Encyclopedia of Education Economics & Finance

Quantile Regression

Contributors: Thurston Domina & Marianne Bitler & Andrew Penner & Emily Penner
Editors: Dominic J. Brewer & Lawrence O. Picus
Book Title: Encyclopedia of Education Economics & Finance
Chapter Title: "Quantile Regression"
Pub. Date: 2014
Access Date: November 17, 2014
Publishing Company: SAGE Publications, Inc.
City: Thousand Oaks,
Print ISBN: 9781452281858
Online ISBN: 9781483346595
DOI: http://dx.doi.org/10.4135/9781483346595.n222
Print pages: 592-595

http://dx.doi.org/10.4135/9781483346595.n222

Many of the analytic techniques that are widely utilized in the economics of education focus on central tendencies. Traditional experimental analyses compare the mean on an outcome of interest for subjects randomly assigned to a treatment group with the mean for subjects randomly assigned to a control group. Regression analyses model the mean value of the outcome variable for fixed values of predictors. Quantile regression and other distributional estimators expand these mean-focused techniques to make it possible to analyze an outcome variable's entire distribution by estimating the relationship between one or more predictor variables and a specific quantile (i.e., percentile) of an outcome.

These approaches have wide applicability in discussions of educational policy and economics, where stakeholders are interested in pursuing both excellence and equity goals. Policymakers expect schools to improve the availability of human capital in a society and to narrow existing social and educational inequalities. However, these two goals are not necessarily mutually reinforcing. Some policies may boost average academic achievement even as they broaden educational inequalities. Others may depress academic achievement even as they narrow inequalities. With their analytic focus on central tendency, traditional econometric techniques may provide limited information about the distribution of educational achievement. Quantile treatment effects and quantile regression provide important opportunities to boost our understanding of educational inequality and the distributional consequences of educational policies and interventions. This entry provides a brief introduction to these distributional research methods and highlights several recent studies that productively used these methods in settings related to educational policy and the economics of education more broadly.

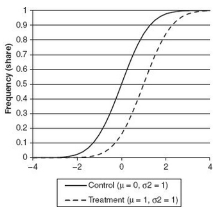
Conceptualizing and Measuring Consequences

To contextualize these distributional estimators, it is useful to first consider the estimation of the mean effects of a simple treatment in an experimental setting. Under the potential outcomes model, each individual has two potential outcomes: one that would occur if that individual were assigned to the treatment group and the other if that individual were assigned to the control group. The fundamental evaluation problem is that the same person cannot simultaneously be in the treatment group and the control group. Experimental research designs address this evaluation problem by randomly assigning subjects to treatment and control conditions. In this setting, evaluators can assume that the odds that a subject is exposed to the treatment are unrelated to any characteristic of the subject and that the mean effect of the treatment is thus the difference between the mean for subjects assigned to the treatment group and the mean for subjects assigned to the control group. Quantile treatment effect estimators expand the logic of these evaluation techniques to investigate the difference in the outcomes of interest in the treatment and control groups at any percentile (or quantile) of the distributions. Intuitively, **[p. 592 ↓]** these estimators can be thought of as comparing the *p*th percentile score of subjects in a treatment group with the *p*th percentile score of subjects in a control group. By compiling these quantile treatment effects estimates across the distribution of an outcome, quantile treatment effects estimators allow inferences regarding a treatment's consequences on the distribution of the outcome.

Mechanically, quantile treatment effect estimation hinges on the comparison of the cumulative distribution function (CDF) of the outcome of interest for treatment and control groups. For any random continuous variable *Z*, the CDF is the proportion of the population for which *Z* is less than or equal to each value *y*. Figure 1 provides CDF illustrations for two groups. The solid line in Figure 1 represents the CDF for a standard normal distribution, and the dashed line is the CDF for a normal of mean 1 but standard distribution of 1. The *y*-axis in this graph represents the cumulative frequencies in the population in each group, and the *x*-axis represents the range of values for the outcome of interest.

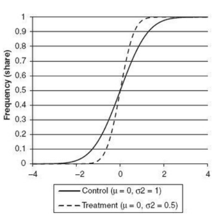
For purposes of illustration, assume that the solid line in Figure 1 represents the test score distribution for the students randomly assigned to the control group in a hypothetical experimental evaluation project, and the dashed line represents the test score distribution for students randomly assigned to the treatment group. To estimate the quantile treatment effect, one subtracts the *x*-value for treatment group at any given quantile from the *x*-value for the control group. In this hypothetical example, the treatment group score is one standard deviation above the control group's at every point in the test score cumulative distribution, so the treatment effect is uniform and equal to 1 everywhere.

Figure 1 Cumulative Distribution Function for Hypothetical Treatment With Uniform Positive Effect Across the Distribution



Such a uniform treatment effect need not necessarily occur. Indeed, the below discussion of applications of quantile treatment effect estimation in educational settings indicates that mean effect estimation may often obscure important variation in treatment effects across the test score distribution. Figure 2 provides a second hypothetical example to illustrate one way in which this could occur. The control group in this example again has a standard normal distribution on the outcome of interest. However, the test score distribution for the treatment group is far more compressed, with a median of 0 and a standard deviation of 0.5. This hypothetical treatment has no average effect on student achievement. However, the hypothetical treatment substantially reduces inequality in student achievement, with negative treatment effects below the median and positive treatment effects above the median.

Figure 2 Cumulative Distribution Function for Hypothetical Treatment With Positive Effect at Bottom and Negative Effect at Top of Distribution



A similar approach can be used to investigate the relationship between predictors and outcome distributions in observational settings, where the assumption of treatment exogeneity is not likely to hold. Nonrandom selection is a particularly vexing problem in educational settings, where educators **[p. 593 ↓]** may handpick students for interventions, or students or parents may self-select into programs or schools. Researchers often employ regression techniques to address this problem, arguing that the mean treatment effect is equal to the difference between students selected into treatment and control groups after controlling for relevant observed characteristics. Just as quantile treatment effect estimation extends traditional means comparisons to estimate the distributional effects of treatments, quantile regression expands the regression approach to estimate the relationship between one or more predictor variables and the distribution of the outcome, conditional on all other predictors. (Roger Koenker provides a more technical description of quantile regression estimation; Lingxin Hao and Daniel Q. Naiman provide a more detailed description of the technique geared to data analysis.)

Quantile regression is particularly useful in contexts in which ceilings, floors, or outliers on the outcome variable threaten to bias traditional regression estimates, since quantile regression estimates at the median and at other points near the center of the distribution are not affected by censoring or measurement error at the tails. These quantile regression models can be interpreted in much the same way as traditional linear regressions, with two important qualifications. First, in interpreting quantile regressions, the analyst should keep in mind that specific quantile regression estimates refer to a specific point on the distribution of the dependent variable. In many cases, it is most informative to compile quantile regression estimates from several points in order to understand the relationship between treatment and outcome across the distribution. Second, it is important to remember that quantile regression models provide estimates of the relationship between given independent variables and the distribution of the dependent variable, conditional on all other independent variables.

Many analysts are comfortable with conditional means in traditional regression settings, where the overall mean is equal to the weighted average of any number of subgroups. However, this is not the case for quantiles (e.g., the overall 10th percentile is not necessarily equal to weighted average of the subgroup's 10th percentiles). The conditional distribution that is the result of a quantile regression is thus different from the unconditional distribution (undoing the conditioning for all but the independent variable that is the focus of the analysis) that is typically of interest. Neglecting that difference can lead to misinterpretation. Joshua Angrist and Jorn-Steffen Pischke provide further discussion of this issue. Inference with quantile regression can be carried out in various ways; Maria Kocherginsky, Xuming He, and Ying Wei lay out guidelines for inference for different sample sizes.

Quantile Treatment Effects and Quantile Regression in Educational Settings

Recent studies utilize quantile treatment effect estimation and quantile regression to deepen our understanding of several educational settings by examining the distribution of educational achievement. For example, Erika Jackson and Marianne Page reevaluate data from the experiment known as Project STAR, or Student/Teacher Achievement Ratio, in which nearly 12,000 early elementary students in 79 Tennessee public elementary schools were randomly assigned to a small class (in which the target enrollment was 13–17 students), a regular-sized class (in which the target enrollment was 22–25 students), or a regular-sized class with a full-time teacher's aide (in which the target enrollment was 22–25 students). Earlier analyses of the STAR data consistently indicate that assignment to a small class improves student achievement test scores, with effect sizes for kindergarten and first graders of approximately 0.2 standard deviations. While Jackson and Page demonstrate that this significant positive effect held across the test score distribution, their quantile treatment effect analyses demonstrate that the effects of small class placement are considerably larger at the top of the test score distribution than at the bottom.

Recent distributional analyses of school accountability policies further demonstrate the potential that these techniques offer for educational researchers. There is strong evidence suggesting that the No Child Left Behind Act and other school accountability policies have small positive average effects on student achievement. But new distributional analyses indicate that these average effects tell only part of the story. Accountability policies give schools incentives to direct teacher attention and other educational resources at students whose scores are just below proficiency thresholds set by states under these policies. Research by Derek Neal and Diane Schanzenbach and by Randall Reback that estimates policy effects on the test score distribution shows that these policies have strong positive effects on students near the proficiency threshold and weaker **[p. 594 ↓]** effects on higher and lower achieving students. These effects might well have been missed had these researchers instead focused on estimation of average effects or subgroup effects.

Other applications of quantile regression provide more descriptive insights into the relationship between education and social inequality. Thomas Lemieux, for example, uses quantile regression to investigate changes in the returns to postsecondary education over time and the extent to which these changes explain recent increases in wage inequality in the United States. His analyses indicate that the returns to postsecondary education increased sharply between the early 1970s and the early 2000s, particularly at the top of the income distribution. Lemieux finds, for example, that relative wages for workers with postgraduate education increased by 39 percentage points at the median between the early 1970s and the early 2000s. At the 90th percentile, the relative wages for workers with postgraduate education increased by 51 percentage points during the same time period.

In other descriptive work, Andrew Penner uses quantile regression models to examine gender differences across the distribution of mathematics achievement in 22 countries. Gender differences vary across countries in both magnitude and shape: In approximately half of the countries, differences do not vary across the distribution, but in several countries, differences are more pronounced at the bottom, while in others, differences are largest at the top. Penner further highlights the relationship between gender inequality in the labor market and gender differences in mathematics achievement at the top of the distribution, arguing that social context shapes the pool of potential scientists.

These studies highlight the ways in which evaluations of average impact miss heterogeneous effects throughout the distribution.

See alsoEconometric Methods for Research in Education; Educational Equity; Policy Analysis in Education

Authors' Note: Research for this entry was supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of

Health under Award Number P01HD065704. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

ThurstonDomina MarianneBitler AndrewPenner EmilyPenner

http://dx.doi.org/10.4135/9781483346595.n222

Further Readings

Angrist, J. D., & Pischke, J. S. (2009). Quantile regression . In Mostly harmless econometrics (pp. 269–292). Princeton, NJ: Princeton University Press.

Hao, L., & Naiman, D. Q. (2007). Quantile regression . Thousand Oaks, CA: Sage.

Jackson, E., & Page, M. E. (2013). Estimating the distributional effects of education reforms: A look at Project STAR . Economics of Education Review , 32, 92–103.

Kocherginsky, M., He, X., & Mu, Y. (2005). Practical confidence intervals for regression quantiles . Journal of Computational and Graphical Statistics , 14(1), 41–55.

Koenker, R. (2007). Quantile regression . Cambridge, UK: Cambridge University Press.

Lemieux, T. (2006). Postsecondary education and increasing wage inequality . American Economic Review , 96(2), 195–199.

Neal, D., & Schanzenbach, D. W. (2010). Left behind by design: Proficiency counts and test-based accountability . Review of Economics and Statistics , 92(2), 263–283.

Penner, A. M. (2008). Gender differences in extreme mathematical achievement: An international perspective on biological and social factors . American Journal of Sociology , 114, S138–S170.

Reback, R. (2008). Teaching to the rating: School accountability and the distribution of student achievement . Journal of Public Economics , 92(5), 1394–1415.