Quantile Regression under misspecification, with an application to the U.S. wage structure

Angrist, Chernozhukov and Fernandez-Val

Reading Group Econometrics

November 2, 2010
Intro: initial problem

- The paper starts with OLS: under misspecification, OLS minimizes the mean-squared error linear approximation.
  - Suppose the true model is $Y_i = g(X_i) + \varepsilon_i$
  - OLS estimation minimizes $\Delta(\beta)^2 = (g(X_i) - X_i\beta)^2$
- The first question of the paper is: What are the properties of quantile regression (QR) when the model is misspecified?
ACFV show that QR estimation is equivalent to the minimization of a weighted mean-squared error linear approximation.

This allows:
- To decompose the contribution of covariates in the estimation.
- To interpret partial regression and to investigate omitted variable bias.

ACFV also derive distribution theory of quantile process.
Framework and notations

- Conditional quantile function:

\[ Q_\tau(Y|X) = \inf\{ y : F_Y(y|X) \geq \tau \} \]

- It solves:

\[ Q_\tau(Y|X) \in \arg\min_{q(X)} E[\rho_\tau(Y - q(X))] \]

with the check function:

\[ \rho_\tau(u)(\tau - 1(u \leq 0))u \]

- Restricting to linear functions: \( q(X) = X'\beta \), the parameter vector solves:

\[ \beta(\tau) = \arg\min_{\beta} E[\rho_\tau(Y - X'\beta)] \]
The QR specification error is then:

\[ \Delta_T(X, \beta) = X' \beta - Q_T(Y|X) \]

The deviation from the true quantile specification is:

\[ \varepsilon_T = Y - Q_T(Y|X) \]

with density \( f_{\varepsilon_T}(e|X) \).
First result (Theorem 1)

- Under assumptions:
  (i) $f_Y(y|X)$ exists.
  (ii) $E(Y), E(Q_\tau(Y|X))$ and $E||X||$ are finite
  (iii) $\beta(\tau)$ is a unique QR estimator

- Then $\beta(\tau)$ solves:

$$\min_\beta E[w_\tau(X, \beta).\Delta_\tau(X, \beta)^2]$$

- with:

$$w_\tau(X, \beta) = \int_0^1 (1 - u)f_{\epsilon_\tau}(u\Delta_\tau(X, \beta)|X)du$$

$$= \int_0^1 (1 - u)f_Y(uX'\beta + (1 - u)Q_\tau(Y|X)|X)du$$
Theorem 1: interpretation of weights

\[ w_\tau(X, \beta) = \int_0^1 (1 - u)f_Y[uX'\beta + (1 - u)Q_\tau(Y|X)|X]du \]

- \( uX'\beta + (1 - u)Q_\tau(Y|X) \) describes a line between the approximation point and the true one.
- Weights are thus a weighted average of the density of \( Y \) in this segment.
- More weight is given to points of the segment close to the true value of the conditional quantile.
- ACFV refer to \( w_\tau(X, \beta) \) as “importance weights”. Overall weights are given by (\( \pi(x) \) is the density of \( X \)):
  \[ w_\tau(x, \beta) \times \pi(x) \]
- Weights can be approximated by:
  \[ w_\tau(X, \beta) = \frac{1}{2}f_Y[Q_\tau(Y|X)|X] + \varrho_\tau(X) \]
- If \( f_Y \) is locally almost constant on the segment, \( \varrho_\tau(X) \approx 0 \) (more importance is given where \( f_Y \) is large).
Second result (Theorem 2)

- Under assumptions:
  (i) $f_Y(y|X)$ exists. $E(Y)$
  (ii) $E(Y)$, $E(Q_\tau(Y|X))$ and $E|X|$ are finite
  (iii) $\beta(\tau)$ is a unique QR estimator

- Then $\bar{\beta}(\tau)$ uniquely solves:

$$
\bar{\beta}(\tau) = \arg\min_\beta E[\bar{\omega}_\tau(X, \bar{\beta}(\tau)).\Delta_\tau(X, \beta)^2]
$$

- with:

$$
\bar{\omega}_\tau(X, \bar{\beta}(\tau)) = 1/2 \int_0^1 f_{\epsilon_\tau}(u\Delta_\tau(X, \bar{\beta}(\tau))|X)du
$$

$$
= 1/2 \int_0^1 f_Y(uX'\bar{\beta}(\tau) + (1 - u)Q_\tau(Y|X)|X)du
$$
Theorem 2: Implications

- Weights do not depend on $\beta$ anymore.
- QR solves a weighted least square approximation.
- As before, weights can be approximated (under smoothness assumption and if $f$ is locally constant):

$$\bar{w}_\tau(X, \tilde{\beta}(\tau)) \approx w_\tau(X, \beta(\tau)) \approx 1/2f_Y[Q_\tau(Y|X)|X]$$
Example from US census

A. $\tau = 0.10$

B. $\tau = 0.50$

C. $\tau = 0.90$

D. $\tau = 0.10$

E. $\tau = 0.50$

F. $\tau = 0.90$
Example from US census

- Minimum distance (MD) estimator is obtained as in Chamberlain (1994) as:
  \[
  \bar{\beta}(\tau) = \arg\min_\beta E[\Delta_\tau(X, \beta)^2]
  \]

- Differences in slope come from difference in weights.

- Importance weights and densities approximations are very close:
  \[
  w_\tau(X, \beta(\tau)) \approx 1/2f_Y[Q_\tau(Y|X)|X]
  \]
## TABLE I

### COMPARISON OF CQF AND QR-BASED INTERQUANTILE SPREADS

<table>
<thead>
<tr>
<th>Census</th>
<th>Obs.</th>
<th>90–10</th>
<th>90–50</th>
<th>50–10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>CQ</td>
<td>QR</td>
<td>CQ</td>
</tr>
<tr>
<td>A. Overall</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1980</td>
<td>65,023</td>
<td>1.20</td>
<td>1.19</td>
<td>0.51</td>
</tr>
<tr>
<td>1990</td>
<td>86,785</td>
<td>1.35</td>
<td>1.35</td>
<td>0.60</td>
</tr>
<tr>
<td>2000</td>
<td>97,397</td>
<td>1.43</td>
<td>1.45</td>
<td>0.67</td>
</tr>
<tr>
<td>B. High school graduates</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1980</td>
<td>25,020</td>
<td>1.09</td>
<td>1.17</td>
<td>0.44</td>
</tr>
<tr>
<td>1990</td>
<td>22,837</td>
<td>1.26</td>
<td>1.31</td>
<td>0.52</td>
</tr>
<tr>
<td>2000</td>
<td>25,963</td>
<td>1.29</td>
<td>1.32</td>
<td>0.59</td>
</tr>
<tr>
<td>C. College graduates</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1980</td>
<td>7,158</td>
<td>1.26</td>
<td>1.19</td>
<td>0.61</td>
</tr>
<tr>
<td>1990</td>
<td>15,517</td>
<td>1.44</td>
<td>1.38</td>
<td>0.70</td>
</tr>
<tr>
<td>2000</td>
<td>19,388</td>
<td>1.55</td>
<td>1.57</td>
<td>0.75</td>
</tr>
</tbody>
</table>

Notes: The sample consist of U.S.-born black and white men aged 40–49. The table shows average measures calculated using the distribution of the covariates in each year. The covariates are schooling, race, and a quadratic function of experience. Sampling weights were used for the 2000 Census.
Partial Quantile Regression

Since QR solve a weighted least square approximation, ACFV use it to derive a partial regression framework and omitted variable bias formula.

- Suppose $X = (X_1, X_2)$. Define $\beta'(\tau) = (\beta'_1(\tau), \beta'_2(\tau))$
- Partialling out is obtained by computing WLS of $Q_\tau(Y|X)$ and $X_1$ on $X_2$:
  - $Q_\tau(Y|X) = X'_2\pi_Q + q_\tau(Y|X)$
  - $X_1 = X'_2\pi_1 + V_1$
  - Weights are $\bar{w}_\tau(X) = \bar{w}_\tau(X, \bar{\beta}(\tau))$
- In a second step, $\beta_1(\tau)$ is obtained minimizing the weighted sum of squares of errors:

$$\beta_1(\tau) = \arg\min_{\beta_1} E[\bar{w}_\tau(X).((q_\tau(Y|X) - V_1\beta_1)^2]$$

or

$$\beta_1(\tau) = \arg\min_{\beta_1} E[\bar{w}_\tau(X).((Q_\tau(Y|X) - V_1\beta_1)^2]$$
Omitted Variable Bias

Using these results, we can derive omitted variable bias formula. Suppose we do not observe $X_2$.

- We obtain $\gamma_1(\tau)$ as:

$$\gamma_1(\tau) = \arg\min_{\gamma_1} E[\rho_\tau(Y - X_1 \gamma_1)]$$

- That can be compared to the results for $\beta_1(\tau)$ in:

$$(\beta_1(\tau), \beta_2(\tau)) = \arg\min_{\beta_1, \beta_2} E[\rho_\tau(Y - X_1' \beta_1 - X_2' \beta_2)]$$

- The difference between the two is given by:

$$\gamma_1(\tau) = \beta_1(\tau) + (E[\tilde{w}_\tau(X)X_1'X_1])^{-1}E[\tilde{w}_\tau(X)X_1'R_\tau(X)]$$
Omitted Variable Bias

The bias, \((E[\tilde{w}_\tau(X)X'_1X_1])^{-1}E[\tilde{w}_\tau(X)X'_1R_\tau(X)]\), depends on:

- \(\tilde{w}_\tau(X) = 1/2 \int_0^1 f_{\varepsilon_\tau}(u.\Delta_\tau(X, \gamma_1(\tau))|X)du\)

- \(\Delta_\tau(X, \gamma_1(\tau)) = X'_1\gamma_1 - Q_\tau(Y|X)\)

- \(\varepsilon_\tau = Y - Q_\tau(Y|X)\)

- \(R_\tau(X) = Q_\tau(Y|X) - X'_1\beta_1\)

Bias is similar to OLS omitted variable bias.
Sampling Properties under misspecification

How does misspecification affects inference?

- As we saw before, QR consistently estimates the approximation of conditional quantile function.
- What about inference on this approximation?
- ACFL take into account the whole QR process:
  - Allows to infer on several quantiles simultaneously.
  - Thus define $\hat{\beta}(.)$ as the set of estimators $\hat{\beta}(\tau), \tau \in T$ a close subset of $]0, 1[$.
Theorem 3

- Under assumptions:
  1. \((Y_i, X_i)\) are iid
  2. \(f_Y(y|X)\) exists, is bounded and uniformly continuous
  3. \(\forall \tau, J(\tau) = E[f_Y(X'\beta(\tau)|X)X'X] > 0\)
  4. \(E||X||^{2+\varepsilon} < \infty\) for some \(\varepsilon > 0\)

- Then:

\[
\sup_{\tau \in T} ||\hat{\beta}(\tau) - \beta(\tau)|| = o_p(1) \tag{1}
\]

\[
J(.)\sqrt{n}(\hat{\beta}(.) - \beta(.)) \xrightarrow{n \to \infty} z(.) \tag{2}
\]

- Where:

\[
z(.) \sim \mathcal{N}(0, \Sigma(\tau, \tau'))
\]

\[
\Sigma(\tau, \tau') = E(z(\tau), z(\tau'))
\]

\[
= E[(\tau - 1\{Y < X'\beta(\tau)\})(\tau' - 1\{Y < X'\beta(\tau')\})X'X]
\]
Theorem 3

\[ \Sigma(\tau, \tau') = E \left[ (\tau - 1\{Y < X'\beta(\tau)\})(\tau' - 1\{Y < X'\beta(\tau')\})X'X \right] \]

Under the assumption that the model is well specified, \( Q_\tau(Y|X) = X'\beta(\tau) \), the Covariance matrix is:

\[ \Sigma_0(\tau, \tau') = [\min(\tau, \tau') - \tau\tau']E \left[ X'X \right] \]

- Joint normality is obtained even with misspecification
- However, misspecification affects the form of the covariance-matrix.
- This raises difficulties to make inference.
Testing hypotheses

ACFV provide a method to test hypotheses of the form:

$$H_0 : \ R(\tau)\beta(\tau) = r(\tau) \ \forall \tau \in T$$

- They use a Kolmogorov statistic (using consistent estimates of $\Sigma_0(\tau, \tau')$ and $J(.)$).
- Under misspecification, this statistic has no well-behaved distribution
- Bootstrap is used to find the quantiles of this statistic.
Example from US census

A. 1980
B. 1990
C. 2000

Quantile Index
Schooling Coefficient (%)
95% Robust Uniform CI
95% Robust Pointwise CI

Quantile Index
Schooling Coefficient (%)
95% Robust Uniform CI
95% Robust Pointwise CI

Quantile Index
Schooling Coefficient (%)
95% Robust Uniform CI
95% Robust Pointwise CI
Example from US census
Example from US census

Figure 6: Unconditional quantiles and conditional quantiles of log-earnings in 1980, 1990, and 2000 censuses (US-born white)
Example from US census

- “robust 95% simultaneous confidence intervals” are computed by subsampling
- More increase in the second half...