Quantile Regression

Quantile Regression

Quantile regression examples

- Consumer economics: effects of household income on food expenditures for low- and high-expenditure households.
- Education: factors affecting student scores along their score distribution.

Quantile regression model explanations

- We use quantiles to describe the distribution of the dependent variable.
- Quantiles and percentiles are synonymous – the 0.99 quantile is the 99th percentile.
- The best-known quantile is the median. The median is the 0.50 quantile.
- The standard Ordinary Least Squares (OLS) models the relationship between one or more independent variables x and the conditional mean of a dependent variable y.
- A quantile regression models the relationship between x and the conditional quantiles of y rather than just the conditional mean of y.
- A quantile regression gives a more comprehensive picture of the effect of the independent variables on the dependent variable.
- The dependent variable is continuous with no zeros or too many repeated values.

Quantile regression model

- The quantile regression is described by the following equation:

$$y_i = x_i^{'}\beta_q + e_i$$

where β_q is the vector of unknown parameters associated with the q^{th} quantile.

- The OLS minimizes $\sum_i e_i^2$, the sum of squares of the model prediction error e_i
- The median regression, also called least absolute-deviation regression minimizes $\sum_i |e_i|$
- The quantile regression minimizes $\sum_i q|e_i| + \sum_i (1-q)|e_i|$, a sum that gives the asymmetric penalties $q|e_i|$ for underprediction and $(1-q)|e_i|$ for overprediction

The qth quantile regression estimator $\widehat{\beta_q}$ minimizes over β_q the objective function

$$Q(\beta_q) = \sum_{i:y_i \geq x_i^{'}\beta}^{N} q\left|y_i - x_i^{'}\beta_q\right| + \sum_{i:y_i < x_i^{'}\beta}^{N} (1-q)\left|y_i - x_i^{'}\beta_q\right|$$

where $0 < q < 1$.

- In contrast to OLS and maximum likelihood, the quantile regression computational implementation uses linear programming methods.
- We have β_q instead of β to make clear that different choices of q estimate different values of β.

Quantile regression coefficients and marginal effects

- The standard conditional quantile is specified to be linear:

$$Q_q(y_i|x_i) = x_i'\beta_q$$

- For the *j*th regressor, the marginal effect is the coefficient for the qth quantile.

$$\frac{\partial Q_q(y|x)}{\partial x_j} = \beta_{qj}$$

- A quantile regression parameter β_{qj} estimates the change in a specified quantile q of the dependent variable *y* produced by a one unit change in the independent variable x_j.
- The marginal effects are for infinitesimal changes in the regressor, assuming that the dependent variable remains in the same quantile.

- Unlike interpretations of OLS regression, the interpretations of quantile regression results need to specify which quantile of the dependent variable they refer to.
- Two types of significance are important for quantile regression coefficients.
    - Quantile coefficients can be significantly different from zero.
    - Quantile coefficients can be significantly different than the OLS coefficients, showing different effects along the distribution of the dependent variable.

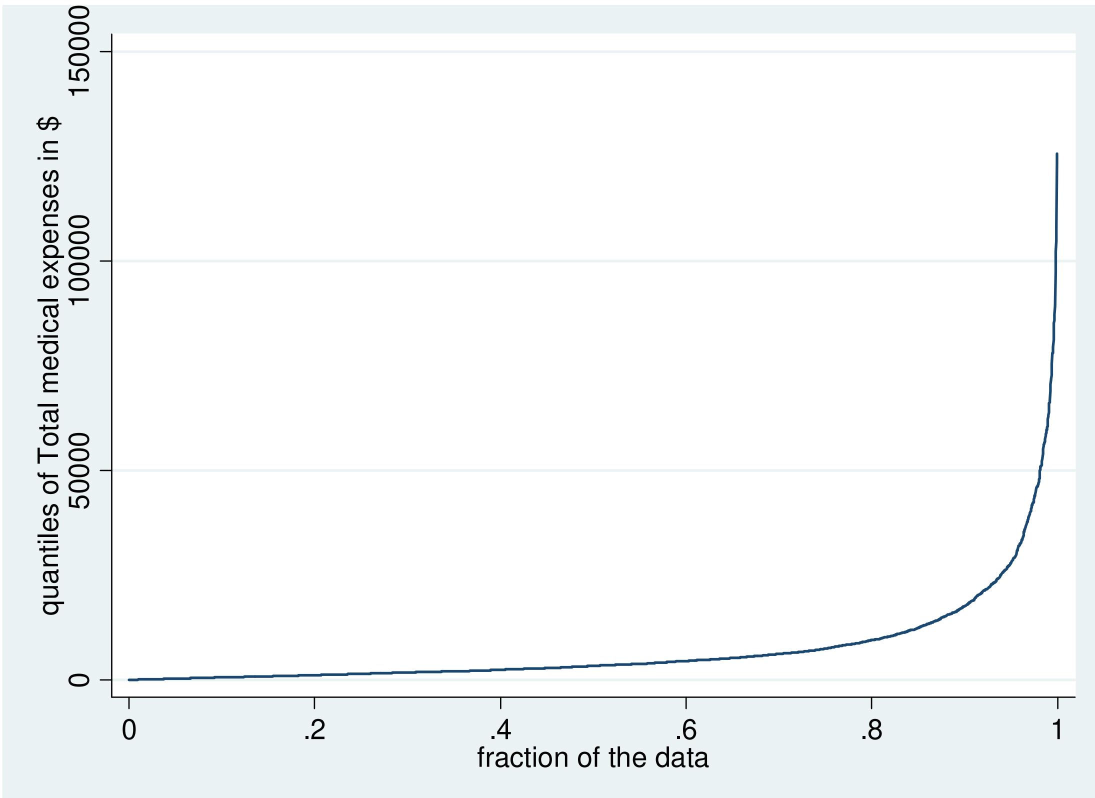
Advantages of the quantile regression

- Flexibility for modeling data with heterogeneous conditional distributions.
- Median regression is more robust to outliers than the OLS regression.
- Richer characterization and description of the data: can show different effects of the independent variables on the dependent variable depending across the spectrum of the dependent variable.

Quantile regression model example

- We would like to study the factors influencing total medical expenditures for people with low-, medium-, and high- expenditures.
- Data are from the Medical Expenditure Panel Survey (MEPS)
- Dependent variable: total medical expenditures
- Independent variables: has supplemental insurance, total number of chronic conditions, age, female, and white
- We estimate an OLS regression, and quantile regressions at the 25^{th}, 50^{th}, and 75^{th} quantile.

Dependent variable by quantiles



Quantile regression coefficients at different quantiles

Total medical expenditures	OLS regression	Quantile regression at 0.25 quantile	Quantile regression at 0.5 quantile	Quantile regression at 0.75 quantile
Supplementary private insurance	585*	453*	687*	708
Number of chronic problems	2528*	782*+	1332*+	2855*
Age	7*	16*	35*	87*
Female	-1239	16+	-260+	-554
White	2193	338	632	801
Intercept	461	-1412	-2252*	-4512

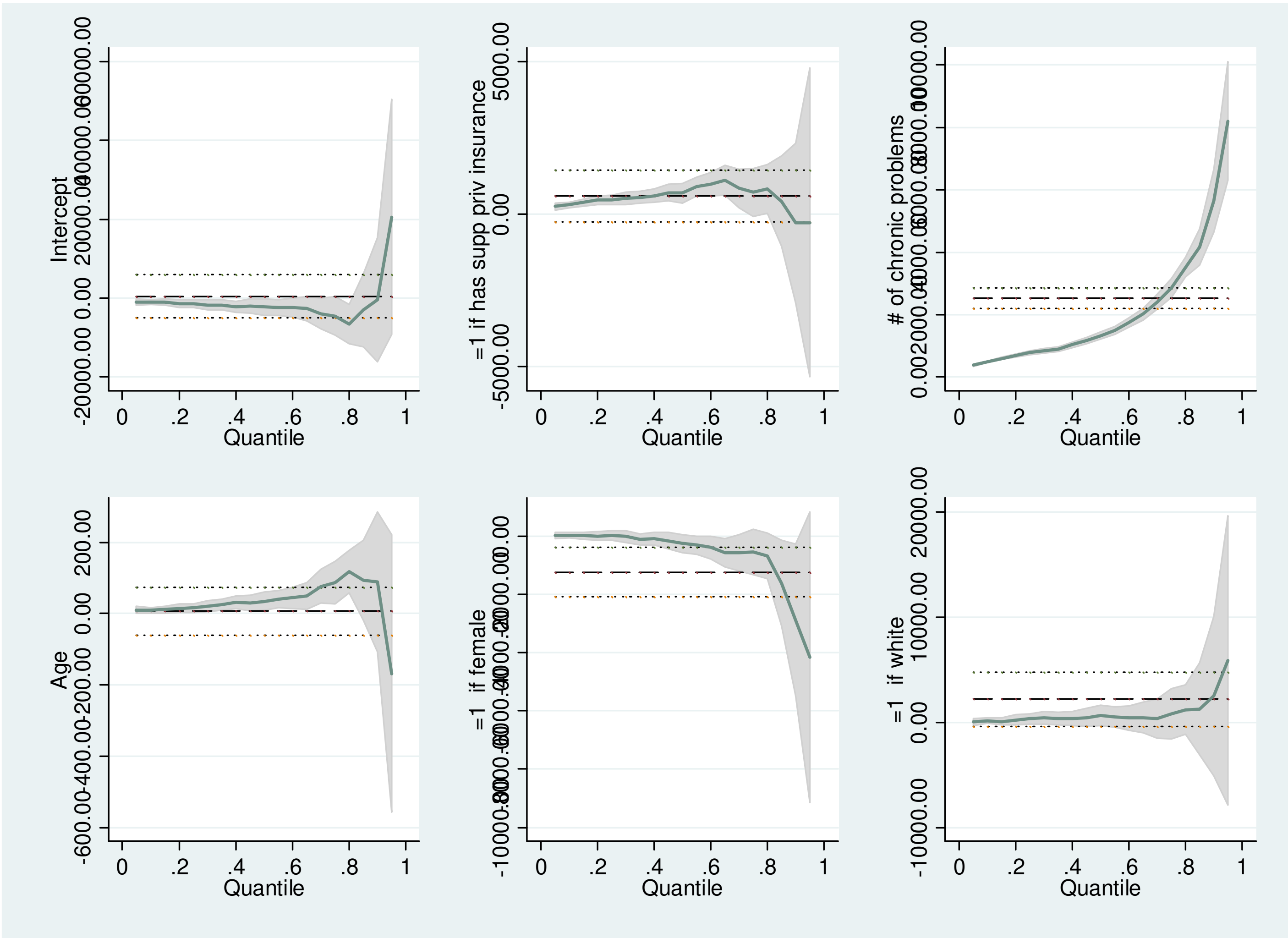
*: Significantly different quantile regression coefficient from zero at the 5% significance level.

+: Significantly different quantile regression coefficients from OLS coefficients at the 5% significance level, when the OLS coefficient is outside of the quantile regression coefficient confidence interval.

- Interpretation of the coefficients: individuals who have one more chronic problem spend $782 more in total medical expenditures for those with low total expenditures (at the 25% quantile) and $2,855 more in total medical expenditures for those with high total expenditures (at the 75% quantile). In other words, the effect of the number of chronic conditions increases for individuals with higher expenditures (higher quantiles).

- There are two types of significant coefficients: those that are significantly different from zero, and the quantile coefficients that are significantly different from the OLS coefficients (outside of the OLS confidence interval). For example, the coefficient on the number of chronic problems at the 75% quantile is significantly different from zero but not significantly different from the OLS coefficient.
- You may want to test for significant differences in coefficients between different quantiles.
- We need to conduct a heteroscedasticity test to justify the use of quantile regression. We find that the Breusch-Pagan test statistic is significantly different than zero, therefore we have heteroscedasticity and are justified in the use of quantile regression.

Quantile regression coefficients (Stata)



- The quantiles of the dependent variable are on the horizontal axis, and the coefficient magnitudes on the vertical axis.
- The OLS coefficient is plotted as a horizontal line with the confidence interval as two horizontal lines around the coefficient line. The OLS coefficient doesn't vary by quantiles.
- The quantile regression coefficients are plotted as lines varying across the quantiles with confidence intervals around them. If the quantile coefficient is outside the OLS confidence interval, then we have significant differences between the quantile and OLS coefficients (as denoted by the $^{+}$ sign in the table).
- The quantile coefficients for the number of chronic problems (independent variable) on total medical expenditures (dependent variable) are significantly different from the OLS coefficients. Moreover, the effect of the number of chronic conditions increases for individuals with higher expenditures (higher quantiles), similarly to what we concluded from the table.