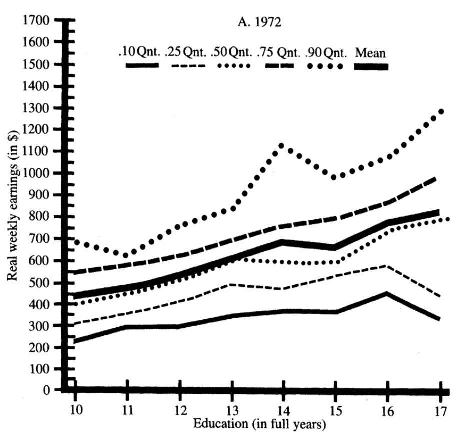
Assignment 3:
Microeconometrics

Daniel Redel (210630)
Instructor: Dr. Pavel Čížek

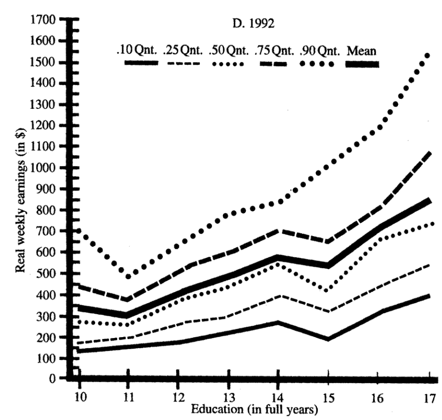
Fall 2022

Quantile Regression

1 According to Buschinksy (1998), a well known observed phenomenon is the substantial **increase in wage inequality** in recent years. This disparity in income across individuals is significant even after controlling for individual characteristics. Additionally, a relevant change in the **return of skills** can be observed, either in the form of changes in education or professional experience. These two examples motivate the idea of studying these income trends **across the wage distribution**.



(a) 1972



(b) 1992

Figure 1: Weekly Earnings for Individuals with 15 Years Experience by Quantile (Buschinksy; 1998)

Figure 1 from the paper illustrates the relevance of studying changes in earnings at different points of the distribution, accounting for the heterogeneity in these changes among individuals. For example, in both cases, we observe how weekly earnings in this sample increased at every quantile with the level of education level. However, there is also a much steeper increase in returns to education at higher quantiles in the year 1992 (figure on the right) than in 1979 (figure on the left). Even more, little increases occurred at the lower part of the distribution.

Quantile regression allows us to do this analysis for each part of the distribution separately, compared to the standard OLS estimator that only looks at the average. Quantile regression gives us the tools to separate the individuals by their observable characteristics and outcomes and see how the returns on skills and final wage structurally differ at 0.10, 0.25, 0.50, 0.75, and 0.90 quantiles.

2 Consider the following linear regression:

$$\begin{aligned} y_i &= x_i'\beta + \varepsilon_i \\ \varepsilon_i &\sim \text{ Laplace}(0,1) \end{aligned}$$

this is because of:

$$\begin{aligned} f_u(\varepsilon_i, \mu = 0, b = 1) &= \frac{1}{2b}\exp\left(-\frac{|\varepsilon_i - \mu|}{b}\right) \\ &= \frac{1}{2 \times 1}\exp\left(-\frac{|\varepsilon_i - 0|}{1}\right) \\ &= \frac{\exp\left(-|\varepsilon_i|\right)}{2} \end{aligned}$$

where $E(\varepsilon_i) = \mu = 0$ and $V(\varepsilon_i) = 2b^2 = 2$ for $i = 1, ..., N$.

We want to compare the asymptotic variance between the median regression and the OLS estimator. For the OLS estimator, the asymptotic variance is defined by:

$$\begin{aligned} \text{Avar}(\hat{\beta}^{OLS}) &= \sigma^2[E(X'X)]^{-1} \\ &= 2[E(X'X)]^{-1} \end{aligned}$$

For the median regression ($\tau = 0.5$), on the other hand, we have the following asymptotic variance[1]:

$$\begin{aligned} \text{Avar}(\hat{\beta}^{MR}) = \Lambda_\tau &= \sigma_\tau^2[E(X'X)]^{-1} \\ \sigma_\tau^2 &= \frac{\tau(1-\tau)}{f_{u_\tau}^2(0)} \end{aligned}$$

As the errors follow the Laplace distribution already described above, we know that the probability density function (pdf) at point 0 is defined by $f_{\varepsilon_\tau}(0) = 0.5$ (assuming ε_i is independent of x_i). From this, we can estimate our Asymptotic Variance:

$$\frac{\tau(1-\tau)}{f_{u_\tau}^2(0)}[E(X'X)]^{-1} = \frac{0.5(1-0.5)}{(0.5)^2}[E(X'X)]^{-1} = 1 \times [E(X'X)]^{-1}$$

Therefore, we see that the asymptotic variance of the median regression is smaller relative to the variances of OLS estimator:

$$\text{Avar}(\hat{\beta}^{MR}) < \text{Avar}(\hat{\beta}^{OLS}) \Leftrightarrow [E(X'X)]^{-1} < 2[E(X'X)]^{-1}$$

[1] Buschinksy (1998), p.98.

3 Let's define and interpret the following equivariance properties:

- **Equivariance Property (8)**: $\hat{\beta}(\theta, \lambda y, X) = \lambda\beta(\theta, y, X)$ for $\lambda \in [0, \infty^+)$.
  - Property (8), also called *scale equivariance*, state that the estimators β that maximises the objective function are linear on y, which implies that if the variable y is scaled by some positive constant, then the estimators should also be scaled by this same constant.
  - Put differently, multiplying the dependent variable y by some constant λ has the effect of multiplying the fitted response β by the same constant. Larger y means equivalently larger β.
  - Using our earnings example, the regression estimates do not change when looking at earnings per week or per year, in dollars or cents.
- **Equivariance Property (9)**: $\hat{\beta}(1-\theta, \lambda y, X) = \lambda\beta(\theta, y, X)$ for $\lambda \in (\infty^-, 0]$.
  - Property (9) is a different version of scale equivariance and states that, when multiplying with a negative constant, the distribution 'flips' —therefore also flip the quantiles— but the estimates stay invariant.
  - Recall that changing the sign on y is the same as exchanging θ for $1-\theta$. "The area under the curve" for quantile θ in y is the same that the one for quantile $1-\theta$ in $-y$.
- **Equivariance Property (10)**: $\hat{\beta}(\theta, y + X\gamma, X) = \beta(\theta, y, X) + \gamma$ for $\gamma \in R^K$.
  - Property (10) is also called *linear shift equivariance*. This property deals with addition rather than multiplication. The idea is that, transforming the dependent variable y by adding a linear combination of the regressor variables, in our case γ, affects the regression estimator only in that γ is added to the estimator based on the original data.
  - For example, if every person in the sample would receive a \$1000 dollar bonus, the conditional mean and intercept would increase by 1000, but the slopes will be unaffected.
- **Equivariance Property (11)**: $\hat{\beta}(\theta, y, XA) = A^{-1}\beta(\theta, y, X)$ for $A_{K\times K}$ is nonsingular.
  - Property (11) or *affine equivariance* resembles the multiplicative (scale) invariance when scaling y, but now concerning the independent variables X. If the matrix X is multiplied, and therefore re-scaled, by some non-singular square matrix $(k \times k)$ of constants A, then the estimators are re-scaled by this same matrix.
  - Here, a linear transformation of the regressor variables does not affect the fitted response variable.

4 For each case, we will derivate the OLS estimator from the standard quadratic loss function to study whether OLS satisfies equivariance properties (8), (10) and (11):

$$\min \sum_{i=1}^{N} (y_i - x_i'\beta)(y_i - x_i'\beta)$$

Let's impose the different equivariance transformations to derive the $\hat{\beta}^{OLS}$ in matrix form.

Scale Equivariance Property (8):

$$\begin{aligned} e'e &= (\lambda y - X\hat{\beta})'(\lambda y - X\hat{\beta}) \\ &= (\lambda y)'(\lambda y) - 2\hat{\beta}'X'(\lambda y) + \hat{\beta}'X'X\hat{\beta} \end{aligned}$$

We take the First Order Conditions (FOC):

$$\begin{aligned} \frac{\partial e'e}{\partial \hat{\beta}} &\Rightarrow -2X'(\lambda y) + 2X'X\hat{\beta} = 0 \\ &\Rightarrow 2X'X\hat{\beta} = 2X'(\lambda y) \\ &\Rightarrow X'X\hat{\beta} = X'(\lambda y) \\ &\Rightarrow (X'X)^{-1}X'X\hat{\beta} = (X'X)^{-1}X'(\lambda y) \\ &\Rightarrow I\hat{\beta} = (X'X)^{-1}X'(\lambda y) \end{aligned}$$

Now we can rewrite this as:

$$\begin{aligned} \hat{\beta}^{OLS}(\lambda y, X) &= (X'X)^{-1}X'(\lambda y) \\ &= (X'X)^{-1}X'\lambda y \\ &= \lambda(X'X)^{-1}X'y \\ &= \lambda\hat{\beta}^{OLS} \end{aligned}$$

The OLS estimator satisfies **Scale Equivariance** (Property 8).

Regression Equivariance Property (10):

$$\begin{aligned} e'e &= ((y + X\gamma) - X\hat{\beta})'((y + X\gamma) - X\hat{\beta}) \\ &= (y + X\gamma)'(y + X\gamma) - 2\hat{\beta}'X'(y + X\gamma) + \hat{\beta}'X'X\hat{\beta} \end{aligned}$$

We take the First Order Conditions (FOC):

$$\begin{aligned} \frac{\partial e'e}{\partial \hat{\beta}} &\Rightarrow -2X'(y + X\gamma) + 2X'X\hat{\beta} = 0 \\ &\Rightarrow 2X'X\hat{\beta} = 2X'(y + X\gamma) \\ &\Rightarrow X'X\hat{\beta} = X'(y + X\gamma) \\ &\Rightarrow (X'X)^{-1}X'X\hat{\beta} = (X'X)^{-1}X'(y + X\gamma) \\ &\Rightarrow I\hat{\beta} = (X'X)^{-1}X'(y + X\gamma) \end{aligned}$$

Now we can rewrite this as:

$$\begin{aligned}
\hat{\beta}^{OLS}(y + X\gamma, X) &= (X'X)^{-1}X'(y + X\gamma) \\
&= (X'X)^{-1}X'y + (X'X)^{-1}X'X\gamma \\
&= (X'X)^{-1}X'y + (X'X)^{-1}X'X\gamma \\
&= (X'X)^{-1}X'y + I\gamma \\
&= \hat{\beta}^{OLS} + \gamma
\end{aligned}$$

The OLS estimator satisfies **Regression Equivariance** (Property 10).

Affine Equivariance Property (11):

$$\begin{aligned}
e'e &= (y - (XA)\hat{\beta})'(y - (XA)\hat{\beta})' \\
&= y'y - 2\hat{\beta}'(XA)'y + \hat{\beta}'(XA)'(XA)\hat{\beta}
\end{aligned}$$

We take the First Order Conditions (FOC):

$$\begin{aligned}
\frac{\partial e'e}{\partial \hat{\beta}} &\Rightarrow -2(XA)'y + 2(XA)'(XA)\hat{\beta} = 0 \\
&\Rightarrow 2(XA)'(XA)\hat{\beta} = 2(XA)'y \\
&\Rightarrow (XA)'(XA)\hat{\beta} = (XA)'y \\
&\Rightarrow [(XA)'(XA)]^{-1}X'X\hat{\beta} = [(XA)'(XA)]^{-1}X'y \\
&\Rightarrow I\hat{\beta} = [(XA)'(XA)]^{-1}(XA)'y
\end{aligned}$$

Now we can rewrite this as:

$$\begin{aligned}
\hat{\beta}^{OLS}(y, XA) &= [(XA)'(XA)]^{-1}(XA)'y \\
&= [A'X'XA]^{-1}A'X'y \\
&= A^{-1}(X'X)^{-1}(A')^{-1}A'X'y \\
&= A^{-1}(X'X)^{-1}X'y \\
&= A^{-1}\hat{\beta}^{OLS}
\end{aligned}$$

The OLS estimator satisfies **Affine Equivariance** (Property 11).

5 Quantile regression can be used to test for homoskedasticity and symmetry of the error distribution.

Test for Homoskedasticity: The idea behind this test is the assumption that, when in presence of homoskedasticity, we should see no difference in the slope coefficients of the different quantile regressions $\hat{\beta}_\tau$ (different intercepts, same slope coefficients). Quantile regression allows us to compare the difference in the error term across the distribution through the differences of each quantile coefficients. In that sense, the null hypothesis can be defined by:

$$H_0 : \hat{\beta}_{\tau_i} = \hat{\beta}_{\tau_j} \text{ where } i \neq j$$

If the errors are homoskedastic, then the coefficients should not any difference, therefore, they should be equal for every quantile. In practice, we can use the estimated coefficients for each quantile and transform it in a way that follows a chi-squared distribution:

$$n(\hat{\beta}_\tau - R\hat{\beta}^R)' A^{-1}(\hat{\beta}_\tau - R\hat{\beta}^R) \to \chi^2_{(pK-p-K+1)}$$

with $\hat{\beta}_\tau$ is the unrestricted vector of p quantile regression estimates and $\hat{\beta}^R$ the restricted coefficient under implied homoskedasticity. The matrix R that multiplies the restricted $\hat{\beta}^R$ allows to compare it against the unrestricted $\hat{\beta}_\tau$, so that the null hypothesis is really $\hat{\beta}_\tau = R\hat{\beta}^R_\tau$.

Test for Symmetry: We can also test for symmetry by exploiting the fact that if the errors are symmetrically distributed, then the mean (if exist) and median of the distribution should coincide. It can also be shown that, under symmetry, the median is the average of two opposite quantiles (τ vs. $1-\tau$). For example, in the case of $\tau = 0.5$, symmetry implies $\hat{\beta}_{0.5} = 0.5 \times (\hat{\beta}_{0.5} + \hat{\beta}_{1-0.5})$. More generally, the null hypothesis can thus written as:

$$H_0 : \hat{\beta}_{0.5} = 0.5 \times (\hat{\beta}_\tau + \hat{\beta}_{1-\tau})$$

Similarly to the homoskedasticity case, we can test this symmetry by transforming the difference between the restricted and unrestricted coefficients to follow a chi-squared distribution:

$$n(\hat{\beta}^U_\tau - R\hat{\beta}^R)' A^{-1}(\hat{\beta}^U_\tau - R\hat{\beta}^R) \to \chi^2_{((p-1)K/2)}$$

6 Table 2 reported in Buschinksy (998) tests for homoskedasticity by way of a χ^2_{84} test, where the null hypothesis states equality among the slope coefficients. The table shows that the null hypothesis is rejected -all the p quantiles being statistically different-, regardless which covariance matrix estimate $A = \hat{\Lambda}_\tau$ is used, for all years and age group. This means that the errors may be heteroscedastic. The value of the test statistic falls in the critical region of the χ^2 distribution with 84 degrees of freedom, at the 0.05 significance level.

We can also say something about wage inequality here. The results in Table 2 not only shows that average wages rise with years of education, but that wage inequality grows as well. The difference in earnings conditional on years of education between the lower and higher quantiles is larger among those with higher level of education, relative to those with lower education levels. In other words, while on average people benefit from extra years of education (via higher wages), this benefit is greatest among the highest earners, and least among the lowest earners.

7 Table 4 reports the returns to education by different "turning-point" levels of education and professional experience. In each of the eight groups the returns to education among people with 25 years of experience are lower than the returns for 15 years of experience, which in turn are lower than the returns to education for 5 years of experience. In simple, returns to education decrease with experience.

These results are consistent with life cycle models developed in human capital theories (Heckman, 1976) or economics of education (Mincer, 1958). The general idea is that, early on in a person's career, the benefits of education are large. As a person grows older, and their career advances, their experience grows, but the returns to education decrease. This is especially true for lower earners. Additional literature (see Haider and Solon, 2006; Bhuller, et al., 2014) tends to confirm the idea that life-cycle effects are an important component in the assessment of the return to education and that constant returns should not be taken for granted.

8 In Table 5, consider the subsample of people of the 18-34 age group with 16 years of education. For year 1972 and 1979, in almost every quantile of the distribution the returns to education for 5 years of experience are nearly twice that of 15 years of experience. However, when compared to what happens in years 1985 and 1992, we observe that this difference has nearly disappeared. Moreover, among the lowest quantiles ($\tau \in \{0.10; 0.25\}$), the returns to education are greater for 15 years of experience than they are for 5.

Then, from 1972 to 1979, the returns to education decreases for both the 5 years and 15 years of experience groups. This decline is greatest among the highest quantiles ($\tau \in \{0.75; 0.90\}$). This tendency reverses between 1979 and 1985. For example, for people with 15 years of experience (1985), the lowest quantiles see a higher return to education than the highest quantiles, where the opposite was true in 1972. In 1992 the differences leveled out, and the returns to education have converged and are relatively high among all quantiles of both experience groups.

In conclusion, all these results confirm the importance of studying changes in earnings using quantile regression to understand the relationship between education and wages across the distribution, rather than averaging over the whole population.