and black have the greatest relative impact on the location-scale results.

To aid in the interpretation of the results presented in Tables 1 and 2, a compact way of presenting information on the OLS and the quantile regression coefficients is with a series of effects displays that appear in Fig. 8, Panels a through n.[13] In each panel, the value of the OLS coefficient is represented by a solid line, the values representing a 95% confidence interval around the OLS coefficient are represented with dashed lines. Note that since these lines are horizontal, it indicates that the value of the OLS coefficient is the same, regardless of the quantile of the distribution of sentence length. The quantile regression coefficients are represented by the black dots, and a 95% confidence band around the quantile regression coefficient estimates is represented by the gray band.

The effects displays reveal that the OLS coefficients do a particularly poor job of indicating the effects of the independent variables for offense severity, prior record score, number of current convictions, jury trial conviction, bench trial conviction, violent, property, or drug offense conviction, and mandatory minimum. The quantile regression effect of prior record score reveals a pattern of increasing effect (Panel b), but never reaches the value of the OLS coefficient. After exhibiting a nearly linear increase in value through the 0.70 quantile, the value of the quantile regression coefficient decreases at the 0.80 quantile, but starts to increase again in the 0.90 quantile. The effect of offense severity (Panel c) reveals a similar pattern of increasing effect. At the lower quantiles, the effect of offense severity is relatively modest (about 1 month for each one unit increase in offense severity score), but as the quantile increases, the effect of offense severity increases in a generally linear fashion. Interestingly, it is not until about the 0.80 quantile that the effect of offense severity is similar to the OLS coefficient for offense severity.

In a similar way, we can decompose the effect of race on sentence length. The OLS coefficients show black offenders are expected to receive sentences that are about 1 month longer than those for white offenders, and no difference between white and Hispanic offenders. When we look at the effects plots for the quantile regression coefficients, we see that the difference in sentence lengths for black and white offenders tends to be of about 0.1–0.5 months. It is not until we move to the 0.90 quantile of sentence length decisions that the quantile regression coefficients start to approximate the OLS coefficients. The difference between Hispanic and white sentence lengths is not significantly different from zero for the full distribution. These results suggest that the difference in sentence lengths between black and white offenders observed in the OLS coefficients is being driven by the most serious sentence lengths.

[13] In much of the literature on quantile regression coefficients, the full range of effects plotted in these figures is referred to as the quantile regression process.

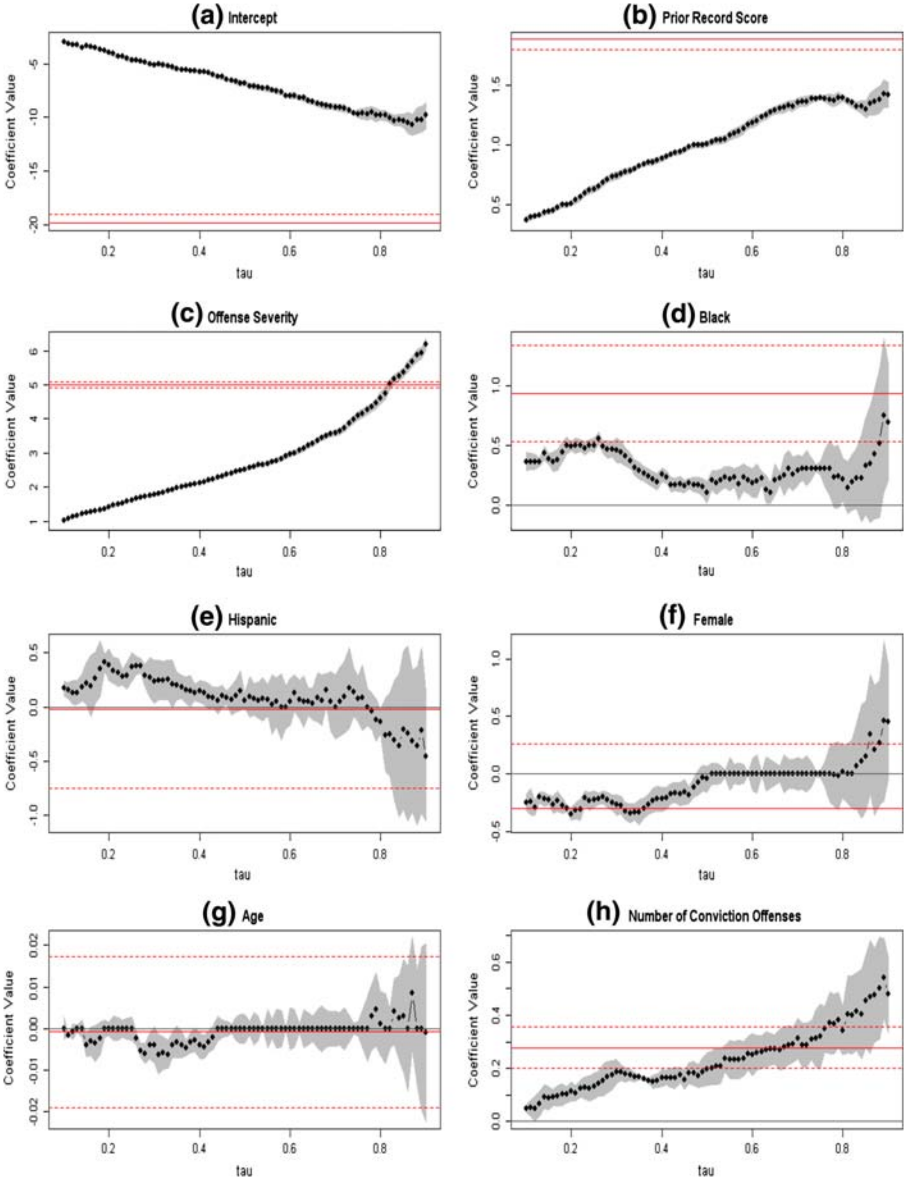

Fig. 8 Quantile regression process for sentence length

Perhaps the greatest discrepancy between an OLS coefficient and the quantile regression coefficients appears for the effect of being convicted through a jury trial (relative to pleading guilty). The OLS effect indicates that offenders convicted through a jury trial receive sentences that are almost 16 months longer than those for offenders who have been convicted through a guilty plea. Yet, when we look at the quantile regression coefficients, we see that the effect ranges from about 1 month (at the 0.10 quantile) to more than 31 months (at the 0.90 quantile). Intuitively, the quantile regression results would seem to

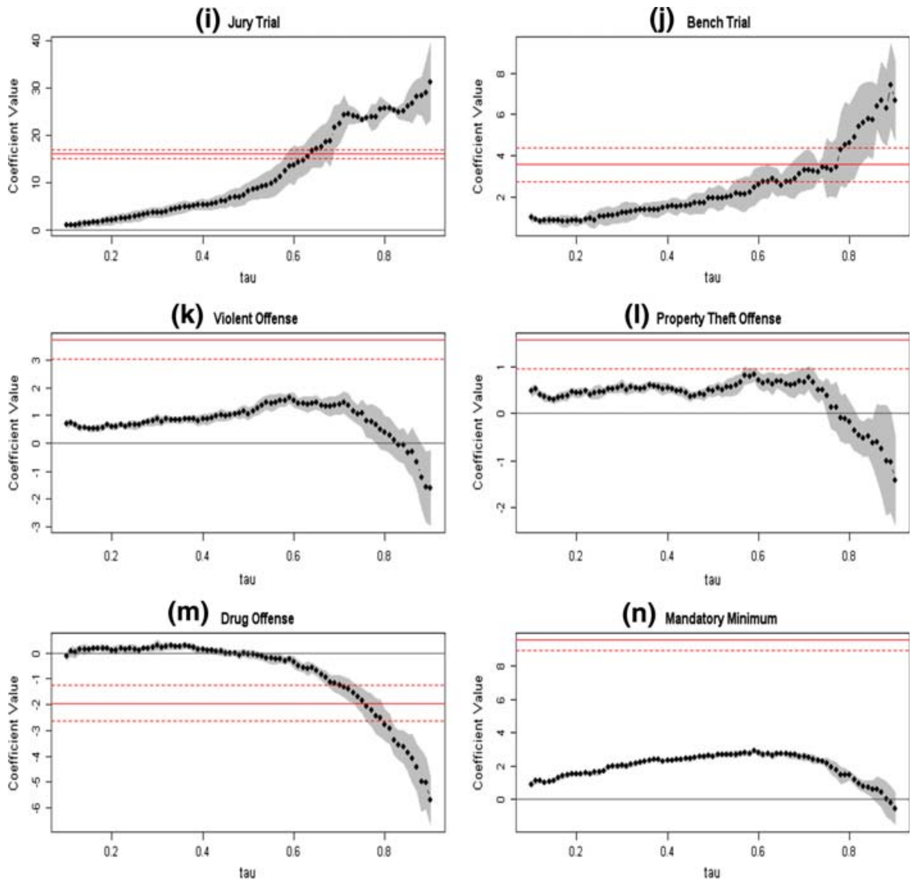

Fig. 8 continued

offer a better reflection of the effect of jury trial conviction—all offenders convicted through a jury trial will not receive sentences that are more than 1 year longer than those received by offenders who plead guilty. Many of the sentences in the Pennsylvania sentencing grid have maximum sentence lengths less than 16 months, so it would seem highly improbable that judges would routinely add approximately 1.3 years to each offender's sentence length, independent of the severity of the conviction offense and prior record score. Rather, we would expect that the "trial penalty" would increase as the maximum sentence length also increased.[14]

To summarize the results thus far, the quantile regression results indicate that the effects of standard predictors of sentence length have differential effects on the sentence length decision that depend on which part of the length distribution is being analyzed. The

[14] Ulmer and Bradley's (2006) analysis of Pennsylvania sentencing data included interaction effects for their jury trial variable with offense severity, prior record score, rape conviction offense, robbery conviction offense, and offender's race (black). Although their inclusion of an interaction effect was a positive step forward in the analysis of the "trial penalty," it still falls to the criticisms noted above regarding fixed effects of the independent variables across the distribution of sentence lengths.

Table 3 OLS and quantile regression results for logged sentence length

Variable	Percentile					
	OLS	10th	25th	50th	75th	90th
Intercept	−0.08	−0.42	−0.28	−0.12	0.11	0.29
Prior record score	0.07	0.06	0.07	0.08	0.07	0.06
Offense severity	0.13	0.14	0.14	0.14	0.13	0.13
Black	0.04	0.06	0.04	0.03	0.02	0.02
Hispanic	0.04	0.06	*0.04*	0.03	*0.02*	*0.01*
Female	−0.05	−0.03	−0.04	*0.00*	−0.05	−0.03
Age ($\times$ 10^{-4})	*−2.00*	*0.00*	*4.70*	*0.00*	*0.00*	*0.00*
Number of conviction offenses	0.01	*0.01*	0.01	0.01	0.01	0.01
Jury trial	0.21	0.12	0.19	0.20	0.19	0.24
Bench trial	0.09	0.10	0.11	0.07	0.07	0.07
Violent offense	0.07	*−0.01*	*0.03*	0.07	0.11	0.09
Property theft offense	0.07	0.03	0.05	0.06	0.09	0.08
Drug offense	0.05	*0.00*	0.04	0.05	0.06	*0.04*
Mandatory minimum	−0.06	0.02	*−0.01*	−0.17	−0.07	*−0.05*

Note: Italicized coefficients are *not* statistically significant at $p < .001$

variation in the effects of the quantile regression coefficients reflects the change in distribution of sentence length across the levels of the independent variables—variation that cannot be modeled easily with a statistical procedure, such as OLS, that models the conditional mean of sentence length by assuming a fixed effect of each of the independent variables. The results also indicate, particularly for the measures of offense severity and jury trial, that an OLS analysis of sentence length decisions provides highly biased estimates that do not reflect the underlying relationships.

Logged Length of Sentence

Table 3 presents the OLS and quantile regression results for log-sentence length regressed on the full set of independent variables. The OLS results for sentence length show all variables to have a statistically significant effect ($p < .001$), except for age of the offender ($p > 0.50$). The only differences between the OLS results in Tables 1 and 3 show that the coefficients for female and Hispanic offenders are now statistically significant ($p < 0.001$). Otherwise, most of the other effects of the independent variables are comparable across the two tables in regard to direction of effect and statistical significance, as well as the relative magnitudes of the effects. The one notable change was the effect of mandatory minimum is now negative, suggesting a 6% decrease in sentence length.

Similar to the quantile regression results presented in Table 1, the pattern of results in Table 3 shows the effects of the case and offender characteristics to vary across quantile. As we might expect, based on the reduction in skew observed in the running boxplots for log-sentence length displayed in Figs. 5 and 7, the relative variablility of the quantile regression coefficients is smaller in Table 3 than in Table 1. The general pattern of statistical significance of the coefficients in Table 3 is also similar, though not identical, to that in Table 1.

The results for the location shift and location-scale hypothesis tests are presented in Table 4. The critical values for the full model and the individual coefficients will be the

Table 4 Location shift and location scale hypothesis test results for logged sentence length

Variable	Location shift	Location-scale
Prior record score	8.56**	14.77**
Offense severity	6.16**	1.26
Black	1.31	3.55**
Hispanic	0.84	0.61
Female	9.42**	21.45**
Age ($\times\ 10^{-4}$)	1.41	2.06
Number of conviction offenses	1.31	0.68
Jury trial	1.79	3.29**
Bench trial	1.81	1.50
Violent offense	2.53*	2.20*
Property theft offense	3.53**	1.86
Drug offense	2.13*	1.77
Mandatory minimum	6.86**	24.48**
Full model	94.20**	244.56**

* $p < 0.05$

** $p < 0.01$

same as in Table 2. The evidence again shows both the location shift and the location-scale hypotheses are rejected for the full model, where the test statistics are 94.202 and 244.560, respectively. Substantively, what these results indicate is the inability of the OLS model and fixed coefficients to represent the effects of the case and offender characteristics across the distribution of log-sentence length, even after adjusting for heteroskedasticity. In regard to individual coefficients, we see that female, prior record score, mandatory minimum, and offense severity have the greatest impact on the location shift, respectively. The location-scale results indicate that mandatory minimum, female, and prior record score have the greatest effects, respectively.

Figure 9 (panels a to n) presents effects displays for the quantile regression results for log-sentence length comparable to those in Fig. 8. Similar to the pattern of results displayed in Fig. 8, the quantile regression coefficients tend to be quite different from the OLS coefficients across the full distribution of sentence lengths. At the same time, there are some interesting differences between the results in Figs. 8 and 9, which represent consequences of using the original (untransformed) sentence length variable or the logged value. For example, the effect of prior record score is clearly curvilinear in Fig. 9, where the effect of prior record score increases from the 0.10 quantile to about the median (where it peaks) and then declines close to the initial value at the 0.90 quantile. More directly, at the 0.10 quantile, each unit change in prior record score is expected to increase sentence length by about 5.8%. By the median ($\tau =\ 0.50$), the effect increased to about 8.0%, but then declined to about 6.1% at the 0.90 quantile.

We observe similar curvilinear patterns for many of the other quantile regression coefficients. For example, the coefficient for female offender is at its lowest value at about $\tau =\ 0.20$, increases in value until the median (where it is not significantly different from zero), decreases through about $\tau =\ 0.60$, but then starts to increase in value again. Substantively, this suggests that at the median of the distribution of sentence length, female offenders receive sentences that are proportionally indistinguishable from those received by male offenders, controlling for other case and offender characteristics. For offenders in either tail of the distribution of sentence lengths, female offenders receive sentences that are proportionally shorter than those received by male offenders.

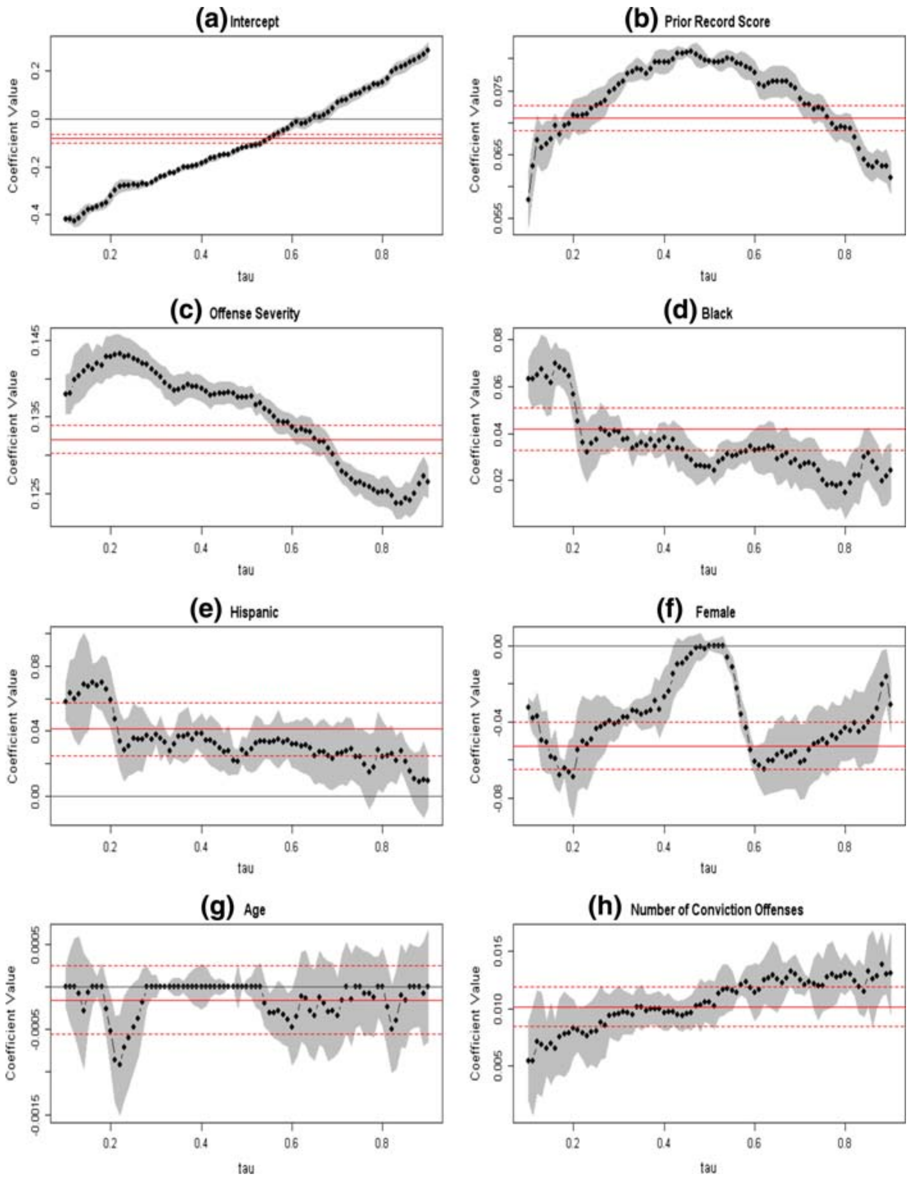

Fig. 9 Quantile regression process for sentence length (logged)

The effect of mandatory minimum starts out as positive at the lower end of the distribution. At $\tau = 0.10$, a mandatory minimum results in sentence lengths that are about 2% longer than those offenses that do not have a mandatory minimum, likely reflecting the mandatory punishments associated with various DUI offenses. The effect of a mandatory minimum then decreases through the median where offenses with a mandatory minimum are about 18% shorter, followed by an increase in magnitude until the quantile regression coefficients are comparable to the OLS coefficient starting at about $\tau = 0.70$.

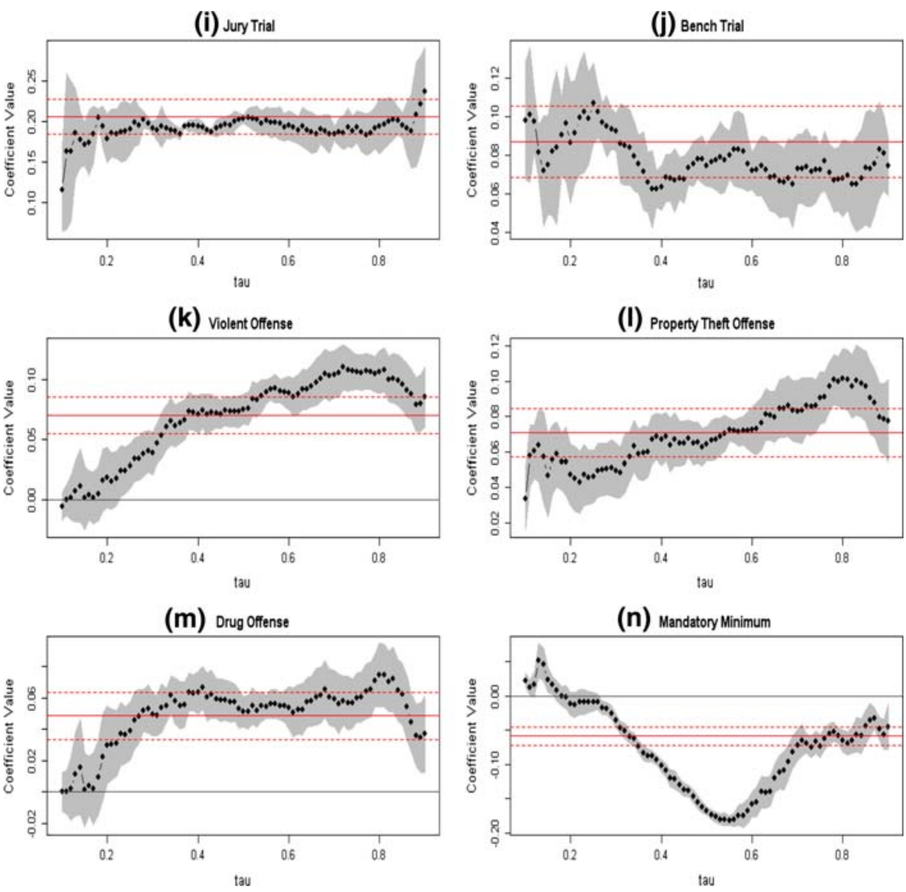

Fig. 9 continued

Substantively, the overall pattern of quantile regression coefficients for mandatory minimum indicate that it has a different meaning among offenders sentenced to relatively short sentences (less than 3 months) than among offenders sentenced to longer and more serious sentences.

The effect of jury trial on log-sentence length observed in Fig. 9 is much less dramatic than that found in Fig. 8 using untransformed sentence length as the dependent variable. The pattern for jury trial seen in Fig. 9 shows that the OLS coefficient does a much better job of measuring the effect of a jury trial conviction on log-sentence length. Where the OLS and quantile regression coefficients differ most is in the tails of the sentence length distribution. More directly, the OLS coefficient indicates that offenders convicted through a jury trial will receive sentences that are about 21% longer than individuals who have plead guilty. At the lower end of the sentence length distribution, the quantile regression results show that at $\tau = 0.10$, a jury trial conviction results in a sentence that is expected to be about 12% longer than those received by offenders who have plead guilty. From approximately $\tau = 0.25$ to $\tau = 0.80$, the OLS and quantile regression coefficients are nearly indistinguishable. At $\tau = 0.90$, however, the quantile regression results indicate that a jury trial conviction results in a sentence that is 24% longer that that received by offenders who have plead guilty.

In sum, the results presented above show the effects of case and offender characteristics to vary by the specific point in the distribution of sentence length (i.e., the quantile) that is being analyzed. The pattern of variability in the effects of the case and offender characteristics holds regardless of whether the dependent variable being analyzed was sentence length or log-sentence length. The key differences in the results for sentence length and log-sentence length are found in the specific patterns of variability and in the magnitude of the discrepancy between the OLS and the quantile regression coefficients.

Summary and Conclusion

The primary focus of this paper has been the analysis of the distribution of sentence lengths as a means of testing for disparity in sentence lengths within a sentencing guidelines system, where the minimum and maximum sentence lengths are presumed to be conditional on the severity of the conviction offense and the prior record score. A problem that has confronted every researcher analyzing sentencing data collected in a state with a guidelines system has been a concern with how to incorporate information about the structure of the guidelines grid into multivariate models analyzing sentence length decisions. Many of the suggestions that have been offered in prior research focus on adding and/or transforming variables in an OLS regression model or tobit model. Although each of these methods offers an improvement over ignoring the structure of the guidelines grid altogether, each is incomplete in the sense that there is an expectation that all of the independent variables will have constant effects on the dependent variable throughout the distribution of the dependent variable. Since the distribution of sentence lengths within each cell of a guidelines grid will tend to be highly skewed, methods such as OLS will be sensitive to the more extreme values within each cell, resulting in undue influence on the magnitude of the regression coefficients. OLS methods will also be less appropriate for discerning the unique and changing effects an independent variable may have at different points of a distribution. I have illustrated in this paper how we can use quantile regression techniques to analyze the distribution of sentence lengths that circumvents many of the difficulties that remain by using more typical OLS regression models.

The results presented above show that while the OLS effects of each of the independent variables on sentence length is assumed to be constant across the distribution of sentence length, the quantile regression results show that this is typically not the case. At the lower end of the distribution—shorter sentence lengths—the effects of the independent variables tend to be small, as we would expect. The range of possible punishments is small for less serious offenses and offenders, so the apparent effect of any case or offender characteristic should be relatively modest. However, as the overall sentence length increases, the effects of the independent variables also change—sometimes decreasing, but generally increasing in absolute magnitude. Again, this is consistent with our expectations: As the range of possible punishments increases, it opens up the possibility for all case and offender characteristics to have a greater effect on sentence length.

One of the more interesting patterns to emerge from the analyses on sentence length (measured as months) and reported in Tables 1 and 2 is the comparability of the OLS coefficients to the quantile regression coefficients. For several of the key independent variables—offense severity, jury trial, bench trial—the OLS coefficients overlap with the quantile regression coefficients at higher segments in the sentence length distribution, and often for a small range of the distribution (see Fig. 8). What this indicates is the cases at the upper end of the sentence length distribution—those with the longest sentences—have the greatest influence in determining the value of the OLS coefficients. In short, the highly skewed nature of the sentence length distribution gives greater weight to the more extereme values. One way of addressing the extreme values of sentence length is to use the logarithm (see, e.g., Fox 2008). Although the logged sentence length distribution exhibited less skew (see Figs. 5, 7), there was again a pattern of results in Tables 3 and 4 that showed the OLS coefficients to be significantly different from most of the quantile regression coefficients. The pattern of differentiation for the OLS and quantile regresison coefficients was more variable for the log-sentence length results. Nevertheless, it mattered little of whether sentence length or log-sentence length was used as the dependent variable, the results were substantially the same: the effects of the case and offender characteristics varied across the distribution of sentence length in ways that cast doubt on the OLS results.

One of the key benefits to using distributional analyses, such as quantile regression techniques, is they are particularly effective at testing for variations in the effects of the independent variables across the distribution of a dependent variable. In the larger context of research on disparities in sentencing due to characteristics such as race-ethnicity and gender, quantile regression models would appear to offer an interesting way of teasing out the effect of each of these characteristics at different quantiles of the distribution of sentence lengths. For example, at lower levels of the distribution, we may not expect to find dramatic differences in sentence lengths by race-ethnicity or gender. In part, this is due to legal constraints on the absolute length of a sentence that can be given to a convicted offender, meaning the range and distribution of possible sentence lengths is constrained. Yet, within this relatively narrow range of sentence lengths, it may be possible to note small differences in sentence length, even after statistically controlling for a variety of other legal characteristics of the offender and the case. More importantly, at the higher levels of the distribution, the range of possible sentence lengths is greatly expanded, creating enhanced opportunities for judges to exercise discretion in their sentencing decisions, which in turn should be reflected in a greater impact of offender and case characteristics. In contrast, the “liberation hypothesis” implies that judges have greater discretion among those offenders who receive the least severe rather than the most severe forms of punishment or supervision (Spohn and Cederblom 1991). OLS and other linear models assuming a fixed effect of the independent variables across the distribution of the dependent variable are unable to detect these kinds of hypothesized changes in the coefficients, while quantile regression models would seem to be well suited to testing these kinds of hypotheses.

Some of the earliest work on the development of quantile regression models was focused on dealing with censored data, where there were many zero values for the dependent variable (Powell 1984, 1986). Quantile regression models represented a way of estimating unbiased effects of the independent variables on the dependent variable, where the influence of the zeroes would be minimal. Practically, the addition of the zero values for the dependent variable simply shifts the non-zero values of the distribution to the right. The ability to test for variation in the quantile regression coefficients is unaffected by the additional zero values on the dependent variable. More recently, Koenker (2005, 2008) has developed the statistical theory and implemented a censored quantile regression procedure that will, among other options, account for zero values on the dependent variable. In light of the ongoing debate in the sentencing literature about how to statistically model the incarceration and the sentence length decisions simultaneously (Bushway et al. 2007), censored quantile regression analysis offers a way to incorporate all sentencing decisions in a jurisdiction into an analysis in a theoretically and empirically justifiable way.

There are at least four other important benefits to the use of distributional analyses that I think are important to highlight. First, by making no assumptions about the distribution of the error terms in the model, quantile regression techniques effectively deal with the non-normal distribution of errors that is inherently built into a sentencing guidelines grid. Second, while I have focused the discussion here on sentencing within a guidelines system, many of the same observations apply to jurisdictions without sentencing guidelines. In all jurisdictions in the United States, sentence length decisions will have a high degree of positive skew—there will be many offenders who receive sentences of a relatively short duration and very few who receive sentences as long as 10 or 20 years (see, e.g., Durose and Langan 2007). A related example of this kind of approach is provided by Rossi and Berk's (1997) study of public views about appropriate punishments (length of sentence) for offenders convicted in federal courts, who were subject to the US federal sentencing guidelines. In large part, their analyses relied on median regression (i.e., $\tau = .50$), recognizing that individual estimates of sentence lengths would be highly skewed and an analysis using OLS would not provide the best estimates in their multivariate models. Third, beyond the issue of sentence length decisions, there are a number of other issues in criminology and criminal justice that rely on the analysis of highly skewed distributions (e.g., bail amount decisions and self-reported delinquent acts). Similar to the research on sentencing decisions, research on these other topics has either transformed the variable (e.g., a logarithm) to reduce the degree of skew or recoded the variable in such a way that the fundamental distribution of the variable is changed (e.g., recoding self-reported number of delinquent acts from a count to something that ranges from 0 to 10 or more). Fourth, the ability of quantile regression models to estimate variable effects of independent variables across the distribution of a dependent variable suggests that this technique would be particularly valuable for testing theories that either directly or indirectly imply non-constant relationships. Beyond criminal justice decision-making research, many theories of crime and delinquency also imply that the effect of a key theoretical variable (e.g., control balance, self-control, peers) may have effects on crime and delinquency that vary by the frequency of criminal and delinquent behavior.

The primary limitation to the application of quantile regression is the need for a sufficiently large number of cases that will allow for the computation of quantiles for many different combinations of the values of the independent variables. However, since many researchers analyzing criminal justice decisions typically rely on samples (or populations) of several thousand cases, this requirement should not prove to be prohibitive. Other researchers may wonder about the prospects for multi-level quantile regression analyses. At present, there is not yet the ability to perform contextual or multi-level analyses. Much of the theoretical development of quantile regression models has occurred in the last 10 years and will require a substantial effort to link these efforts to the body of work on multi-level modeling. Nevertheless, these models offer a potentially valuable way for researchers to understand more fully the relationships between their independent and dependent variables independent of a multi-level analysis.

References

Buchinsky M (1994) Changes in the U.S. wage structure 1963–1987: application of quantile regression. Econometrica 62:405–458

Buchinsky M (1998) The dynamics of changes in the female wage distribution in the USA: a quantile regression approach. J Appl Econom 13:1–30

Bushway SD, Piehl AM (2001) Judging judicial discretion: legal factors and racial discrimination in sentencing. Law Soc Rev 35:733–764
Bushway S, Johnson BD, Slocum LA (2007) Is the magic still there? The use of the Heckman two-step correction for selection bias in criminology. J Quant Criminol 23:151–178
Conley T, Galenson D (1998) Nativity and wealth in mid-nineteenth-century cities. J Econ History 58: 468–493
Durose MR, Langan PA (2007) Felony sentences in state courts, 2004. USGPO, Washington, DC
Eide ER, Showalter MH (1999) Factors affecting the transmission of earnings across generations: a quantile regression approach. J Hum Resour 34:253–267
Eide ER, Showalter MH, Sims DP (2002) The effects of secondary school quality on the distribution of earnings. Contemp Econ Policy 20:160–170
Engen RL, Gainey RR (2000a) Modeling the effects of legally relevant and extralegal factors under sentencing guidelines: the rules have changed. Criminology 38:1207–1230
Engen RL, Gainey RR (2000b) Conceptualizing legally relevant factors under guidelines: a reply to Ulmer. Criminology 38:1245–1252
Fox J (2008) Applied regression analysis and generalized linear models, 2nd edn. Sage, Los Angeles
Gilchrist WG (2000) Statistical modelling with quantile functions. Chapman and Hall, Boca Raton, FL
Handcock MS, Morris M (1999) Relative distribution methods in the social sciences. Springer, New York
Hao L, Naiman DQ (2007) Quantile regression. Sage, Thousand Oaks, CA
Holleran D, Spohn C (2004) On the use of the total incarceration variable in sentencing research. Criminology 42:211–240
Johnson BD (2005) Contextual disparities in guidelines departures: courtroom social contexts, guidelines compliance, and extralegal disparities in criminal sentencing. Criminology 43:761–796
Johnson BD (2006) The multilevel context of criminal sentencing: integrating judge- and county-level influences. Criminology 44:259–298
Koenker R (2005) Quantile regression. Cambridge University Press, New York
Koenker R (2008) Quantile regression in R: a vignette. http://www.econ.uiuc.edu/~roger/research/rq/vig.pdf
Koenker R, Bassett G (1978) Regression quantiles. Econometrica 46:33–50
Koenker R, Bassett G (1982) Robust tests for heteroskedasticity based on regression quantiles. Econometrica 50:43–61
Koenker R, Hallock KF (2001) Quantile regression. J Econ Perspect 15:143–156
Koenker R, Xiao Z (2002a) Inference on the quantile regression process. Econometrica 70:1583–1612
Koenker R, Xiao Z (2002b) Inference on the quantile regression process: electronic appendix. http://www.econ.uiuc.edu/~roger/research/inference
McCullagh P, Nelder J (1989) Generalized linear models, 2nd edn. Chapman and Hall, New York
Mello M, Perrelli R (2003) Growth equations: a quantile regression exploration. Quart Rev Econ Financ 43:643–667
Miethe TD, Moore CA (1986) Racial differences in criminal processing: the consequences of model selection on conclusions about differential treatment. Sociol Quart 27:217–237
Miller PW (2005) The role of gender among low-paid and high-paid workers. Aust Econ Rev 38:405–417
Mora MT, Davila A (2006) Hispanic ethnicity, gender, and the change in the LEP-earnings penalty in the united states during the 1990s. Soc Sci Quart 87:1295–1318
Mustard DB (2001) Racial, ethic, and gender disparities in sentencing: evidence from the U.S. federal courts. J Law Econ XLIV:285–314
Ostrom BJ, Ostrom CW Jr, Hanson RA, Kleiman M (2007) Trial courts as organizations. Temple University Press, Phildelphia, PA
Powell JL (1984) Least absolute deviation estimation for the censored regression model. J Econom 25:303–325
Powell JL (1986) Censored regression quantiles. J Econom 32:143–155
R Development Core Team (2003) R: a language and environment for statistical computing. http://www.R-project.org
Rossi PH, Berk RA (1997) Just punishments: federal guidelines and public views compared. De Gruyter, New York
Spohn C, Cederblom J (1991) Race and disparities in sentencing: a test of the liberation hypothesis. Justice Quart 8:305–327
Stata Corporation (2007) Base reference manual. Stata Press, College Station, TX
Steen S, Engen RL, Gainey RR (2005) Images of danger and culpability: racial stereotyping, case processing, and criminal sentencing. Criminology 43:435–468
Tonry M (1996) Malign neglect: race, crime, and punishment in America. Oxford University Press, New York
Tukey JW (1977) Exploratory data analysis. Addison-Wesley, Reading, MA

Ulmer JT (1997) Social worlds of sentencing: court communities under sentencing guidelines. SUNY Press, Albany, NY

Ulmer JT (2000) The rules have changed—so proceed with caution: a comment on Engen and Gainey's method for modeling sentencing outcomes under guidelines. Criminology 38:1231–1244

Ulmer JT, Bradley MS (2006) Variation in trial penalties among serious violent offenses. Criminology 44:631–670

Ulmer JT, Johnson B (2004) Sentencing in context: a multilevel analysis. Criminology 42:137–177

Ulmer JT, Kurlychek MC, Kramer JH (2007) Prosecutorial discretion and the imposition of mandatory minimum sentences. J Res Crime Delinq 44:427–458

Walker TR (2000) Economic opportunity on the urban frontier: wealth and nativity in early San Francisco. Explor Econ Hist 37:258–277

Wheeler S, Weisburd D, Bode N (1982) Sentencing the white-collar offender: rhetoric and reality. Am Sociol Rev 47:641–659