Quantile Regression

Katrin Hohl, Methodology Institute, London School of Economics

"How does explanatory variable X affect dependent variable Y?" is a standard type of research question in quantitative criminology and the social sciences at large. If the dependent variable is continuous, linear regression is typically used to answer it.

Linear regression estimates the expected change of *the mean* of the dependent variable for every one-unit change in the explanatory variable. To give an example, on average the mean level of trust increases by 1.5 points (on a scale ranging from 10 to 50) for every one-point increase in the rating of how well the police are keeping the public informed (a 7-point likert scale). It does not estimate how the rest of the distribution of the dependent variable responds to the one-unit change in X. But what about the effect of police information provision on the 10 percent least trusting citizens, or the 10 percent with the greatest trust in the police? Does keeping the public informed enhance the perceived trustworthiness of the police in all members of the public by the same amount, e.g. by 1.5 points as predicted for the mean? Or are there any differences in the effect of police information provision on the least, 'average' and most trusting citizens?

Standard linear regression precludes the analysis of how parts of the distribution of Y, other than its mean, respond to changes in the explanatory variables. Depending on whether or not normally distributed error terms are assumed, changes in X are either assumed to have the same effect on all parts of the distribution of Y, or the regression model does not say anything about how the rest of the distribution of Y responds to changes in X. Yet, many research questions - like the example given above - require analysing the effect of X on some location on the distribution of Y other than its mean (i.e. some quantile), or on the entirety of the distribution of Y, including its tails, spread and skew. In such cases, quantile regression is a useful method of analysis.

Like standard linear regression, quantile regression is appropriate for continuous dependent variables. The explanatory variables can have any measurement level. The main difference between quantile regression and conventional regression is that instead of predicting the effect of a marginal (one-unit) change in the explanatory variable X on the *mean* of the dependent variable Y, quantile regression predicts the effect of a one-unit change in the explanatory variable on some researcher-defined *quantile* of the distribution of Y, for example, the 10^{th}quantile, the first quartile, the median or the 95^{th} quantile. Combining the results of a series of quantile regression analyses beginning in the left tail (e.g. the 5^{th} quantile) and predicting quantiles in regular intervals up to the 95^{th} quantile in the right tail of the distribution of the dependent variable allows describing the relationship between explanatory and dependent variable along the entire distribution of the dependent variable. This then allows detecting when some parts of the distribution respond to changes on the explanatory variable differently than others and determining the effect of the explanatory variables on the variance and skew of the dependent variable.

Whilst standard linear regression uses ordinary least squares (OLS) to estimate regression coefficients, quantile regression minimises the sum of absolute vertical distances weighted by the quantile, where data points above the fitted line are weighted by the quantile θ and data points below the line are weighted by (1-θ). The Stata commands ‘qreg’, ‘sqreg’ and ‘bsqreg’ produce basic quantile regression results. There exist a number of post-estimation commands to test the statistical significance of differences in regressions coefficients at different quantiles and to produce graphs that visualise the results.

Quantile regression might be considered in addition or instead of standard linear regression in the following situations:

- *Research questions that are about a quantile, not the mean:* Sometimes it is not the mean, but the extremes of the distribution that are of immediate concern. For example, the poorest 10 percent of the population. In this case, predicting the effect of the explanatory variables on the 10^{th} quantile answers the research question more directly than standard linear regression.
- *Outliers and skewed distributions:* The mean is often not a good measure of central tendency when the dependent variable is highly skewed or has influential outliers that cannot be removed from the dataset. In this case, the median is a better measure of central tendency. By extension, (quantile) regression on the median may provide a better summary of the relationship than (standard linear) regression on the mean.
- *Differences in the effect of the explanatory variable on the low, average and high end of the dependent variable:* Differences in the way different parts of the distribution respond to changes on the explanatory might be informative with regard to the research question and have important implications for policy making. For example, a policy intervention might enhance confidence in the police in those with average levels of trust in the police, but might be ineffective in the 25 percent least trusting citizens.
- *Effects on the distribution – the variance and skew - of the dependent variable:* The study of how a particular intervention or a set of explanatory variables impact on social- or any other form of inequality falls into this category.

This article is based on:

Hohl, K. *In preparation*. Enhancing confidence in the police: how information provision affects trust in the police amongst the least, ‘average’ and most police-trusting citizens.

Further reading

Hao, L. and Naiman, D. 2007. *Quantile Regression*. London: Sage.

Hohl, K. 2009. Beyond the average case: The mean focus fallacy of standard linear regression and the use of quantile regression for the social sciences. *Working paper. Available on SSRN:* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1434418

Koenker, R. 2005. *Quantile Regression*. Cambridge: Cambridge University Press.