• Quantile regression (Koenker and Bassett, 1978) is increasingly used by practitioners but it is still not part of the standard econometric/statistics courses.

• Road map:
  • general introduction to quantile regression
  • two topics from recent research:
    • models with time-invariant individual (“fixed effects”) effects
    • structural quantile function.

• I will present the approach to these problems proposed by Machado and Santos Silva (2019), and illustrate the use of the corresponding Stata commands *xtqreg* and *ivqreg2*. 
2. Conditional quantiles

- For $0 < \tau < 1$, the $\tau$-th quantile of $y$ given $x$ is defined by
  
  \[ Q_y(\tau|x) = \min\{\eta | P(y \leq \eta|x) \geq \tau\}. \]
3. Basics of quantile regression

- Quantile regression estimates $Q_y(\tau|x)$.
- Throughout we assume linearity: $Q_y(\tau|x) = x'\beta(\tau)$.
- With linear quantiles, we can write

$$y = x'\beta(\tau) + u(\tau); \quad Q_{u(\tau)}(\tau|x) = 0.$$ 

- Note that the errors and the parameters depend on $\tau$.
- For $\tau = 0.5$ we have the median regression.
- We need to restrict the support of $x$ to ensure that quantiles do not cross.
4. Inference

- The estimator of $\beta (\tau)$ is defined by

$$
\hat{\beta} (\tau) = \arg \min_b \frac{1}{n} \left\{ \sum_{y_i \geq x_i' b} \tau |y_i - x_i' b| + \sum_{y_i < x_i' b} (1 - \tau) |y_i - x_i' b| \right\}.
$$

- The F.O.C. can be written as

$$
\frac{1}{n} \sum_{i=1}^{n} \left( (\tau - 1 \left( (y_i - x_i' \hat{\beta} (\tau)) < 0 \right)) \right) x_i = 0.
$$

- $\hat{\beta} (\tau)$ is invariant to perturbations of $y_i$ that do not change the sign of $(y_i - x_i' \hat{\beta} (\tau))$.

- $\hat{\beta} (\tau)$ can be estimated by linear programming (see \texttt{qreg}).
• Asymptotic theory is **non-standard** because the objective function is not differentiable.

• However, under certain regularity conditions, \( \hat{\beta}(\tau) \) has standard properties:

\[
\sqrt{n} (\hat{\beta}(\tau) - \beta(\tau)) \xrightarrow{d} \mathcal{N}(0, D^{-1} A D^{-1}),
\]

\[
D = \mathbb{E} \left[ f_{u(\tau)}(0 \mid x_i) x_i x_i' \right], \quad A = \mathbb{E} \left[ (\tau - 1(u(\tau) \leq 0))^2 x_i x_i' \right].
\]

• It is possible to estimate \( A \) and \( D \) under different assumptions (see \texttt{qreg} and \texttt{qreg2}).
• The main advantage of quantile regression is the **informational gains** they provide.

• Quantiles are “**robust**” measures of location and are estimated using a “**robust**” estimator.

• Quantiles and means have very **different** properties.
  
  • Quantiles are not **additive**; the quantile of the sum is not the sum of the quantiles.
  
  • Quantiles are **equivariant** to non-decreasing transformations; for example, if $y_i$ is non-negative with

    $$Q_{y_i} (\tau | x_i) = \exp \left( x_i' \beta \left( \tau \right) \right),$$

    then,

    $$Q_{\ln(y_i)} (\tau | x_i) = x_i' \beta \left( \tau \right).$$
• The plain-vanilla quantile regression estimator has been extended to different settings:
  • Censored regression; Powell (1984)
  • Binary data; Manski (1975, 1985), Horowitz (1992)
  • Ordered data; M.-j. Lee (1992)
  • Count data; Machado and Santos Silva (2005)
  • Corner-solutions data; Machado, Santos Silva, and Wei (2016)
  • Clustering; Parente and Santos Silva (2016)

• Two areas of active research are:
  • quantile regressions with time-invariant individual ("fixed") effects, and
  • structural quantile function.
Consider a location-scale model

\[ y_i = x_i' \beta + (x_i' \gamma) u_i, \]

where \( x_i \) and \( u_i \) are independent and \( \Pr(x_i' \gamma > 0) = 1 \).

In this case the mean and all conditional quantiles are linear

\[ Q_y(\tau|x) = x_i' \beta + (x_i' \gamma) Q_u(\tau|x_i) = x_i' \beta(\tau) \]

\[ \beta(\tau) = \beta + \gamma Q_u(\tau). \]

In this model, the information provided by \( \beta, \gamma \), and \( Q_u(\tau) \) is equivalent to the information provided by regression quantiles.
Machado and Santos Silva (2019) noted that, assuming $E(U) = 0$ and using the normalization $E(|U|) = 1$, $\beta$ and $\gamma$ are identified by conditional expectations:

\[ E[y_i|x_i] = \beta_0 + \beta_1 x_i \]

\[ E[|y_i - \beta_0 - \beta_1 x_i| | x_i] = \gamma_0 + \gamma_1 x_i \]

$Q_u(\tau|x_i)$ can be estimated from the scaled errors

\[ \frac{y_i - \beta_0 - \beta_1 x_i}{\gamma_0 + \gamma_1 x_i} \]

This provides a way to estimate quantile regression using two OLS regressions and the computation of a univariate quantile.
8. Panel data

- Suppose now that we are interested in estimating

\[ Q_{yit}(\tau|x_{it}, \eta_i) = x_{it}' \beta(\tau) + \eta(\tau)_i, \text{ with } i = 1, \ldots, n; \ t = 1, \ldots, T. \]

- As in mean regression, “fixed effects” can be important.
• Estimation of quantile regression with fixed effects is difficult because there is no transformation that can be used to eliminate the incidental parameters.

• Therefore, due to the incidental parameter problem, consistency requires that both $n \to \infty$ and $T \to \infty$.

• For fixed $T$, the only realistic option is the "correlated random effects" (Mundlak) estimator; see Abrevaya and Dahl (2008).

• Roger Koenker (2004) and Canay (2011) proposed estimators based on the assumption that $\eta (\tau)_i = \eta_i$ but this goes against the spirit of quantile regression.
• Kato, Galvão, and Montes-Rojas (2012) studied the properties of quantile regression in a model where the fixed effects are explicitly included as **dummies**.

• The estimator is consistent and asymptotically normal when both \( n \to \infty \) and \( T \to \infty \) with \( n^2 \left[ \ln(n) \right]^3 / T \to 0 \).

• This is an issue because in many applications \( n \) is much larger than \( T \) (e.g. for \( T = 40 \), \( n = 100 \), \( n^2 \left[ \ln(n) \right]^3 / T = 24,416 \)).

• An alternative is to use the quantiles-via-moments estimator.
Consider the location-scale model for panel data

\[ y_{it} = \alpha_i + x_{it}' \beta + (\delta_i + x_{it}' \gamma) u_{it} \]

\[ \eta (\tau)_i = \alpha_i + \delta_i Q_u(\tau), \quad \beta (\tau) = \beta + \gamma Q_u(\tau), \]

where \( x_i \) and \( u_i \) are independent and \( \Pr ((\delta_i + x_{it}' \gamma) > 0) = 1 \).

Estimation is performed using two fixed effects regressions (\texttt{xtreg}) and computing a univariate quantile.

Consistency requires \((n, T) \to \infty \) with \( n = o(T) \).

For fixed \( T \) the estimator will have a bias but:

- simulations suggest that the bias is negligible for \( n / T \leq 10 \);
- the bias can be removed using \texttt{jackknife}.

The estimator is implemented in the \texttt{xtqreg} command (available from SSC)
xtqreg depvar [indepvars] [if] [in] [, options]

- **quantile(#[#[# ...]])**: estimates # quantile; default is quantile(.5)

  - **id**: specifies the variable defining the panel

  - **ls**: displays the estimates of the location and scale parameters
9. Endogeneity

- Suppose that we have a structural relationship defined by

\[
y = d\alpha + x'\beta + u,
\]
\[
d = \delta(x, z, v)
\]

where \(v\) may not be independent of \(u\)

- We are interested in

\[
S_y(\tau|d, x) = d\alpha(\tau) + x'\beta(\tau),
\]

the structural quantile function such that:

- \(\Pr[y < S_y(\tau|d, x)|z, x] = \tau\),
- \(S_y(\tau|d, x) = Q_y(\tau|z, x) \neq Q_y(\tau|d, x)\).
• **Chernozhukov and Hansen (2008)** propose an estimator of $S_Y(\tau|d, x)$ based on the observation that

$$Q_{y_d\alpha(\tau)}(\tau|z, x) = x'\beta(\tau) + z\gamma(\tau)$$

with $\gamma(\tau) = 0$.

• We can implement the estimator by:
  
  • estimating $\beta(\tau)$ and $\gamma(\tau)$ for a range of values of $\alpha(\tau)$
  
  • and choosing as estimates the ones corresponding to the value of $\alpha(\tau)$ for which $\gamma(\tau)$ is in some sense closer to zero.

• Chernozhukov and Hansen (2008) prove the consistency and asymptotic normality of the estimator.

• The estimator is difficult to implement when there are multiple endogenous variables, but there have been a number of recent developments on this.
• Again, the quantile-via-moments estimator can be useful.
• Consider a location-scale structural relationship

\[ y = d\alpha + x'\beta + (d\delta + x'\gamma)\ u, \quad d = \delta(x, z, v), \]

where \( v \) may not be independent of \( u \) but \( u \) is independent of \( x \) and \( z \).
• Because \( S_y(\tau|d, x) \) is such that \( \Pr[y < S_y(\tau|d, x)|z, x] = \tau, \)

\[
\begin{align*}
S_y(\tau|d, x) & = d\alpha + x'\beta + (d\delta + x'\gamma)\ Q_u(\tau) \\
& = d(\alpha + \delta Q_u(\tau)) + x(\beta + \gamma Q_u(\tau)).
\end{align*}
\]
• GMM can be used to estimate the structural parameters:

\[
E \left[ \left( \frac{y_i - d\alpha - x'\beta}{d\delta + x'\gamma} \right) \bigg| z_i \right] = 0,
\]

\[
E \left[ \left( \frac{|y_i - d\alpha - x'\beta|}{d\delta + x'\gamma} - 1 \right) \bigg| z_i \right] = 0.
\]

• \( Q_u(\tau) \) can be estimated from the standardized errors

\[
(y_i - d\hat{\alpha} - x'\hat{\beta}) / (d\hat{\delta} + x'\hat{\gamma}).
\]

• The estimator has the usual properties.

• The estimator is implemented in the \texttt{ivqreg2} command (available from SSC)
ivqreg2 depvar [indepvars] [if] [in] [, options]

quantile(#[#[# ...]]): estimates # quantile; default is quantile(.5)

instruments(varlist): list of instruments, including control variables; by default no instruments are used and restricted quantile regression is performed

ls: displays the estimates of the location and scale parameters
• Quantile regression can be very useful and it is now easy to implement in a variety of cases.

• In some contexts, however, quantile regression can be challenging.

• The Method of Moments-Quantile Regression estimator can be useful in some of these cases.

• `xtqreg` and `ivqreg2` make it easy to estimate quantile regressions with “fixed effects” or endogenous variables.


