Tales from the Tails: Examining the Effect of Inequality at the Extremes of the Mortality Distribution

Carla Shoff
Department of Agricultural Economics and Rural Sociology
and The Population Research Institute
The Pennsylvania State University
13 Armsby Building
University Park, PA 16802-6211 U.S.A.
cms534@psu.edu

Vivian Yi-Ju Chen
Department of Statistics
Tamkang University
151 Ying-Chuan Road
Tamsui, Taipei County 251 Taiwan
viviyjchen@stat.tku.edu.tw

Tse-Chuan Yang
Social Science Research Institute
The Pennsylvania State University
803 Oswald Tower
University Park, PA 16802-6211 U.S.A.
tuy111@psu.edu

Stephen A. Matthews
Department of Sociology, Department of Anthropology,
and the Population Research Institute
The Pennsylvania State University
University Park, PA 16802-6211 U.S.A.
matthews@pop.psu.edu

Draft: Full paper will be completed by the end of the year.

Abstract

Quantile Regression (QR) is increasingly used in economics, but uptake by demographers, including mortality researchers, is limited. A goal of this paper is to provide an empirical application of QR to mortality research, specifically exploring county-level associations between inequality and mortality in the US. The inequality/mortality association is well documented, but QR is appropriate when the research question is whether inequality has a greater influence in the counties with high mortality compared to those with lower mortality. QR reveals the associations between predictors of mortality vary across counties depending on where they are located in the mortality distribution and a non-linear relationship between inequality and mortality; underscoring the fact that differentials associated with inequality are more important at upper quantiles than implied by OLS-based findings. This has implications for public policy designed to reduce health disparities including the need for targeting not one model fits all (one tale fits all).

Introduction

The goal of this paper is to explore county-level associations between inequality and mortality in the US using quantile regression (QR). The association between inequality and mortality is well documented, but the QR framework is appropriate when the question of interest is whether inequality has a greater influence in the counties with high mortality rates compared to those with lower mortality rates.

Since being introduced in 1978 by Koenker and Bassett, Quantile Regression (QR) is a technique increasingly used in financial and economic research, medicine, and ecology. However, its uptake by demographers in general and mortality researchers specifically, has been limited. QR is not yet an established component of the quantitative training of demographers, but we believe it has broad applicability in population science. A goal of this paper is to provide an empirical application of QR applied to mortality research, and as such, help raise awareness of QR among population scientists.

Data

Mortality

County mortality rates are derived from the *Compressed Mortality Files* (CMF), 1989-1998 and 1999-2003, from the National Center for Health Statistics (NCHS) to calculate five-year (1998-2002) mortality rates (NCHS 2003; NCHS 2006) standardized with 2000 US age-sex population structure. We keep the rate unstandardized by race, but control for race/ethnic variables as a separate category.

Inequality

The Gini index is used to measure the distribution of household income in a county. The Gini coefficient is defined as a ratio with values between 0 and 1. The numerator is the area between the Lorenz Curve of the household income distribution and the uniform distribution line, and the denominator is the area under the uniform distribution line. In other words, a smaller Gini coefficient indicates a more even distribution of household income. The Gini index is calculated by multiplying the Gini coefficient by 100, and hence the value spans from 0 indicating perfect income equity, to 100 indicating extreme income inequality.

Race/ethnicity

Race/ethnic groups are included in the analysis. The percent of the county population that is Hispanic, the percent black, and the percent other race are used in the analysis. While Latinos and African Americans are known to be deprived relative to whites, prevailing literature suggests that Hispanics have lower mortality rates than whites, while blacks have higher mortality rates.

Rurality

In this study, rurality is measured by six variables derived from the *2000 Census of Population and Housing SF3*. Factor analysis indicated that the six variables could be summarized into three dimensions of residence: industrial structure, denseness, and exogenous economic integration (EEI). We calculated the factor scores with the regression method and used them as indicators of rural/urban residence.

The first dimension, industrial structure, comprises only one variable: percentage of the population ages 16 and over employed in farming, forestry, and fishing (factor loading is .934). The second dimension, denseness, consists of three variables related to the total population of a county: population density, which is the total population divided by land area (factor loading is .931), road density, which is the length of major roads per squared kilometer (factor loading is .800), and percentage of workers commuting by public transportation (factor loading is .947).

Higher scores reflect greater population density. The third characteristic of rurality is exogenous economic integration (EEI), which indicates the economic influence by neighboring metropolitan areas. Two variables are identified to capture the idea: percentage of workers traveling over an hour to work (factor loading is .866), and percentage of workers who work outside their county of residence (factor loading is .821). The more integrated county is expected to have a higher score and would be more economically dependent on the adjacent counties.

SES

We begin to describe the social structure of a county with social affluence and concentrated disadvantage. The former comprises the following variables: log of per capita income, percentage of the population ages 25 and over with a bachelor degree or higher, and percentage of population ages 16 and over employed in professional, administrative, and managerial positions, and the percentage of families with incomes over 75,000 dollars. A principal component factor analysis was used to reduce the variables and account for multicollinearity.

In contrast to social affluence, concentrated disadvantage consists of the subsequent covariates: poverty rate, percentage of persons receiving public assistance, unemployment rate, and percentage of female-headed households with children. They are considered as one indicator of concentrated disadvantage because the principal factor analysis indicates that 72 percent of the variance is shared by these variables.

Social Capital

We draw on recent endeavors by Rupasingha et al. (2006), who have developed a social capital index for US counties that pulls together a number of widely recognized indicators of this concept. Along with the social capital index, we use two additional measures of social capital:

safety and residential stability. Safety is a factor score based on the incidence of a variety of crimes, and is used to reflect the absence of mutual trust and the sense of safety (and thus weaker social capital). To reduce random variation, five-year average rates are calculated for 1998-2002 from the FBI's *Uniform Crime Reports*. Since this concept is measured in the inverse, it is expected to have a positive effect on mortality. Finally, social capital is higher among homeowners, implying that a stable environment is good for residents' interaction and facilitates the development of social capital. Hence, we include a residential stability index that is created by combining the percentage of the population living at the same address in 1995, the percentage of owner-occupied housing units, and the percentage of people living in mobile homes, respectively, and then averaging the three z-scores. The *2000 Census of Population and Housing SF3 Files* enables the calculation of residential stability.

QR vs OLS Regression

Standard OLS techniques concentrate on estimating the mean of the dependent variable subject to the values of the independent variables. Usually, variables are included as non-centered regressors. Quantile regression allows us to center the regressor around different quantiles (for example, regressors are centered around the median at the 0.5 quantile). This adds value to estimation results, especially because the distribution of mortality across counties is likely to be skewed. Given a set of explanatory variables, quantile regression estimates mortality rates conditional on the selected quantile. For example, it allows us to evaluate whether predictors are significant when we examine observations centered around percentiles in the lower and upper tails of the mortality distribution. The resulting coefficients give an estimate of the impact on counties with relatively low/high mortality rates values. By estimating the model at different quantiles, one can trace the entire conditional distribution of mortality rates given a set of predictors. An advantage of employing a QR estimation method is that the regression coefficient vector is not sensitive to outlying values of mortality rates, as the QR objective function is a weighted sum of absolute deviations.

Descriptive Analysis Results

The descriptive statistics of the variables used in this study are shown in Table 1. The average age-sex standardized mortality rate was 8.90 per 1,000 population in US counties. While the mortality variation was not large (standard deviation was only 1.38), the range was wide. Moreover, according to the US Census Bureau (2001), the overall Gini index at the household level in 2000 was 0.46, which was very close to our county-level estimate of 0.43. Although the maximal inequality was 0.61, the small value of the standard deviation (0.04) clearly suggested that most counties had a Gini coefficient around the mean. In the year 2000, on average, the counties' population was 8.72 percent Black and 6.16 percent Hispanic. However, as documented elsewhere i.e. (Jensen and Yang 2009), the geographical distribution of minorities, especially those of Hispanic-origin, varies greatly in the US, as illustrated by the large standard deviations of the race/ethnicity variables in Table 1. Since other socioeconomic and social capital variables were derived from factor analysis, they have a mean close to 0 and a variance of 1.

Quantile Regression Results

We ran five models to explore how the association between inequality and mortality varies with the inclusion of nested sets of explanatory variables and the results are displayed in Figure 1. In the figures, the x-axis represented the percentile of mortality and the y-axis is the magnitude of the association with inequality. The shaded gray area is the 95 percent confidence interval of the estimate at each percentile. The horizontal long dashed line represented the estimated *mean* association derived from an OLS regression, and the shorter dashed line indicated the 95 percent confidence interval of those estimates.

Without controlling for any other covariates (Model I), the association between inequality and mortality increased, and quite dramatically so across the mortality distribution; although the marginal effect of inequality declined. Mortality varied from approximately 1.5 deaths per 1,000 population in the lower tail to 15+ deaths per 1,000 at the high end (see Figure 1). The inclusion of race/ethnicity variables in Model II reduced the absolute level of the association with inequality but the increasing trend from low to high mortality remained. In Model III, rurality was incorporated, but does not greatly alter this pattern found in Model II.

The socioeconomic variables in Model IV reduced the association of inequality on mortality by half, especially for those counties with a mortality rate greater than 9.95 (the 80th percentile). The downward trend between inequality and mortality after the 80th percentile is interesting because it shows how the strong association between inequality and mortality in the high mortality counties are partly the result of their socioeconomic conditions. Explicitly, after controlling for SES, the relationship between inequality and mortality was no longer linear. In addition, the association for inequality was not significant for the lower tail of the distribution, (i.e. for those counties with a mortality rate below 7.36 deaths per 1,000 population; the 10th percentile), inequality was not a significant predictor of mortality.

Model V adds to the mix social capital variables and these further reduced the magnitude of the association for inequality (Model IV compared to Model V). The insignificant range at the low end of the distribution expanded to approximately the 15th percentile (7.57 deaths per 1,000 population). This may not sound like a big change, but what this translates into is that income inequality was not significantly associated with mortality for over 450 counties in the US. In Model V, the declining trend in the upper tail remained, but the slope became smoother compared to Model IV. The final QR graph in Figure 1 clearly offered an example of how OLS estimates can be misleading. Indeed, the quantile regression method suggested a potential curvilinear association of inequality on mortality, an association that has not drawn much attention in the literature.

Overall, income inequality was positively associated with mortality. Our QR analysis provides a new insight into how inequality is associated with mortality across US counties. After controlling for the other independent variables, income inequality had a detrimental effect on mortality for those counties with the highest mortality rates (i.e. higher than 7.6 deaths per 1,000 population; the 15th percentile). In the lowest 15 percent of the distribution, there seemed to be no significant association with inequality. The strongest association could be found for those counties with a mortality rate of about 9.95 deaths per 1,000 population (the 80th percentile), above which the effect of inequality on mortality decreased.

After identifying the potential curvilinear association between inequality and mortality, we focus our discussion on the findings in Model V. In Table 2 we list QR parameter estimates at selected percentiles. For the purpose of brevity, we reported the results (with the exception of inequality) using graphs (Figure 2). Several interesting findings emerge from our analysis.

First, the intercept of Model V may be interpreted as the estimated conditional quantile function of the mortality distribution of a county with all the explanatory variables set at their mean values. Note that except for the Gini and race/ethnicity covariates, all the other variables are centered, which is the preferable treatment for the independent variables (Koenker and

Hallock 2001). Compared to the OLS estimates, the intercepts from the quantile regression were relatively low and demonstrated a rapidly increasing trend after the 80th percentile.

Second, with respect to the race/ethnicity variables, the relationship for percentage Black was relatively stable across the distribution, and quite consistent with OLS results. As has been suggested in the literature, counties with a higher percent Black population tend to have higher mortality rates than those counties with a low percent. Note that the confidence intervals of the quantile regression estimates overlapped with the OLS estimates, indicating that percent Black exerted a pure association with the conditional mortality distribution. While Latinos and African Americans are sometimes seen to have comparable socioeconomic statuses, published research suggests that being Hispanic is negatively associated with mortality, which is known as the Hispanic Paradox (Abraido-Lanza, Dohrenwend et al. 1999). Our results not only confirmed this finding, but also clearly demonstrate that the association of percent Hispanic on mortality was stronger at the lower quintile. Specifically, the association of percent Hispanic with mortality almost doubled for those counties with a mortality rate below 7.81 deaths per 1,000 population (the 20th percentile) in contrast to those above the threshold.

Third, we found no significant effect of denseness on mortality. However, EEI was positively associated with mortality, but this relationship was only significant for those counties above the threshold of 7.57 deaths per 1,000 population (the 15th percentile). In the upper 15 percentile, the association of EEI with mortality increased rapidly. That is, being more economically integrated with neighboring counties was associated with higher mortality. In addition, primary industries were negatively associated with mortality. The effect of primary industries on mortality was stronger in the lower half of the distribution and declined through to the 90th percentile (10.56 deaths/1,000 population), above which it was insignificant. Our findings here corresponded to several recent studies (McLaughlin, Shannon Stokes et al. 2001; McLaughlin, Stokes et al. 2007).

The socioeconomic variables, affluence and disadvantage, exerted the expected effects on mortality. Their point estimates at different percentiles were significant across the mortality distribution and were quite close to the OLS estimates, with very little variation. One conspicuous trend shared by affluence and disadvantage was that their impacts on mortality changed greatly between the 80th and the 90th percentiles. In general, a county with superior socioeconomic composition was related to lower mortality after controlling for other factors, which in part supports the argument that social conditions are fundamental determinants of health (Link and Phelan 1995).

Finally, as expected, residential stability, safety, and the social capital index were found to be significant predictors of county mortality rates. Specifically, a unit increase in residential stability could reduce mortality by approximately 21 deaths per 100,000 population in a county. For those counties above the 70th percentile, the decrease could be as large as 40 deaths. In addition, the association of safety with mortality was relatively uniform across the distribution, or about 10 to 15 deaths per 100,000 population. The social capital index demonstrated a declining trend (from -0.2 to -0.05) in county-level mortality. The counties at the lower half of the mortality distribution benefitted from the social capital index more than those counties at the upper half of the mortality distribution.

Conclusion

In sum, by using the quantile regression method, our analyses revealed that the association between inequality and morality was curvilinear. The strongest impact was found around the 80th percentile, above which the effect decreased. In addition, the association of inequality on mortality in the lower 15 percent of the counties was not significant after controlling for the other predictors in the model. Only a few of the predictors demonstrated a pure association on mortality, including percent Black and safety. A conventional OLS modeling and mean regression approach would fail to provide these insights into how these predictors influenced mortality.

Our research is substantively driven by concerns in health inequality and social stratification. In this paper we demonstrate how an emergent technique—quantile regression—can help demographers interested in the study of disparities in mortality. In our empirical example, the QR approach reveals that association between predictors of mortality varies across counties depending on where they are located in the mortality rate distribution. Indeed, our results identify a curvilinear relationship between inequality and mortality and thus underscore the fact that differentials associated with inequality are even more important at upper quantiles than implied by prior OLS-based findings. This has implications for public policy designed to reduce social inequality and the need for nuanced targeting rather than one model fits all (or one tale fits all).

Works Cited

Abraido-Lanza, A. F., B. P. Dohrenwend, et al. (1999). "The Latino mortality paradox: A test of the "salmon bias" and healthy migrant hypotheses." American Journal of Public Health **89**: 1543-1548.

Jensen, L. and T.-C. Yang, Eds. (2009). Taken by surprise: New immigrants in the rural United States. International Migratino and Rural Areas-Global Perspectives. Hampshire, UK, Ashgate Publishing, Ltd.

Koenker, R. and K. F. Hallock (2001). "Quantile Regression." Journal of Economic Perspectives **15**(4): 143-156.

Link, B. G. and J. Phelan (1995). "Social conditions as fundamental causes of disease." Journal of Health and Social Behavior **35(**Extra Issue: Forty Years of Medical Sociology: The State of the Art and Directions for the Future (1995)): 80-94.

McLaughlin, D. K., C. Shannon Stokes, et al. (2001). "Residence and Income Inequality: Effects on Mortality Among U.S. Counties." Rural Sociology **66**(4): 579-598.

McLaughlin, D. K., C. S. Stokes, et al. (2007). Differential Mortality Across the United States: The Influence of Place-Based Inequality. The Sociology of Spatial Inequality. L. M. Lobao, G. Hooks and A. R. Tickamyer. Albany, State University of New York Press.

NCHS (2003). "Compressed Mortality File, 1989-98 (machine readable data file and documentation, CD-ROM series 20, No. F)." National Center for Health Statistics: Hyattsville, Maryland.

NCHS (2006). "Compressed Mortality File, 1999-2002 (machine readable data file and documentation, CD-ROM series 20, No.F." National Center for Health Statistics: Hyattsville, Maryland.

Rupasingha, A., S. J. Goetz, et al. (2006). "The Production of Social Capital in US Counties." The Journal of Socio-Economics **35**: 83-101.

US Census Bureau (2001). Money Income in the United States: 2000. I. S. Branch**:** 1-26.

Table 1. Descriptive Statistics of Mortality and Independent Variables (N=3,109)

Variable	Mean	Median	Maximum	Minimum	Std Dev
Dependent Variable					
Mortality	8.898	8.848	19.777	0.000	1.377
Inequality					
Gini	0.434	0.432	0.605	0.314	0.038
Race/Ethnicity					
Black	8.715	1.638	86.078	0.000	14.504
Other Races	3.494	1.863	93.583	0.000	6.757
Hispanic	6.156	1.753	98.104	0.000	12.116
Rurality					
Denseness	0.000	-0.171	28.704	-0.605	1.000
EEI	0.000	-0.115	4.535	-1.918	1.000
Primary Industries	0.000	-0.200	8.652	-2.891	1.000
SES					
Affluence	0.000	-0.185	6.011	-2.428	1.000
Disadvantage	0.000	-0.183	9.056	-2.536	1.000
Social Capital					
Residential Stability	0.000	0.049	1.701	-4.152	0.589
Safety	0.000	-0.201	12.119	-1.370	1.000
Social Capital Index	0.003	-0.120	7.656	-4.063	1.294

Table 2. Quantile Regression Results Predicting Mortality (N=3109)

	Quantile					
	0.20	0.40	0.50	0.60	0.80	0.90
Inequality						
Gini	**2.58****	**3.65****	**3.99****	**4.18****	**5.15****	**4.49****
	(1.03,4.13)	(2.52,4.78)	(2.7,5.29)	(2.75,5.61)	(3.62,6.68)	(2.75,6.23)
Race/Ethnicity						
Black	**0.01****	**0.01****	**0.01****	**0.01****	**0.01****	**0.01****
	(0.00,0.01)	(0.01,0.01)	(0.01,0.02)	(0.01,0.01)	(0.01,0.02)	(0.00,0.02)
Other Races	**-0.01**	**0.00**	**0.00**	**0.01**	**0.02**	**0.01**
	(-0.02,0.01)	(-0.01,0.01)	(-0.01,0.01)	(0.00,0.02)	(0.01,0.03)	(-0.01,0.03)
Hispanic	**-0.02****	**-0.02****	**-0.02****	**-0.02****	**-0.02****	**-0.02****
	(-0.03,-0.02)	(-0.02,-0.02)	(-0.02,-0.01)	(-0.02,-0.01)	(-0.02,-0.01)	(-0.03,-0.01)
Rurality						
Sparseness	**0.04**	**0.05**	**0.04**	**0.03**	**0.00**	**0.03**
	(-0.03,0.10)	(-0.01,0.11)	(-0.01,0.10)	(-0.05,0.12)	(-0.11,0.10)	(-0.09,0.15)
Isolation	**0.13****	**0.20****	**0.20****	**0.19****	**0.18****	**0.22****
	(0.07,0.20)	(0.16,0.23)	(0.16,0.24)	(0.15,0.22)	(0.13,0.23)	(0.17,0.27)
SES						
Industrial	**-0.40****	**-0.33****	**-0.31****	**-0.26****	**-0.19****	**-0.07**
	(-0.48,-0.32)	(-0.38,-0.28)	(-0.35,-0.27)	(-0.31,-0.22)	(-0.25,-0.14)	(-0.14,0.01)
Affluence	**-0.40****	**-0.41****	**-0.43****	**-0.41****	**-0.43****	**-0.36****
	(-0.48,-0.32)	(-0.46,-0.36)	(-0.48,-0.38)	(-0.46,-0.35)	(-0.50,-0.37)	(-0.41,-0.31)
Disadvantage	**0.38****	**0.34****	**0.34****	**0.34****	**0.30****	**0.41****
	(0.29,0.47)	(0.26,0.42)	(0.25,0.42)	(0.24,0.43)	(0.21,0.39)	(0.31,0.51)
Social Capital						
Stability	**-0.15****	**-0.18****	**-0.17****	**-0.16****	**-0.21****	**-0.24****
	(-0.24,-0.06)	(-0.24,-0.12)	(-0.25,-0.09)	(-0.24,-0.08)	(-0.31,-0.10)	(-0.36,-0.11)
Safety	**0.13****	**0.15****	**0.15****	**0.15****	**0.13****	**0.14****
	(0.08,0.18)	(0.11,0.19)	(0.10,0.19)	(0.12,0.19)	(0.08,0.19)	(0.09,0.19)
Social Capital	**-0.17****	**-0.14****	**-0.12****	**-0.14****	**-0.14****	**-0.11****
Index	(-0.23,-0.11)	(-0.18,-0.11)	(-0.17,-0.07)	(-0.18,-0.11)	(-0.20,-0.08)	(-0.16,-0.07)

Note: *$p \leq 0.05$; **$p \leq 0.01$; The numbers in parentheses indicate the 95 percent confidence intervals.

Figure 1. Models I-V Exploring the Impact of Inequality on Mortality Using Nested Sets of Explanatory Variables

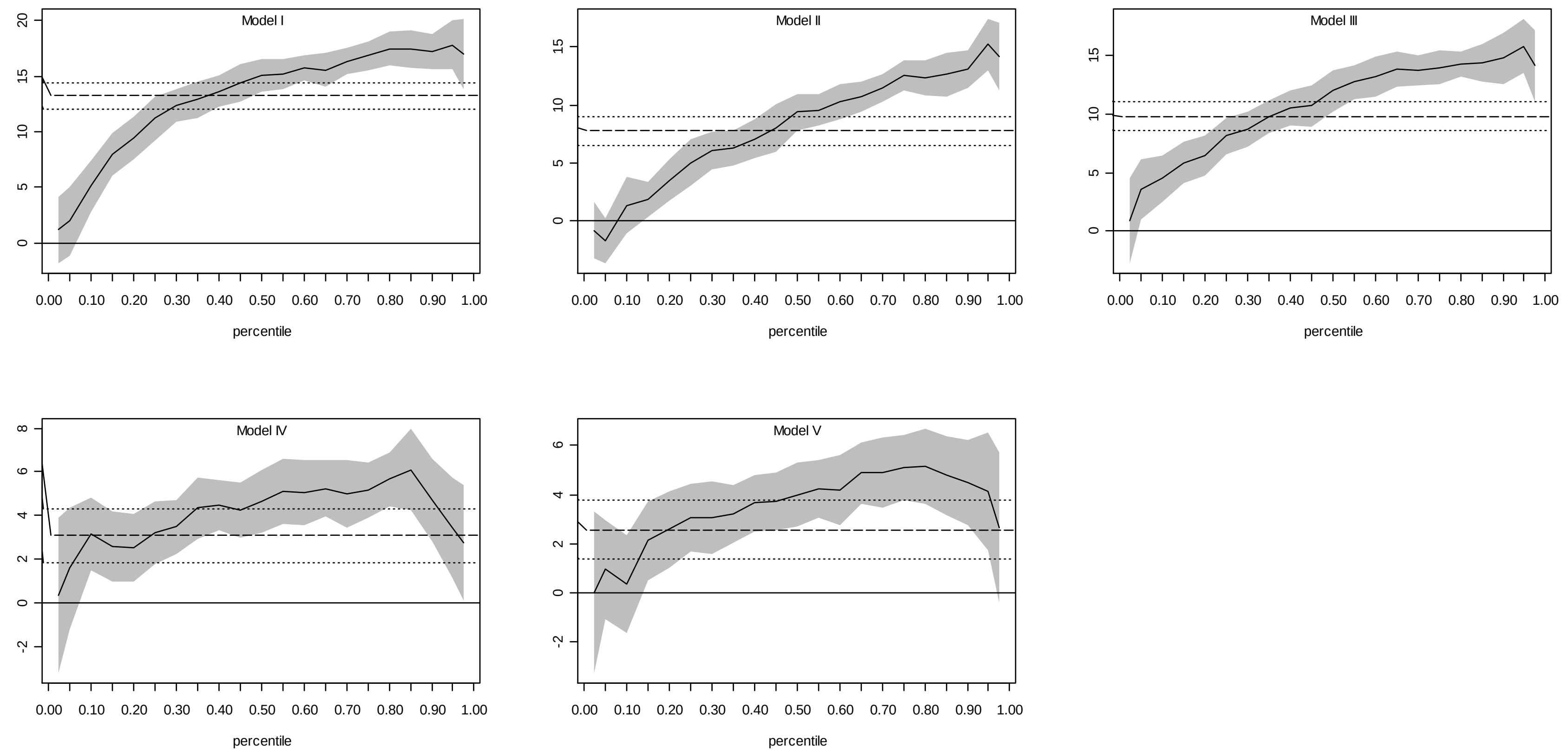

Figure 2. Model V for Each of the Explanatory Variables (Excluding Inequality)

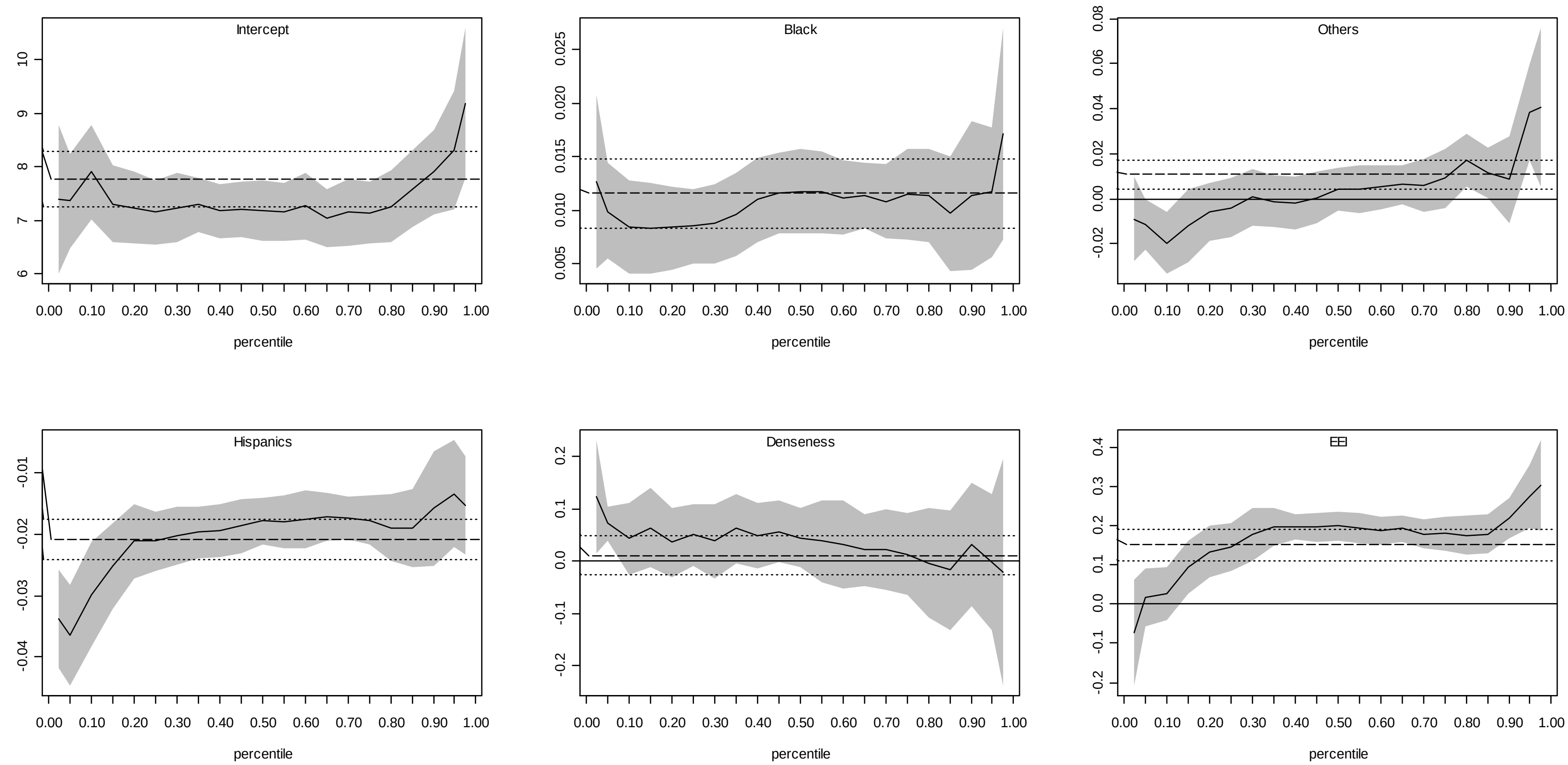

Figure 2 (Cont.). Model V for Each of the Explanatory Variables (Excluding Inequality)