Recent Advances in Quantile Regression Models

A Practical Guideline for Empirical Research

Moshe Buchinsky

ABSTRACT

This paper provides a guideline for the practical use of the semi-parametric technique of quantile regression, concentrating on cross-section applications. It summarizes the most important issues in quantile regression applications and fills some gaps in the literature. The paper (a) presents several alternative estimators for the covariance matrix of the quantile regression estimates; (b) reviews the results for a sequence of quantile regression estimates; and (c) discusses testing procedures for homoskedasticity and symmetry of the error distribution. The various results in the literature are incorporated into the generalized method of moments framework. The paper also provides an empirical example using data from the Current Population Survey, raising several important issues relevant to empirical applications of quantile regression. The paper concludes with an extension to the censored quantile regression model.

I. Introduction

The semi-parametric technique of *quantile regression* has recently received a lot of attention in both theoretical and empirical research. A number of papers suggest new estimators that deal with various extensions of the original quantile regression model. Other papers deal with practical estimation problems such as the estimation of the covariance matrix for the quantile regression estimates, the performance of the various estimates in small samples, and such like. More impor-

Moshe Buchinsky is a professor of economics at Brown University and a researcher at the National Bureau of Economic Research. *He dedicates this paper to his wife, Claudia. He wishes to thank Don Andrews, Bo Honoré, Ariel Pakes and Jim Powell for many discussions leading to writing this paper, and Richard Blundell and two anonymous referees for their valuable comments and suggestions. He also thanks Phillip Leslie for carefully reading and commenting on an earlier draft of the paper. He acknowledges great benefits over the years from discussions with Gary Chamberlain to whom he wishes to acknowledge special thanks. He takes full responsibility for any remaining errors in this paper. He greatly appreciates financial support from the NSF under grant SBR-9320386.*

tantly, empirical researchers have successfully applied the model, and its various modifications, to a wide range of issues in economics. The goals of this paper are to: (a) summarize some of the most important issues in quantile regression applications; (b) offer some further tools for facilitating various stages of the analysis; and (c) provide a useful guideline for empirical researchers.

The quantile regression model introduced by Koenker and Bassett (1978b), extends the notion of ordinary quantiles in a location model to a more general class of linear models in which the conditional quantiles have a linear form. A well known special case of quantile regression is the least absolute deviation (LAD) estimator (Koenker and Bassett 1978a), which fits medians to a linear function of covariates. LAD estimation is potentially attractive for the same reason that the median may be a better measure of location than the mean.

In an important generalization of the quantile regression model, Powell (1984 and 1986) introduced the censored quantile regression model. This model consistently estimates conditional quantiles when observations on the dependent variable are censored. For example, in the Current Population Survey (CPS) data set many of the variables are top coded for confidentiality.

Useful features of the quantile regression and censored quantile regression models can be summarized as follows: (a) the models can be used to characterize the entire conditional distribution of a dependent variable given a set of regressors; (b) the quantile regression model has a linear programming representation (LP) which makes estimation easy; (c) like the LAD minimand, the quantile regression objective function is a weighted sum of absolute deviations, which gives a robust measure of location, so that the estimated coefficient vector is not sensitive to outlier observations on the dependent variable; (d) when the error term is non-normal, quantile regression estimators may be more efficient than least squares estimators; (e) potentially different solutions at distinct quantiles may be interpreted as differences in the response of the dependent variable to changes in the regressors at various points in the conditional distribution of the dependent variable; (f) L-estimators, based on a linear combination of quantile estimators (for example, Portnoy and Koenker (1989) adaptive L-estimator) are, in general, more efficient than least squares estimators.

The purpose of this paper is to facilitate the practical use of quantile regression models by making recent theoretical developments operationally feasible. Where the literature is ambiguous, I attempt to clarify important ideas. Where there are gaps in the literature, I attempt to fill them. In the process I incorporate results obtained in different studies into the same generalized method of moments (GMM) framework, and provide the various asymptotic results based on this framework. The paper concentrates on cross-section applications, where the observations are assumed to be independently and identically distributed (i.i.d.).[1]

In addition, I discuss an empirical example—the estimation of a log wage regression—using data from the CPS for several representative years. The emphasis in this example is on the analysis of changes in the returns to education at distinct points of the log wage distribution for several age groups. I provide several tests that help characterize the form of dependence between the error term and the re-

1. Relatively little literature considers quantile regressions in the context of time series, for example Weiss (1991) and Bai (1995).

gressors. Finally, several sensitivity analyses are provided, in which I evaluate the performance of various alternative covariance matrix estimators.

The paper is organized as follows. Section II motivates the empirical example provided in this paper and describes the data used. Section III reviews the basic quantile regression results and fits them into the GMM framework. Section IV presents and discusses several alternative estimators for the covariance matrix of the quantile regression estimates. In particular, it provides alternative estimators which are valid under different assumptions on the nature of the dependence between the error term and the regressors. Section V reviews the results for a sequence of quantile regression estimates. Section VI of the paper discusses procedures for testing homoskedasticity and symmetry of the error distribution via the minimum distance (MD) framework. Section VII presents detailed results of the empirical example introduced in Section II, raising several important issues relevant to empirical applications of quantile regression. Section VIII describes an extension of the quantile regression model to the censored quantile regression model. The censoring problems discussed are of a different nature and therefore require different solutions.

II. An Empirical Example—Analysis of Weekly Earnings

A. *Motivation and Background*

It is well known in the labor economic literature that the U.S. wage structure went through enormous changes over the past few decades. One of the most well known observed phenomenon is the increase in wage inequality, even after controlling for individual's characteristics. Another important phenomenon is the increase in the return to skills (namely, education and experience) since the early 1980s.

These two examples indicate that some major changes occurred across the wage distribution. Therefore, it is essential to examine such changes at different points of the distribution. To show the importance of this analysis, Figure 1 depicts the weekly earnings for individuals with 15 years of experience at various quantiles of the earnings distribution.[2] As can be seen from this figure, earnings generally increased at all quantiles as the education level rose. Nevertheless, there are some major differences in the increases at the various quantiles of the distribution across the four years. For example, we can see that there was a much steeper increase in the return to education at the higher quantiles of the distribution in 1992 than in 1979. In fact, very little increases occurred at the lower part of the distribution. Note also that the mean of weekly earnings is consistently above the median of weekly earnings, indicating that the wage distribution is right-skewed, and more so in the latter years.

This example illustrates the potential importance of investigating changes in earnings at different points of the distribution. Clearly, it is not enough to investigate changes in the mean when the entire shape of the distribution changes dramatically. In this paper I provide an empirical example for the estimation of a log wage regres-

2. The exact explanation of the data used to form these graphs and the variables used in the analysis is provided below.

sion.[3] I estimate the regressions at five interesting quantiles of the log wage distribution, namely .10, .25, .50, .75, and .90 quantiles.

Specifically, I investigate the differential changes in the returns to education (or more precisely the derivatives of the conditional quantiles with respect to education) at distinct points of the (log) wage distribution. The years chosen can be viewed as "turning-point" years in the history of changes in the wage structure in the United States. The emphasis of the empirical example given here is on the evolution of the returns to education across these years.

As part of the analysis I introduce a few tests to examine the implications of the results. First, I test for equality among the slope coefficients of the various quantile regressions. Second, as a consequence of the results from of the first test, I estimate a multiplicative heteroskedasticity model, which implies a certain structure on the quantile parameter vectors, and I report the results of that test for the validity of this model.

B. The Data

The extract used for the example presented here is from the March Current Population Survey (March-CPS) for the years 1973, 1980, 1986, and 1993. The extract contains all males between the ages of 18 and 64 who satisfied the following restrictions: (a) they worked at least 13 weeks in the proceding year; (b) they earned at least $50 per week in 1987 prices; (c) they did not attend school; and (d) they were not self-employed. Because the March-CPS variables used here (weeks worked and total earnings) reflect activity in the year preceding the sample year, I refer to the actual years of earnings (namely, 1993 CPS sample is referred to as 1992, etc.). All nominal data are deflated by the implicit price deflator of personal consumption expenditures for gross domestic produce (see The Economic Report of the President, 1995).

Table 1a reports the number of observations used in each of the four sample years, for each of the age groups. As can be seen there are between 5,600 and 19,000 observations for each age group, ranging. This means that the estimation of the coefficients is quite accurate. In addition, deviations among the various estimates for the asymptotic covariance matrix are more likely to indicate the existence of differences in the true covariance matrices rather than differences in the estimates which come from having small samples.

Table 1b reports some basic statistics for the main variables used in the analysis. In general, we see that for the whole sample the education level rises from 12 years (on the average) to 13 years. This increase is not uniform across the various age groups. For example, larger increases are observed for the middle age group than for the younger group. This is due to the fact that more people acquire higher education later on in life or returned to school because of changing returns to education. Note also that more people are living in metropolitan areas (see SMSA) in the later sample years and that people in the southern region have greater representation in the years after 1979. The larger percentage of representation of nonwhite people in

3. For other applications of quantile regressions see Buchinsky (1994), Buchinsky (1995b), Chamberlain (1994), Evans and Schwab (1994), Horowitz and Neumann (1987), and Poterba and Reuben (1994).

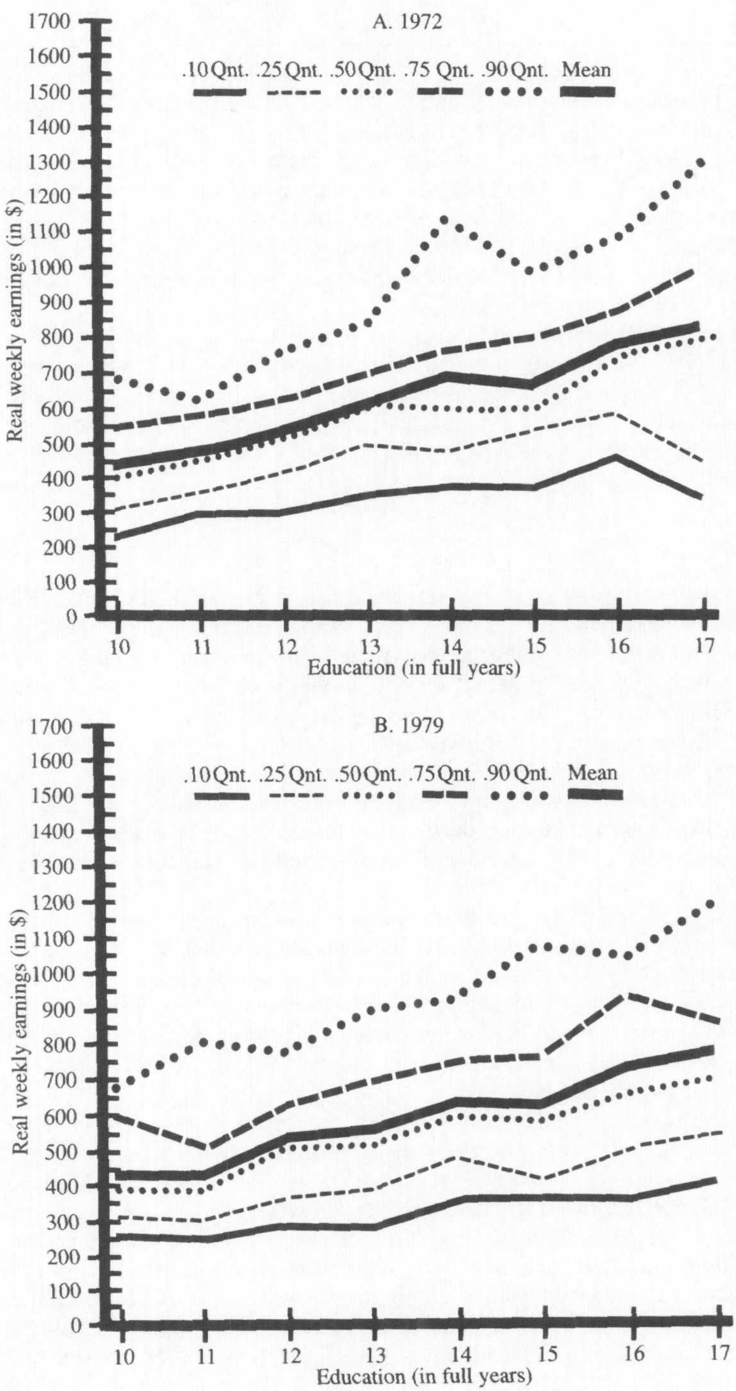
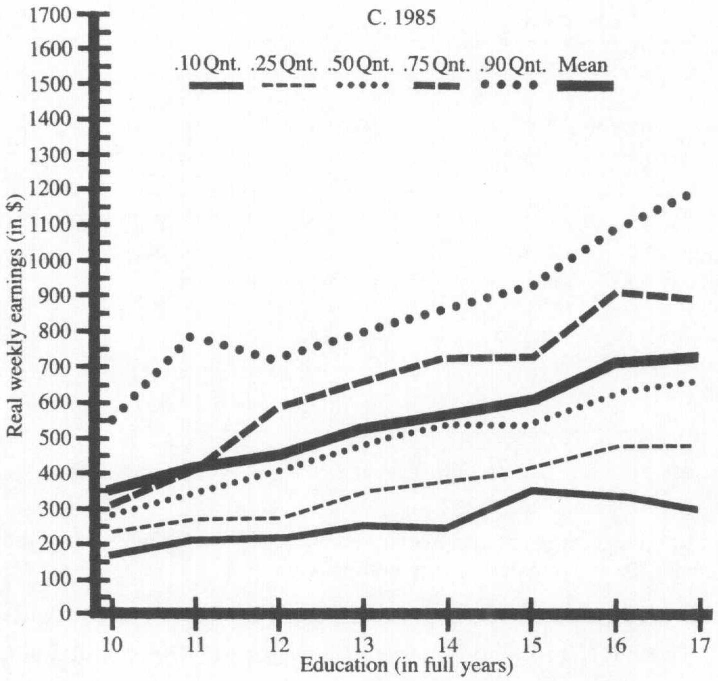


Figure 1

Weekly Earnings for Individuals with 15 Years Experience by Quantile


C. 1985
.10 Qnt. .25 Qnt. .50 Qnt. .75 Qnt. .90 Qnt. Mean
Real weekly earnings (in $)
Education (in full years)

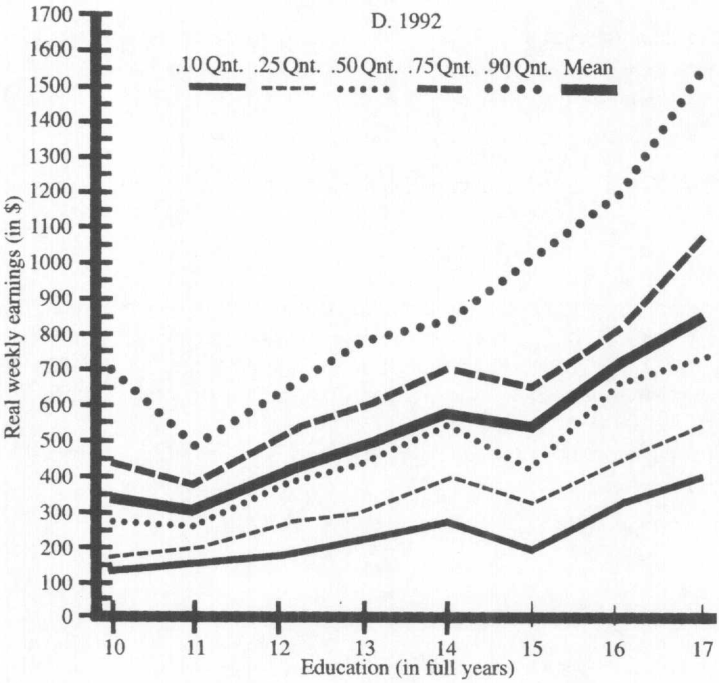

D. 1992
.10 Qnt. .25 Qnt. .50 Qnt. .75 Qnt. .90 Qnt. Mean
Real weekly earnings (in $)
Education (in full years)

Table 1a
Number of Observations

	Year			
Age Group	1972	1979	1985	1992
Total	30,959	37,718	32,748	32,254
18–34	12,913	19,258	15,933	14,325
35–49	10,063	10,834	10,763	12,282
50–64	7,983	7,626	6,052	5,647

Note: The samples include all males between the ages of 18 and 64. For further detail see the text.

the later sample years is because of changes in the CPS sampling method rather than a real change in the population race composition.

III. Quantile Regression—Basic Model and Features

A. *The Model*

The quantile regression model, first introduced by Koenker and Bassett (1978b), can be viewed as a *location model*. Specifically, let (y_i, x_i), $i = 1, \ldots, n$, be a sample from some population, where x_i is a $K \times 1$ vector of regressors. It is assumed that

$$\Pr(y_i \leq \tau | x_i) = F_{u_\theta}(\tau - x_i'\beta_\theta | x_i), \quad i = 1, \ldots, n.$$

This relation—in a different and perhaps more familiar formulation—can be rewritten as

$$(1) \quad y_i = x_i'\beta_\theta + u_{\theta_i}, \quad \text{Quant}_\theta(y_i|x_i) = x_i'\beta_\theta,$$

where $\text{Quant}_\theta(y_i|x_i)$ denotes the conditional quantile of y_i, conditional on the regressor vector x_i.[4] If $F_{u_\theta}(\cdot)$ was known then various techniques could be used to estimate β_θ. However, here the distribution of the error term u_{θ_i} is left unspecified. As is implied by (1), it is only assumed that u_{θ_i} satisfies the quantile restriction $\text{Quant}_\theta(u_{\theta_i}|x_i) = 0$.

In general, the θth sample quantile $(0 < \theta < 1)$ of y, say $\hat{\mu}_\theta$, solves

$$\min_b \left\{ \sum_{i:y_i \geq b} \theta |y_i - b| + \sum_{i:y_i < b} (1 - \theta)|y_i - b| \right\}.$$

4. Note that it is assumed here that both x_i and y_i are observed with no error and that Equation (1) is correctly specified. Problems such as measurement error and omitted variables are not discussed in this paper, and are, by and large, unsolved in the literature. If (1) is not correctly specified (namely, is not linear) then one can view the model as the best linear predictor for the conditional quantile.

The analogue of the linear model for the θth quantile is defined in a similar manner. That is, $\hat{\beta}_\theta$, the estimator for β_θ in (1)—termed the θth quantile regression—solves

$$(2)\quad \min_\beta \frac{1}{n}\left\{\sum_{i:y_i \geq x_i'\beta} \theta|y_i - x_i'\beta| + \sum_{i:y_i<x_i'\beta} (1-\theta)|y_i - x_i'\beta|\right\} = \min_\beta \frac{1}{n}\sum_{i=1}^{n} \rho_\theta(u_{\theta_i}),$$

where $\rho_\theta(\lambda) = (\theta - I(\lambda < 0))\,\lambda$ is the *check function*, and $I(\cdot)$ is the usual indicator function.

The θth quantile regression problem in (2) can be rewritten as

$$(3)\quad \min_\beta \frac{1}{n}\sum_{i=1}^{n} (\theta - 1/2 + 1/2\ \mathrm{sgn}(y_i - x_i'b))(y_i - x_i'b).$$

The $K \times 1$ vector of first-order conditions (F.O.C.) for the problem in (3) is given by[5]

$$(4)\quad \frac{1}{n}\sum_{i=1}^{n} (\theta - 1/2 + 1/2\ \mathrm{sgn}(y_i - x_i'\hat{\beta}_\theta))x_i = 0.$$

In fact, it can be shown that the F.O.C., as specified in (4), implies a moment function which fits into the GMM framework. Define the moment function as

$$(5)\quad \psi(x_i, y_i, \beta) = (\theta - 1/2 + 1/2\ \mathrm{sgn}(y_i - x_i'\beta))x_i.$$

It is straightforward to show that under certain regularity conditions $E[\psi(x_i, y_i, \beta_\theta)] = 0$. This establishes the validity of $\psi(\cdot)$ in (5) as a moment function. The GMM framework can be used, therefore, to establish consistency and asymptotic normality of $\hat{\beta}_\theta$, the estimator of β_θ. Specifically, under certain regularity conditions, it can be shown that[6]

$$\sqrt{n}(\hat{\beta}_\theta - \beta_\theta) \xrightarrow{L} N(0, \Lambda_\theta)$$

where

$$(6)\quad \Lambda_\theta = \theta(1-\theta)(E[f_{u_\theta}(0|x_i)x_ix_i'])^{-1}E[x_ix_i'](E[f_{u_\theta}(0|x_i)x_ix_i'])^{-1}.$$

5. In general these first-order conditions cannot hold exactly. Nevertheless, as $n \to \infty$ the left hand side of this equation converges to 0, or stated more precisely, the left hand side of the equation is $o_p(n^{-1/2})$.

6. The form of Λ_θ is derived in Powell (1984) for the censored quantile regression model. A general derivation of the asymptotic distribution can be obtained using Huber's (1967) framework. Note that here

$$E[f_{u_\theta}(0|x_i)x_ix_i'] = \partial E[\psi(x_i, y_i, \beta_\theta)]/\partial\beta_\theta',$$

and

$$\theta(1-\theta)E[x_ix_i'] = E[\psi(x_i, y_i, \beta_\theta)\psi(x_i, y_i, \beta_\theta)'].$$

Table 1b
Raw Statistics for Main Variables

	All			18–34			35–49			50–64		
Variable	Mean	Std	Med	Mean	Std	Med	Mean	Std	Med	Mean	Std	Med
1972												
Log wage	6.09	0.63	6.16	5.88	0.62	5.98	6.28	0.57	6.30	6.19	0.62	6.21
Education	12.0	3.1	12.0	12.5	2.6	12.0	12.0	3.3	12.0	11.1	3.4	12.0
Experience	21.1	13.8	20.0	7.6	5.1	7.0	24.2	5.6	24.0	38.9	5.6	39.0
Education × experience	2.36	1.54	2.24	0.89	0.56	0.85	2.79	0.68	2.76	4.18	1.04	4.20
Dummies												
North east	0.24	0.43	0.00	0.23	0.42	0.00	0.24	0.43	0.00	0.26	0.44	0.00
North central	0.29	0.45	0.00	0.29	0.45	0.00	0.28	0.45	0.00	0.29	0.45	0.00
South	0.17	0.38	0.00	0.17	0.38	0.00	0.17	0.38	0.00	0.17	0.38	0.00
SMSA	0.28	0.45	0.00	0.29	0.45	0.00	0.28	0.45	0.00	0.29	0.45	0.00
Part-time	0.16	0.37	0.00	0.20	0.40	0.00	0.11	0.31	0.00	0.17	0.38	0.00
Part-year 1	0.06	0.24	0.00	0.09	0.29	0.00	0.04	0.19	0.00	0.05	0.21	0.00
Part-year 2	0.06	0.24	0.00	0.10	0.30	0.00	0.03	0.17	0.00	0.04	0.20	0.00
Race	0.08	0.28	0.00	0.09	0.28	0.00	0.09	0.29	0.00	0.07	0.26	0.00
1979												
Log wage	6.04	0.61	6.11	5.85	0.62	5.93	6.27	0.52	6.33	6.22	0.57	6.28
Education	12.5	2.9	12.0	12.8	2.5	12.0	12.6	3.2	12.0	11.7	3.4	12.0
Experience	18.0	13.3	15.0	7.3	4.7	7.0	22.9	5.6	23.0	38.4	5.5	38.0
Education × experience	2.14	1.55	1.80	0.90	0.56	0.84	2.77	0.67	2.70	4.37	1.04	4.44
Dummies												
North east	0.20	0.40	0.00	0.19	0.39	0.00	0.21	0.41	0.00	0.24	0.42	0.00
North central	0.24	0.43	0.00	0.25	0.43	0.00	0.23	0.42	0.00	0.25	0.43	0.00
South	0.28	0.45	0.00	0.28	0.45	0.00	0.30	0.46	0.00	0.28	0.45	0.00
SMSA	0.57	0.50	1.00	0.56	0.50	1.00	0.56	0.50	1.00	0.59	0.49	1.00
Part-time	0.04	0.20	0.00	0.07	0.25	0.00	0.01	0.10	0.00	0.02	0.14	0.00
Part-year 1	0.07	0.25	0.00	0.09	0.29	0.00	0.04	0.20	0.00	0.05	0.22	0.00
Part-year 2	0.06	0.24	0.00	0.09	0.29	0.00	0.03	0.17	0.00	0.04	0.19	0.00
Race	0.10	0.30	0.00	0.10	0.30	0.00	0.11	0.31	0.00	0.09	0.29	0.00

1985												
Log wage	6.01	0.68	6.07	5.76	0.67	5.83	6.26	0.58	6.32	6.21	0.63	6.25
Education	12.9	2.8	12.0	12.9	2.4	12.0	13.3	3.0	12.0	12.2	3.3	12.0
Experience	18.0	12.4	16.0	7.8	4.7	8.0	21.9	5.4	21.0	37.9	5.4	38.0
Education × experience	2.23	1.52	1.96	0.97	0.56	0.96	2.80	0.64	2.72	4.50	1.02	4.56
Dummies												
North east	0.24	0.43	0.00	0.23	0.42	0.00	0.24	0.43	0.00	0.28	0.45	0.00
North central	0.24	0.43	0.00	0.24	0.43	0.00	0.23	0.42	0.00	0.24	0.43	0.00
South	0.29	0.46	0.00	0.29	0.46	0.00	0.30	0.46	0.00	0.29	0.45	0.00
SMSA	0.76	0.43	1.00	0.76	0.43	1.00	0.76	0.43	1.00	0.76	0.43	1.00
Part-time	0.05	0.22	0.00	0.08	0.28	0.00	0.01	0.12	0.00	0.03	0.17	0.00
Part-year 1	0.07	0.25	0.00	0.09	0.28	0.00	0.05	0.21	0.00	0.05	0.21	0.00
Part-year 2	0.07	0.25	0.00	0.09	0.29	0.00	0.04	0.19	0.00	0.05	0.22	0.00
Race	0.11	0.32	0.00	0.11	0.32	0.00	0.12	0.32	0.00	0.10	0.30	0.00
1992												
Log wage	5.95	0.69	6.00	5.66	0.67	5.70	6.18	0.60	6.21	6.17	0.66	6.24
Education	13.0	2.7	13.0	12.8	2.4	12.0	13.3	2.7	13.0	12.7	3.1	12.0
Experience	18.6	11.6	17.0	8.3	4.8	8.0	22.0	5.0	22.0	37.0	5.3	36.5
Education × experience	2.35	1.50	2.16	1.02	0.56	1.04	2.87	0.66	2.86	4.59	0.99	4.62
Dummies												
North east	0.23	0.42	0.00	0.22	0.42	0.00	0.23	0.42	0.00	0.25	0.43	0.00
North central	0.24	0.43	0.00	0.24	0.43	0.00	0.24	0.43	0.00	0.25	0.43	0.00
South	0.29	0.46	0.00	0.29	0.45	0.00	0.30	0.46	0.00	0.29	0.45	0.00
SMSA	0.76	0.43	1.00	0.77	0.42	1.00	0.75	0.44	1.00	0.76	0.43	1.00
Part-time	0.05	0.22	0.00	0.09	0.28	0.00	0.02	0.13	0.00	0.04	0.19	0.00
Part-year 1	0.06	0.24	0.00	0.08	0.27	0.00	0.04	0.20	0.00	0.05	0.22	0.00
Part-year 2	0.07	0.25	0.00	0.09	0.29	0.00	0.04	0.20	0.00	0.05	0.23	0.00
Race	0.13	0.33	0.00	0.13	0.34	0.00	0.12	0.33	0.00	0.11	0.32	0.00

Note: The samples include all males between the ages of 18 and 64. For further details see the data description section.

If $f_{u_\theta}(0|x) = f_{u_\theta}(0)$ with probability 1 (namely, the density of the error term u_θ evaluated at 0 is independent of x), then Λ_θ in (6) simplifies to[7]

$$(7)\quad \Lambda_\theta = \frac{\theta(1-\theta)}{f^2_{u_\theta}(0)}(E[x_i x_i'])^{-1}.$$

1. Interpretation of Quantile Regression Estimation

Note that since the θth conditional quantile of y given x is given by $\text{Quant}_\theta(y_i|x_i) = x_i'\beta_\theta$, its estimate is given by $\widehat{\text{Quant}_\theta}(y_i|x_i) = x_i'\hat{\beta}_\theta$. As one increases θ continuously from 0 to 1, one traces the entire conditional distribution of y, conditional on x. In practice, given that any data set contains only a finite number of observations, only a finite number of quantile estimates will be numerically distinct, although this number can be quite large. Note also that the various quantile regression estimates are correlated. I discuss below the joint asymptotic distribution of these quantile regression estimates.

How can the quantile's coefficients be interpreted? Consider the partial derivative of the conditional quantile of y with respect to one of the regressors, say j, namely, $\partial\text{Quant}_\theta(y_i|x_i)/\partial x_{ij}$. This derivative is to be interpreted as the marginal change in the θth conditional quantile due to marginal change in the jth element of x. If x contains K distinct variables, then this derivative is given simply by β_{θ_j}, the coefficient on the jth variable. One should be cautious in interpreting this result. It does not imply that a person who happens to be in the θth quantile of one conditional distribution will also find himself/herself at the same quantile had his/her x changed.

2. Efficient Estimation

The quantile regression estimator described above is not an efficient estimator for β_θ. An efficient estimator can be obtained by solving

$$\min_\beta \frac{1}{n}\sum_{i=1}^{n} f_{u_\theta}(0|x)(\theta - 1/2 + 1/2\,\text{sgn}(y_i - x_i'\beta))(y_i - x_i'\beta).$$

However, this estimation procedure requires the use of an estimate for the unknown density $f_{u_\theta}(0|x)$. For more details see Newey and Powell (1990).

3. Equivariance Properties

The quantile regression estimator has several important equivariance properties which help facilitate the computation procedure. Denote the set of feasible solutions to the problem defined in (3) by $B(\theta, y, X)$. Then for every $\hat{\beta}_\theta \equiv \hat{\beta}(\theta, y, X) \in B(\theta, y, X)$ we have (Koenker and Bassett 1978b, Theorem 3.2):

$$(8)\qquad \hat{\beta}(\theta, \lambda y, X) = \lambda\hat{\beta}(\theta, y, X), \qquad \lambda \in [0, \infty),$$

7. This is exactly the result obtained by Koenker and Bassett (1978b). In general, if $f_{u_\theta}(\cdot|x)$ is independent of x, then all quantiles should have parameter vectors that differ only in their intercepts. A test of equality among the slope parameters is discussed in Section V.

(9) $\hat{\beta}(1-\theta, \lambda y, X) = \lambda\hat{\beta}(\theta, y, X), \qquad \lambda \in (-\infty, 0],$

(10) $\hat{\beta}(\theta, y + X\gamma, X) = \hat{\beta}(\theta, y, X) + \gamma, \quad \gamma \in \Re^K,$

(11) $\hat{\beta}(\theta, y, XA) = A^{-1}\hat{\beta}(\theta, y, X), \qquad A_{K\times K}$ is nonsingular.

Because, as is shown below, the quantile regression is a linear programming problem, manipulation based on (8)–(11) can be used to significantly reduce the number of simplex iterations.

B. Linear Programming Representation of Quantile Regression

The problem in (2) can be shown to have a *linear programming* (LP) representation. This feature has important consequences from both theoretical and practical standpoints.

To see that the problem (2) is an LP problem note that y_i can be rewritten as a function of only positive elements:

$$y_i = \sum_{j=1}^{K} x_{ij}\beta_{\theta_j} + u_{\theta_i} = \sum_{j=1}^{K} x_{ij}(\beta^1_{\theta_j} - \beta^2_{\theta_j}) + (\varepsilon_{\theta_i} - v_{\theta_i}),$$

where $\beta^1_{\theta_j} \geq 0$, $\beta^2_{\theta_j} \geq 0$ $(j = 1, \ldots, K)$, and $\varepsilon_{\theta_i} \geq 0$, $v_{\theta_i} \geq 0$ $(i = 1, \ldots, n)$. When written in matrix notation the problem in (2) takes the familiar *primal problem* of LP (for example, Franklin 1980):

$$\min_z c'z \quad \text{subject to:} \quad Az = y,\ z \geq 0.$$

where $A = (X, -X, I_n, -I_n)$, $y = (y_1, \ldots, y_n)'$, $z = (\beta^{1'}, \beta^{2'}, u', v')'$, $c = (0', 0', \theta \cdot l', (1-\theta) \cdot l')'$, $X = (x_1, \ldots, x_n)'$, I_n is an n dimensional identity matrix, $0'$ is a $K \times 1$ vector of zeros, and l is an $n \times 1$ vector of ones. Furthermore, the *dual problem*, of LP is (approximately) the same as the F.O.C. specified in (4) and is given by

$$\max_w w'y \quad \text{subject to:} \quad w'A \leq c'.$$

The *duality theorem* implies that feasible solutions exist for both the primal and the dual problems, if the design matrix X is of full column rank. The *equilibrium theorem* of LP guarantees then that this solution is optimal.

The LP representation of the quantile regression problem has several important implications from both theoretical and practical standpoints. First, it is guaranteed that the quantile regression estimate will be obtained in a finite number of simplex iterations.[8] Second, unlike the case of the mean type regression, the parameter vector estimate is robust to outliers. That is, if $y_i - x_i'\hat{\beta}_\theta > 0$, then y_i can be increased toward ∞, or if $y_i - x_i'\hat{\beta}_\theta < 0$, y_i can be decreased toward $-\infty$, without altering the solution $\hat{\beta}_\theta$. In other words, it is not the magnitude of the dependent variable that matters but on which side of the estimated hyperplane is the observation.

8. Typically, the number of iterations required is relatively small for an efficient LP algorithm such as that proposed by Barrodale and Roberts (1973).

In many economic situations the dependent variable is censored (at some known value). The implication of the above property is that if, for example, an observation is right-censored, then for as long as it has a positive residual one need not know the true value of the dependent variable.[9]

1. Computation Algorithm

Several algorithms have been proposed in the literture for solving the LP problem, the most attractive of which seems to be Barrodale and Roberts' (1973) algorithm.[10] The main advantage of the Barrodale-Roberts algorithm is that it significantly reduces the number of simplex transformations relative to the other known algorithms. The computation time can be further reduced by modifying the LP algorithm to take advantage of the equivariant property (10). If a good initial value of $\hat{\beta}_\theta$, say $\hat{\beta}^0_\theta$, is known in advance, it can be used to speed up the computational time, since it can place the observations at the "right" side of the hyperplane. Let $y_R = y - X\hat{\beta}^0_\theta$ and let $\hat{\beta}^R_\theta$ be the estimate from a quantile regression of y_R on X, then by property (10), $\hat{\beta}_\theta = \hat{\beta}^R_\theta + \hat{\beta}^0_\theta$. Obtaining $\hat{\beta}^R_\theta$ and $\hat{\beta}^0_\theta$ can be much faster than obtaining $\hat{\beta}_\theta$ directly.

One possible initial value is an adjusted least squares (LS) estimate, where the constant is replaced by the $[n\theta]$th order statistic of the residuals $\hat{e}_1, \ldots, \hat{e}_n$. An alternative initial value can be obtained from a quantile regression based on only a small fraction of the sample; this can be especially useful when the data set is very large.

IV. Asymptotic Covariance Matrix Estimation

Equations (6) and (7) give formulas for the asymptotic covariance matrix for $\hat{\beta}_\theta$ under two alternative assumptions about $f_{u_\theta}(0|x)$. Problems in estimating the covariance matrix in (6) arise mainly with regard to $f_{u_\theta}(0|x)$, or alternatively $E[f_{u_\theta}(0|x)xx']$. Consequently, estimation of Λ_θ depends on the subjective decision of the econometrician, and there is no decisively "best" path to follow. Each approach has some advantages and disadvantages and each entails some degree of arbitrariness, the effects of which can be quite important. In this section I detail several estimators for Λ_θ and then in the empirical example in Section VII, I provide a comparison of the various methods.[11]

9. In cases where some of the observations which are right-censored have negative residuals, or some of the observations which are left-censored have positive residuals, the quantile regression estimate is biased. In this case one has to resort to the censored quantile regression model as is discussed later.

10. Barrodale and Roberts' algorithm is for the median regression, namely, $\theta = 1/2$, but, with a minor modification, can be adapted for any quantile regression (for example, Koenker and D'Orey 1987). Other algorithms include Barrodale (1968), Robers and Ben-Israel (1969), Robers and Robers (1970), Abdelmalek (1971), and Armstrong, Frome and Kung (1979).

11. For a comprehensive discussion and examination of the various estimators for the asymptotic covariance matrix see Buchinsky (1995a). In summary, the design matrix bootstrap performs the best, but it is also the most computer intensive method.

A. *Order Statistic Estimator*

This estimator is valid when $f_{u_\theta}(0|x) = f_{u_\theta}(0)$, that is, when the independence assumption holds.[12] Recall that under this assumption the asymptotic covariance matrix simplifies to

$$\Lambda_\theta = \sigma_\theta^2(E[xx'])^{-1},$$

where

$$\sigma_\theta^2 = \frac{\theta(1-\theta)}{f_{u_\theta}^2(0)}.$$

The second term of this asymptotic variance is easily estimated by $\hat{E}(xx') = 1/n \sum_{i=1}^n x_i x_i'$. The first term, σ_θ^2, can be extracted from a confidence interval constructed from the $[n\theta]$th order statistic of $\hat{u}_{\theta_1}, \ldots, \hat{u}_{\theta_n}$. In general, an exact confidence interval can be computed for the θth quantile of a random variable $Y \sim F_Y(\cdot)$ (for example, Mood, Graybill, and Boes 1974). Specifically,

$$(12)\quad \Pr(y_{(j)} \leq \xi_\theta \leq y_{(k)}) = \Pr(y_{(j)} \leq \xi_\theta) - \Pr(y_{(k)} < \xi_\theta),$$

where $y_{(j)}$ and $y_{(k)}$ are the jth and kth order statistics of $y_1, \ldots, y_n$, respectively. Note that

$$(13)\quad \Pr(y_{(j)} \leq \xi_\theta) = \Pr(j \text{ or more observations} \leq \xi_\theta) = \sum_{i=j}^{n} \binom{n}{i} \theta^i (1-\theta)^{n-i}.$$

Similarly,

$$(14)\quad \Pr(y_{(k)} < \xi_\theta) = \sum_{i=k}^{n} \binom{n}{i} \theta^i (1-\theta)^{n-i}.$$

Substituting (13) and (14) into (12) yields

$$(15)\quad \Pr(y_{(j)} \leq \xi_\theta \leq y_{(k)}) = \sum_{i=j}^{k-1} \binom{n}{i} \theta^i (1-\theta)^{n-i}.$$

Constructing a symmetric confidence interval of level $1 - \alpha$ for ξ_θ is straightforward. Let $j = [n\theta - l]$, $k = [n\theta + l]$, and let $X \sim B(n, \theta)$. Then,

$$(16)\quad \Pr(y_{([n\theta - l])} \leq \xi_\theta < y_{([n\theta + l])}) = \Pr([n\theta - l] \leq X < [n\theta + l])$$

$$\approx \Pr\left(\left|\frac{X - n\theta}{\sqrt{n\theta(1-\theta)}}\right| \leq \frac{l}{\sqrt{n\theta(1-\theta)}}\right).$$

Because $(X - n\theta)/\sqrt{n\theta(1-\theta)} \xrightarrow{L} N(0, 1)$, equating the probability in (16) to $1 - \alpha$ gives $l = Z_{1-\alpha/2}\sqrt{n\theta(1-\theta)}$. Matching the length of the exact confidence interval

12. This estimator was suggested by Chamberlain (1994). Similar computation techniques, using an exact confidence interval for order statistics, can be found in Efron (1982) and Huber (1981).

in (15) with that of the asymptotic normal confidence interval yields an estimate for σ_θ^2:

$$(17)\quad \hat{\sigma}_\theta^2 = \frac{n(y_{([n\theta+l])} - y_{([n\theta-l])})^2}{4Z_{1-\alpha/2}^2}.$$

Essentially what we implicitly obtained is an estimator for $f_{u_\theta}(0)$ given by $\hat{f}_{u_\theta}(0) = 2Z_{1-\alpha/2}\sqrt{\theta(1-\theta)}/\sqrt{n}(y_{([n\theta+l])} - y_{([n\theta-l])})$. This estimate can be used when constructing an estimate for the covariance matrix of a sequence of quantile regressions as discussed below.

B. Bootstrap Estimators

There are two alternative ways to employ the bootstrap method proposed by Efron (1979), based on fundamentally different assumptions about the form of the asymptotic covariance matrix of $\hat{\beta}_\theta$. While one technique (*Design Matrix Bootstrap Estimator*) provides a consistent estimator of the asymptotic matrix under more general conditions, the other (*Error Bootstrap Estimator*) yields a consistent estimator only under the idependence assumption. A third bootstrap estimator is the *Sigma Estimator*. In this estimator only part of the covariance matrix is estimated using the bootstrap technique, namely $\hat{\sigma}_\theta^2$. Bootstrap samples for the design matrix bootstrap estimator are drawn from the empirical joint distribution of x and y, F_{nxy}; and for the error bootstrap estimator from the empirical distribution of x, F_{nx}, and u_θ, $F_{n\hat{u}_\theta}$.[13]

1. Design Matrix Bootstrapping Estimator

Let (y_i^*, x_i^*), $i = 1, \ldots, n$, be a randomly drawn sample from the empirical distribution F_{nxy}. It follows from the model in (1) that $y^* = X^*\beta_\theta + u_{\theta_i}^*$, where $y^* = (y_1^*, \ldots, y_n^*)'$ and $X^* = (x_1^*, \ldots, x_n^*)'$. Let $\hat{\beta}_\theta^*$ denote the bootstrap estimate obtained from a quantile regression of y^* on X^*. This process can be repeated B times, to yield bootstrap estimates $\hat{\beta}_{\theta 1}^*, \ldots, \hat{\beta}_{\theta B}^*$. The bootstrap estimator of Λ_θ is given then by

$$(18)\quad \hat{\Lambda}_\theta^{DMB} = \frac{n}{B}\sum_{j=1}^{B}(\hat{\beta}_{\theta j}^* - \bar{\beta}_\theta^*)(\hat{\beta}_{\theta j}^* - \bar{\beta}_\theta^*)',$$

where $\bar{\beta}_\theta^* = 1/B\,\Sigma_{j=1}^{B}\,\hat{\beta}_{\theta j}^*$. Alternatively, one can use $\hat{\beta}_\theta$ as the pivotal value instead of $\bar{\beta}_\theta^*$.

This is a consistent estimator of the asymptotic covariance of $\hat{\beta}_\theta$ given in (6) in the sense that the conditional distribution of $\sqrt{n}(\hat{\beta}_\theta^* - \hat{\beta}_\theta)$ weakly converges to the unconditional distribution of $\sqrt{n}(\hat{\beta}_\theta - \beta_\theta)$.[14]

13. Note that the empirical distribution $F_{n\hat{u}_\theta}$ is based on $u_{\theta i} = y_i - x_i'\hat{\beta}_\theta$, $i = 1, \ldots, n$, and not $\hat{u}_{\theta i}$, $i = 1, \ldots, n$, which are not observed.

14. For a general derivation of this results see Bickel and Freedman (1981). For an application of this result in a mean regression model see Freedman (1981).

2. Error Bootstrap Estimator

Under the independence assumption it is possible to perform the bootstrap estimation procedure by resampling from the marginal empirical distributions F_{nx} and $F_{n\hat{u}_\theta}$. Let $u_\theta^* = (u_{\theta_1}^*, \ldots, u_{\theta_n}^*)'$ be a randomly drawn sample of size n from the empirical distribution $F_{n\hat{u}_\theta}$ and let $X^* = (x_1^*, \ldots, x_n^*)'$ be a randomly drawn sample from the empirical distribution F_{nx}. Define $y^* = X^*\hat{\beta}_\theta + u_\theta^*$. This starred data is then used to solve the quantile regression problem, the solution of which is a bootstrap estimator, say $\hat{\beta}_\theta^*$. As with the design matrix bootstrapping procedure, this is repeated B times, to yield B bootstrap estimators $\hat{\beta}_{\theta j}^*$ ($j = 1, \ldots, B$). The estimator of Λ_θ, as the asymptotic covariance matrix of $\hat{\beta}_\theta$, is then obtained in a fashion identical to that of the design matrix bootstrap method given in (18).

This is a consistent estimator of Λ_θ only under the independence assumption. If the independence assumption does not hold, the resampling scheme destroys any relationship that might exist between u_θ and x, making this method invalid. As is clear from the discussion above, there is no advantage to using an error bootstrap estimator over the design matrix estimator since both require the same computation time and both hold under independence. In small samples, however, the performance of the error bootstrap estimator might be better if the independence assumption is satisfied.

3. Percentile Estimation

The percentile method is an alternative way of using the information from any single bootstrap method. One can extract the variance of an estimated coefficient from a confidence interval directly computed for that coefficient. Any one of the bootstrap methods provides a sequence of bootstrap estimates $\hat{\beta}_{\theta 1}^*, \ldots, \hat{\beta}_{\theta B}^*$. Let $\hat{\beta}_{\theta(1)}^*, \ldots, \hat{\beta}_{\theta(B)}^*$ be the ordered sequence of $\hat{\beta}_{\theta 1}^*, \ldots, \hat{\beta}_{\theta B}^*$, element by element. Then, $1 - \alpha$ confidence interval for the kth coefficient β_θ^k ($k = 1, \ldots, K$), can be constructed as follows: Let L_B^k and U_B^k be the $[B\alpha/2]$th and $[B(1 - \alpha/2)]$th order statistics of the kth element of $\hat{\beta}_{\theta 1}^*, \ldots, \hat{\beta}_{\theta B}^*$, respectively, where $[\lambda]$ denotes the largest integer greater than or equal to λ. Note that

$$\Pr(L_B^k - \hat{\beta}_\theta^k \leq \hat{\beta}_\theta^{*k} - \hat{\beta}_\theta^k \leq U_B^k - \hat{\beta}_\theta^k) \approx \Pr(L_B^k - \hat{\beta}_\theta^k \leq \hat{\beta}_\theta^k - \beta_\theta^k \leq U_B^k - \hat{\beta}_\theta^k) \approx 1 - \alpha,$$

where $k = 1, \ldots, K$. Therefore, an (approximate) $1 - \alpha$ confidence interval for β_θ^k is given by

$$(19) \quad [2\hat{\beta}_\theta^k - U_B^k, 2\hat{\beta}_\theta^k - L_B^k].$$

Since $\hat{\beta}_\theta^k$ has an asymptotic normal distribution, matching the length of the asymptotic confidence interval with the length of the confidence interval in (19) gives the asymptotic variance for $\hat{\beta}_\theta^k$:

$$(20) \quad \hat{\sigma}^2_{\hat{\beta}_\theta^{(K)}} = \frac{n(U_B^k - L_B^k)^2}{4Z_{1-\alpha/2}^2}, \; k = 1, \ldots, K.$$

Note that this particular example provides a confidence intervals for each parameter in β_θ. It can be used, however, to provide a confidence interval for any statistic of interest that is based on a combination of the quantile regression coefficients.

4. Sigma Bootstrap Estimator

This estimator also relies on the independence assumption and on the form of the asymptotic covariance matrix given in (7). More precisely, the technique combines the special representation of Λ_θ with a nonparametric estimation of σ_θ^2, by employing the bootstrap method. One can obtain B bootstrap estimates for q_θ, the θth quantile of u_θ, $\hat{q}_{\theta 1}^*, \ldots, \hat{q}_{\theta B}^*$, from B bootstrap samples—each of size n—drawn from the empirical distribution $F_{n\hat{u}_\theta}$. An estimator for σ_θ^2 is obtained then by

$$\hat{\sigma}_{\theta_B}^2 = \frac{n}{B} \sum_{j=1}^{B} (q_{\theta j}^* - \hat{q}_\theta^*)^2,$$

where $\hat{q}_\theta^* = 1/B \sum_{j=1}^{B} \hat{q}_{\theta j}^*$.

C. Kernel Estimator

Powell (1986) considered a kernel estimator for $E[f_{u_\theta}(0|x)xx']$. This estimator takes the form

$$(21) \quad \hat{E}(f_{u_\theta}(0|x)xx') = (c_n n)^{-1} \sum_{i=1}^{n} k(\hat{u}_{\theta_i}) x_i x_i',$$

where $k(\cdot)$ is some kernel function and $c_n = o_p(1)$ is the kernel bandwidth.[15] The term $E(xx')$ is estimated as before by $\hat{E}(xx') = \sum_{i=1}^{n} x_i x_i'/n$.

A practical problem arises regarding the choice of the kernel bandwidth that would preferably be data dependent. No direct method (that I am aware of) allows one to optimally choose c_n. Nevertheless, note that the top left-hand element of the matrix in (21) is an estimate of the density $f_{u_\theta}(0)$:

$$\hat{f}_{u_\theta}(0) = (c_n n)^{-1} \sum_{i=1}^{n} k(\hat{u}_{\theta_i}).$$

Fortunately, for this particular problem, there are a number of cross-validation methods for optimally choosing c_n (for example, least-squares, log likelihood, and so on). Therefore, one can first determine c_n via a cross validation technique and then use the optimally chosen c_n to estimate $E(f_{u_\theta}(0|x)xx')$ in (21).

Note that under the independence assumption it is only necessary to estimate $f_{u_\theta}(0)$, in addition to $E(xx')$. This is immediately available from the first step described above. The estimate of the covariance matrix Λ_θ is then given by

$$(22) \quad \hat{\Lambda}_\theta = \frac{\theta(1-\theta)}{\hat{f}_{u_\theta}^2(0)} \left(\frac{1}{n} \sum_{i=1}^{n} x_i x_i' \right)^{-1}.$$

15. Powell (1986) considered a one-side kernel function due to the problem of censoring. In general, one can use a two-side kernel function.

V. Sequence of Quantile Regressions

So far I have discussed the estimation of a single quantile regression for a specific value of θ. In practice one would like to estimate several quantile regressions at distinct points of the conditional distribution of the dependent variable. Because these quantile regressions are estimated using the same data with different weighting schemes, they ought to be correlated. This section outlines the estimation of a finite sequence of quantile regressions and provides its asymptotic distribution.

Consider the model given by (2) (dropping the i subscript for simplicity) for p alternative θ's,

$$y = x'\beta_{\theta_j} + u_{\theta_j} \quad \text{and} \quad \text{Quant}_{\theta_j}(u_{\theta_j}|x) = 0, \quad j = 1, \ldots, p.$$

Without loss of generality assume that $0 < \theta_1 < \theta_2 < \cdots < \theta_p < 1$. Let

$$(23) \quad \psi(x, y, \beta_1, \ldots, \beta_P)' = (\psi_1(x, y, \beta_1)', \ldots, \psi_p(x, y, \beta_p)'),$$

where

$$\psi_j(x, y, \beta) = (\theta_j - 1/2 + 1/2 \,\text{sgn}(y - x'\beta))x, \quad j = 1, \ldots, p$$

define the p moment functions for the θ_1th through θ_pth quantile regressions, respectively. Let $\beta' = (\beta_1', \ldots, \beta_p')$ and let $\beta_\theta' = (\beta_{\theta_1}', \ldots, \beta_{\theta_p}')$ be the population's true values.

Under some regularity conditions, $E[\psi(x, y, \beta_{\theta_1}, \ldots, \beta_{\theta_p})] = 0$. From the analogy principle (for example, Manski 1988) of estimation the estimator $\hat{\beta}_\theta$ for β_θ is obtained as a solution to

$$\frac{1}{n}\sum_{i=1}^{n} \psi(x_i, y_i, \hat{\beta}_{\theta_1}, \ldots, \hat{\beta}_{\theta_p}) = 0.$$

However, one need not solve for $\hat{\beta}_{\theta j}$ ($j = 1, \ldots, p$) simultaneously. In fact the estimation amounts to solving the problem in (3) for each quantile separately, as there are no cross-restrictions imposed on $\hat{\beta}_{\theta_1}, \ldots, \hat{\beta}_{\theta P}$.

Under some regularity conditions (see Powell 1984) $\hat{\beta}_\theta$ has an asymptotic normal distribution, that is, $\sqrt{n}(\hat{\beta}_\theta - \beta_\theta) \xrightarrow{L} N(0, \Lambda_\theta)$, where $\Lambda_\theta = \{\Lambda_{\theta_{jk}}\}_{j,k=1,\ldots,p}$ and

$$(24) \quad \Lambda_{\theta_{jk}} = (\min\{\theta_j, \theta_k\} - \theta_j\theta_k)(E[f_{u_{\theta_j}}(0|x)xx'])^{-1} E[xx'](E[f_{u_{\theta_k}}(0|x)xx'])^{-1}.$$

Again, note that if $f_{u_{\theta_j}}(0|x) = f_{u_{\theta_j}}(0)$ ($j = 1, \ldots, p$), then (24) simplifies to

$$(25) \quad \Lambda_\theta = \Omega_\theta \otimes (E[xx'])^{-1},$$

where (see Koenker and Bassett (1978b), Theorem 4.2) $\Omega_\theta = \{\Omega_{\theta jk}\}_{j,k=1,\ldots,p}$ and

$$\Omega_{\theta_{jk}} = \frac{\min\{\theta_j, \theta_k\} - \theta_j\theta_k}{f_{u_{\theta_j}}(0) f_{u_{\theta_k}}(0)}.$$

Note that the estimated conditional quantiles, conditional on x, are given by $x'\hat{\beta}_{\theta_1}, \ldots, x'\hat{\beta}_{\theta_p}$. Since the estimates $\hat{\beta}_{\theta_j}$ ($j = 1, \ldots, p$) for the p quantiles are obtained separately at every quantile, it need not be the case that $x'\hat{\beta}_{\theta_j} > x'\hat{\beta}_{\theta_k}$ if θ_j

$> \theta_k$. In fact, one can always find a vector x_0 such that $x_0'\hat{\beta}_{\theta_j} > x_0'\hat{\beta}_{\theta_k}$, namely, the conditional quantiles cross each other. This may not be of any practical consequence, because there may not be such a vector within the relevant range of plausible x's.

VI. Tests for Homoskedasticity and Symmetry

In the previous sections the asymptotic properties of the regression quantiles were presented. In this section several practical questions pertaining to these estimators are discussed. In particular, I consider tests that can be performed having a sequence of quantile estimates at hand. Note that if $f_{u_\theta}(0|x) = f_{u_\theta}(0)$ then any two quantile parameter vectors, β_{θ_1} and β_{θ_2}, should differ only in their intercepts but not in their slope coefficients. If the distribution of u_θ was symmetric, an alternative structure is implied for the coefficient vectors. This section examines tests for equality of the slope coefficients and symmetry using the minimum distance (MD) framework.

A. *Test for Homoskedasticity*

The main difference between the technique used here and that followed by Koenker and Bassett (1982) is that the covariance matrix for the test statistic used here is valid under non-local alternative hypotheses, while theirs is valid only under local alternatives.[16]

In the MD framework followed here, first the slope coefficients are estimated under the restrictions implied by homoskedasticity. That is, the restricted coefficient vector minimizes, with respect to β^R,

$$(26) \quad Q(\beta^R) = (\hat{\beta}_\theta - R\beta^R)'A^{-1}(\hat{\beta}_\theta - R\beta^R),$$

where A is a weight matrix with $A \rightarrow \Psi$, a positive definite matrix and $\hat{\beta}_\theta = (\hat{\beta}_{\theta_1}', \ldots, \hat{\beta}_{\theta_p}')'$ is the unrestricted vector of p quantile regression estimates.[17] Note that $\beta_\theta^R = (\beta_{\theta_1 1}, \ldots, \beta_{\theta_p 1}, \beta_2, \ldots, \beta_k)'$ is a $(p + K - 1) \times 1$ vector of restricted parameters. The *restriction matrix* R is given by

$$R' = (R_1, \ldots, R_p) \quad \text{where} \quad R_j = \begin{pmatrix} e_j & 0_m \\ 0_v & I_{K-1} \end{pmatrix},$$

e_j is a $p \times 1$ vector of zeros except for 1 in the jth place, 0_v is a $(K - 1) \times 1$ vector of zeros, 0_m is a $p \times (K - 1)$ matrix of zeros, and I_{K-1} is the identity matrix of order $K - 1$. Note that the intercepts from the alternative quantile regressions $(\beta_{\theta_1 1}, \ldots, \beta_{\theta_p 1})$, need not be equal.

The asymptotic distribution of the optimal MD estimator, $\hat{\beta}_\theta^R$, is given by $\sqrt{n}(\hat{\beta}_\theta^R - \beta_\theta^R) \xrightarrow{L} N(0, \Lambda_\theta^R)$, where $\Lambda_\theta^R = (R'\Lambda_\theta^{-1}R)^{-1}$. A test statistic is immediately available

16. Koenker and Bassett considered a particular model of multiplicative heteroskedasticity that implies a certain structure on the quantile parameter vectors.

17. Note that if $\psi = \Lambda_\theta$ [defined in (25)] then the resulting estimate for β_θ^R is the optimal MD estimator whose asymptotic covariance matrix is given below. If $\psi \neq \Lambda_\theta$ then the asymptotic covariance matrix for the MD estimator is given by $\Lambda_\theta^R = (R'\psi^{-1}R)^{-1}R'\psi^{-1}\Lambda_\theta\psi^{-1}R(R'\psi^{-1}R)^{-1}$.

from the MD framework. For $\Psi = \Lambda_\theta$, defined in (25), we have, under the null hypothesis (of equality among the slope coefficients):

$$(27)\quad n(\hat{\beta}_\theta - R\hat{\beta}_\theta^R)'A^{-1}(\hat{\beta}_\theta - R\hat{\beta}_\theta^R) \xrightarrow{L} \chi^2_{(pK-p-K+1)}.$$

B. Test for Symmetry

As was noted by Newey and Powell (1987) it is possible to test for symmetry. Let $\hat{\beta}_\theta$ and $\hat{\beta}_{(1-\theta)}$ be the estimated parameter vectors for the θth and $(1-\theta)$th quantile regressions, respectively. Symmetry implies then that $\hat{\beta}_{.50} = .5(\beta_\theta + \beta_{(1-\theta)})$, where $\beta_{.50}$ is the median $(\theta = .50)$ parameter vector.

Let $\beta_{\theta_1}, \ldots, \beta_{\theta_p}$ be the parameter vectors associated with the $\theta_1, \ldots, \theta_p$ quantiles, respectively, and let p be an odd number. Furthermore, in order to test for symmetry we also need that $\theta_{p-j} = 1 - \theta_{j+1}$ for $j = 0, \ldots, (p-1)/2 - 1$, and the middle quantile $\beta_{\theta_{1+(p-1)/2}}$ is for $\theta_{1+(p-1)/2} = .50$. From the symmetry assumption it follows that $\beta_{\theta_{p-j}} = 2\beta_{.50} - 2\beta_{\theta_{j+1}}$.

Let $\hat{\beta}_\theta^U = (\hat{\beta}'_{\theta_1}, \ldots, \hat{\beta}'_{\theta_p})'$, be the stacked vector of the p unrestricted quantile estimates and let $\beta_\theta^R = (\beta'_1, \ldots, \beta'_{1+(p-1)/2})'$ be the restricted population parameter. Then an estimate for β_θ^R is obtained as a solution to

$$(28)\quad \min_b (\hat{\beta}_\theta^U - Rb)'A^{-1}(\hat{\beta}_\theta^U - Rb),$$

where the restriction matrix R is given by $R = (R'_1, R'_2)'$. R_1 is an identity matrix of dimension $(1 + (p-1)/2)K$, and

$$R_2 = \begin{pmatrix} -I_K & 0_K & 0_K & \ldots & 0_K & 2I_K \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0_K & 0_K & 0_K & \ldots & -I_K & 2I_K \end{pmatrix}$$

is a $K(p-1)/2 \times K(1 + (p-1)/2)$ matrix.

The asymptotic distribution of the test statistic, under the null hypothesis, is similar to that given in (27). That is, for $A \xrightarrow{p} \Lambda_\theta$ as $n \to \infty$, we have

$$(29)\quad n(\hat{\beta}_\theta^U - R\hat{\beta}_\theta^R)'A^{-1}(\hat{\beta}_\theta^U - R\hat{\beta}_\theta^R) \xrightarrow{L} \chi^2_{((p-1)K/2)},$$

where $\hat{\beta}_\theta^R$ denotes the MD estimate.

A few remarks about the testing procedure described above are in line. First, note that the outcome of the tests is either rejection or nonrejection of the null hypothesis—we do not obtain any measure of the power of the test against alternative hypotheses. The setup in Koenker and Bassett (1982) and Newey and Powell (1987) enables them to compute local power under local alternative hypotheses. Second, note that in the case of the homoskedasticity test, if the null hypothesis is not rejected, then the MD framework provides an optimal way to combine the different slope coefficients. Finally, note that one needs to choose a weight matrix in (26) or (28). This matrix can be any matrix (including a nonrandom matrix) whose limit is a nonstochastic positive definite matrix Ψ. For A in (27) or (29), one can use one of the asymptotic covariance estimates described in Section IV. For the homoskedasticity test all covariance matrices are valid, but for the symmetry test only the estimates for the general formula in (6) can be used.

C. *Testing Using the GMM Framework*

An alternative testing procedure can be applied here using Hansen's (1982) GMM method. The moment function defined in (23) is a $pK \times 1$ vector. Under the null hypothesis of either symmetry or heteroskedasticity (of the form discussed above) there are a fewer number of parameters to be estimated, say q, than pK. Hansen's GMM framework provides an estimator for β_θ^R, say $\hat{\beta}_\theta^R$, which is obtained as a solution to

$$(30) \quad \min_b \left(\frac{1}{n}\sum_{i=1}^{n} \psi(x_i, y_i, b)\right)' A^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} \psi(x_i, y_i, b)\right).$$

An efficient estimator can be obtained if A is chosen that $A \xrightarrow{p} E[\psi(x, y, \beta_\theta)\psi(x, y, \beta_\theta)']$ as $n \to \infty$, where $\psi(x, y, \beta_\theta)$ is defined in (23). This framework provides us with a straightforward testing procedure. Under the null hypothesis (of either heteroskedasticity or symmetry) we have

$$n\left(\frac{1}{n}\sum_{i=1}^{n} \psi(x_i, y_i, \hat{\beta}_\theta^R)\right)' A^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} \psi(x_i, y_i, \hat{\beta}_\theta^R)\right) \xrightarrow{L} \chi^2(pK - q) \quad \text{as } n \to \infty.$$

The drawback of this approach is that solving the problem in (30) can be difficult since it does not have an LP representation. Note that, because of the linearity of the conditional quantiles, the estimates obtained from (30), with an optimal A, are (asymptotically) equivalent to those obtained by the optimal MD discussed above. Therefore, in the following empirical example, I report only the test statistic for the MD method.

VII. Empirical Results

A. *General Setup and Results*

In this example, I adopt a linear model of the type introduced by Mincer (1974). The dependent variable is log of weekly wage.[18] The set of independent variables includes: education, education squared, experience, experience squared, an interaction term between education and experience, three regional dummy variables, metropolitan area dummy variable, interaction variables between each of the regional dummy variables and the metropolitan area dummy variable, part-time dummy variable, two part-year dummy variables, a dummy variable for race, and interaction terms between the race dummy variable and education, experience and the part-

18. This variable is defined as the natural log of total annual earnings divided by the number of weeks worked.

time and part-year dummy variables.[19] As mentioned above, I compute five quantile regressions, namely .10, .25, .50, .75 and .90, using the same set of independent variables in each regression.

After obtaining estimates for the coefficient vectors from the five regressions, one should first test whether or not they are statistically different from each other. As explained above, if the model was truly a location model, then all the slope coefficients would be the same. Note that, under independence, all the covariance estimates described above are valid. Table 2 reports the test statistics for each year and age group. This table shows that the null hypothesis (of equality among the slope coefficients) is overwhelmingly rejected, regardless which covariance matrix estimate is used, for all years and age groups. This test provides no indication, however, as to the form of the dependence between the error term and the regressors.

To better evaluate the results, I estimated the multiplicative heteroskedasticity model considered in Koenker and Bassett (1982). This model implies a certain structure on the parameter estimate at the various quantiles. Consider the model given by (dropping the i subscript) $y = x'\beta + \sigma(x;\gamma)\varepsilon$, where $\varepsilon \sim$ i.i.d.$(0, \sigma^2_\varepsilon)$ independent of x. Let also $\sigma(x;\gamma) = 1 + x'\gamma$. Then the conditional quantile of y is given by

$$Q_\theta(y|x) = x'(\beta + \gamma Q^\varepsilon_\theta) + Q^\varepsilon_\theta = x'\delta_\theta,$$

where $\delta_\theta = \beta + (\gamma + e'_1)Q^\varepsilon_\theta$, $e_1 = (1, \ldots, 0)$, and Q^ε_θ denotes the θth quantile of ε. Note that: (a) we cannot identify both β_1 and γ_1 (corresponding to the constant terms); and (b) not all of the Q^ε_θ's can be identified. I therefore introduce normalization, $\gamma_1 = 0$ and $Q^\varepsilon_{.50} = 0$, and estimate the remaining coefficients using the MD framework. In this specific model, the set of restricted parameters are nonlinear functions of a smaller subset of parameters, namely, $\delta_\theta = \delta(\mu)$, where $\mu' = (\beta_1, \ldots, \beta_K, \gamma_2, \ldots, \gamma_k, Q^\varepsilon_{\theta_1}, Q^\varepsilon_{\theta_2}, Q^\varepsilon_{\theta_4}, Q^\varepsilon_{\theta_5})$. Given the quantile regression estimates, an efficient estimator for μ is given by

$$\hat{\mu}_\theta = \arg\min_\mu (\hat{\beta}_\theta - \delta(\mu))'\hat{\Lambda}_\theta^{-1}(\hat{\beta}_\theta - \delta(\mu)),$$

and its asymptotic covariance is given by

$$\Lambda_\mu = \left(\frac{\partial\delta(\mu)}{\partial\mu}\Lambda_\theta^{-1}\frac{\partial\delta(\mu)}{\partial\mu'}\right)^{-1}.$$

Under the null hypothesis of multiplicative heteroskedasticity it follows then that

$$n(\hat{\beta}_\theta - \delta(\hat{\mu}_\theta))'\hat{\Lambda}_\theta^{-1}(\hat{\beta}_\theta - \delta(\hat{\mu}_\theta)) \xrightarrow{L} \chi^2(5K - (2K + 3)).$$

19. Education is defined as the number of full years of education. Experience is defined as Experience = max{age − education − 6,0}. There are three regional dummy variables for the CPS regions. Northeast, North Central and South (the West region is the excluded region). Each of these dummy variables takes the value 1 if the individual resided in that region and 0 otherwise. The metropolitan area dummy variable takes the value 1 if the individual resided in a metropolitan area (according to the CPS definition) and 0 otherwise. The part-time dummy takes the value 1 if an individual worked less than 30 hours per week and 0 otherwise. The first part-year dummy takes the value 1 if the individual worked between 26 and 39 weeks and 0 otherwise. The second part-year dummy takes the value 1 if an individual worked up to 26 weeks and 0 otherwise.

Table 2
χ^2 Test for Equality Among Slope Parameters

		Age Group		
Variance Estimate	All	18–34	35–49	50–64
1972				
Order statistic	2,649.5	1,295.6	721.2	988.8
Homoskedastic kernel	2,067.0	958.7	573.5	682.3
General kernel	1,457.0	659.8	439.3	512.6
Design matrix bootstrap	3,971.4	1,006.8	601.5	519.1
1979				
Order statistic	1,631.6	1,009.9	821.6	929.2
Homoskedastic kernel	1,435.7	759.1	663.5	589.7
General kernel	1,021.8	600.2	470.6	330.9
Design matrix bootstrap	1,636.2	896.0	719.2	429.2
1985				
Order statistic	1,138.4	698.6	835.6	598.2
Homoskedastic kernel	1,019.8	524.9	655.5	416.2
General kernel	674.6	408.2	361.9	299.5
Design matrix bootstrap	1,139.0	559.1	448.6	424.7
1992				
Order statistic	745.3	474.9	721.8	606.2
Homoskedastic kernel	651.9	329.0	504.5	427.8
General kernel	506.2	292.2	408.3	268.8
Design matrix bootstrap	609.1	484.5	440.0	305.0

Note: The reported statistic is $n\,(\hat{\beta}_\theta - \hat{\beta}_\theta^R)'\,\hat{\Lambda}_\theta^{-1}\,(\hat{\beta}_\theta - \hat{\beta}_\theta^R)$, where $\hat{\beta}_\theta$ is a stacked vector of all unrestricted coefficient estimates, $\hat{\Lambda}_\theta$ is its asymptotic covariance estimate, and $\hat{\beta}_\theta^R$ is the restricted parameter vector estimate (obtained by minimum distance). This statistic has an asymptotic χ^2 distribution with $5k - 5 - (k - 1) = 84$ degree of freedom.

Table 3 reports the test statistics for each year and age group using the order statistic and design matrix bootstrap estimators for the unrestricted covariance matrix Λ_θ. The null hypothesis is again rejected in all cases, but there is a clear indication that the rejection is not as strong as in the previous test. Also note that the test statistics are larger when they are based on order statistic estimates for Λ_θ than when they are based on design matrix bootstrap estimates. This is not surprising since the order statistic estimator is not valid as is implied by the results of the previous test. The results seem to indicate that there exists a more complicated type of dependence between u_θ and x.

B. Returns to Education

Let's now focus our attention on the evolution of the returns to education over time across the different quantiles for the various age groups. The return to education

Table 3
χ^2 Test for Multiplicative Heteroskedasticity Model

	Age Group			
Variance Estimate	All	18–34	35–49	50–64
1972				
Order statistic	501.6	251.6	155.8	250.9
Design matrix bootstrap	492.8	154.9	119.7	103.6
1979				
Order statistic	330.3	249.7	175.5	273.3
Design matrix bootstrap	329.9	238.1	134.2	107.9
1985				
Order statistic	283.5	207.8	269.1	239.5
Design matrix bootstrap	249.1	160.4	173.9	131.2
1992				
Order statistic	198.7	202.2	211.8	211.1
Design matrix bootstrap	169.3	175.6	136.0	117.5

Note: The reported statistic is $n\,(\hat{\beta}_\theta - \delta(\hat{\mu}_\theta))'\,\hat{\Lambda}_\theta^{-1}\,(\hat{\beta}_\theta - \delta(\hat{\mu}_\theta))$, where $\hat{\beta}_\theta$ is a stacked vector of all unrestricted coefficient estimates, $\hat{\Lambda}_\theta$ is its asymptotic covariance estimate, and $\hat{\mu}_\theta$ is the restricted parameter vector estimate (obtained by minimum distance). This statistic has an asymptotic χ^2 distribution with $5k - (2k - 1) - 4 = 63$ degree of freedom.

(ed) is defined as the derivative of the conditional quantile with respect to education, $\partial\text{Quant}_\theta(y|x)/\partial\text{ed}$. Because the wage equation includes a squared term of education, interaction terms between education and experience (ex) and between education and a race dummy (ra), this derivative is given by

$$\frac{\partial\text{Quant}_\theta(y|x)}{\partial\text{ed}} = \beta_e + 2\beta_{e^2}\text{ed} + \beta_{\text{ex}}\text{ex} + \beta_{\text{ra}}\text{ra}.$$

Table 4 reports the returns to education at five quantiles for the model estimated using the entire sample. The returns are evaluated for white males at the levels of education and experience denoted in the table. We can see from this table that the returns at all quantiles declined significantly during the period from 1972 to 1979. By 1985 the returns at all quantiles increased to unprecedented highs; reaching levels which are twice as large as in 1979. Note, however, the distinct features across the various quantiles. For example, in 1972 and 1979 the returns for the high school graduates (12 years of education) at the entry and mid-career levels (five and 15 years of experience, respectively) are lower at the higher quantiles. For experienced workers (25 years of experience) with high school education, this pattern is completely reversed.

In contrast, the returns for the college graduates (16 years of education) are always higher at the higher quantiles, especially in the years when the return to education

Table 4
Return to Education for the Entire Sample, by Education and Experience Groups

		1972						1979					
		Quantile						Quantile					
Ed	Ex	.10	.25	.50	.75	.90	LS	.10	.25	.50	.75	.90	LS
12	5	10.1	9.9	9.4	9.0	8.6	9.3	9.2	9.0	8.6	7.7	7.5	8.4
		(.36)	(.08)	(.03)	(.27)	(.69)	(.06)	(.49)	(.17)	(.18)	(.02)	(.10)	(.10)
12	15	8.1	8.1	8.0	8.2	8.6	8.1	7.7	7.7	7.4	6.9	7.0	7.3
		(.39)	(.10)	(.03)	(.26)	(.69)	(.05)	(.51)	(.18)	(.19)	(.02)	(.10)	(.11)
12	25	6.1	6.3	6.6	7.4	8.6	6.8	6.2	6.4	6.2	6.2	6.6	6.1
		(.42)	(.12)	(.04)	(.26)	(.69)	(.06)	(.54)	(.19)	(.20)	(.03)	(.10)	(.12)
16	5	9.2	10.1	10.4	11.6	12.9	10.6	7.3	8.0	7.3	8.2	9.0	7.9
		(.40)	(.08)	(.05)	(.36)	(.90)	(.09)	(.58)	(.20)	(.21)	(.02)	(.14)	(.11)
16	15	7.3	8.3	9.0	10.8	12.9	9.4	5.8	6.7	6.2	7.5	8.5	6.8
		(.43)	(.10)	(.05)	(.36)	(.90)	(.09)	(.60)	(.21)	(.22)	(.03)	(.14)	(.12)
16	25	5.3	6.5	7.6	10.0	12.9	8.1	4.4	5.3	5.0	6.8	8.1	5.6
		(.46)	(.12)	(.05)	(.35)	(.90)	(.08)	(.62)	(.22)	(.22)	(.03)	(.14)	(.13)

		1985						1992					
		Quantile						Quantile					
Ed	Ex	.10	.25	.50	.75	.90	LS	.10	.25	.50	.75	.90	LS
12	5	10.7	11.5	11.8	11.5	11.2	11.3	12.7	13.3	13.6	13.5	13.6	13.2
		(.05)	(.10)	(.18)	(.14)	(.10)	(.07)	(.27)	(.11)	(.08)	(.11)	(.19)	(.08)
12	15	9.3	9.9	10.1	9.8	9.9	9.7	11.2	11.7	11.9	11.9	12.1	11.6
		(.08)	(.11)	(.19)	(.15)	(.10)	(.09)	(.26)	(.11)	(.08)	(.10)	(.18)	(.08)
12	25	7.8	8.3	8.3	8.2	8.7	8.2	9.7	10.1	10.2	10.3	10.6	9.9
		(.10)	(.13)	(.20)	(.17)	(.10)	(.10)	(.26)	(.11)	(.09)	(.10)	(.18)	(.08)
16	5	11.7	11.4	11.2	11.5	12.7	11.5	15.0	15.2	15.3	15.5	16.1	15.1
		(.08)	(.10)	(.19)	(.13)	(.16)	(.07)	(.39)	(.17)	(.12)	(.17)	(.27)	(.12)
16	15	10.2	9.9	9.5	9.9	11.4	9.9	13.5	13.6	13.6	13.9	14.5	13.4
		(.09)	(.11)	(.20)	(.15)	(.16)	(.08)	(.38)	(.17)	(.13)	(.16)	(.27)	(.12)
16	25	8.8	8.3	7.8	8.2	10.1	8.4	12.0	11.9	12.0	12.3	13.0	11.8
		(.11)	(.13)	(.22)	(.17)	(.16)	(.09)	(.37)	(.16)	(.13)	(.15)	(.26)	(.12)

Note: Ex denotes potential experience, Ed denotes education. The column titled LS gives the mean return to education computed via least-squares. The numbers in parentheses are standard errors computed using the design matrix bootstrap estimator. The standard error for the least-squares estimates are computed using White's method.

was relatively low. As the returns at the various quantiles increased during the 1980s, there is considerable convergence of these returns resulting from sharper increases at the lower quantiles. In the last sample year (1992) we clearly see that the returns for all education-experience combinations are higher at the higher quantiles. For most skill groups there are differences of close to one percentage point between the returns at the .90 quantile and the .10 quantile.

The results in Table 4 also indicate that the returns for the more experienced workers are significantly lower, at all quantiles, than for the less experienced workers. Workers at the entry level have a higher return to their education at each and every quantile. As they become more experienced, the return to education declines. This result is consistent with life-cycle labor supply models which predict a decline in the rate of return over one's career. It is apparent from Table 4, however, that larger differences exist at the bottom of the wage distribution than at the top, especially during the early sample years.

The next three tables present the returns to education computed from the quantile (and least-squares) regressions carried out for each of the three age groups considered in this study. Tables 5, 6 and 7 present the results for the 18–34, 35–49, and 50–64 age groups, respectively. There are several differences between the various age groups that cannot be seen from Table 4. When estimated separately, the 18–34 age group, which contains of all the new entrants, exhibits higher returns at the entry level and lower returns at the midcareer level. In general, sharper declines in the returns are apparent between 1972 and 1979. While the rise in the returns to education at all quantiles can be clearly seen in 1985 and 1992, there are sharper increases at the higher quantiles. Especially for the college graduates the returns seem to be higher at the lower quantiles in 1979 and 1985. The sharper increases at the upper quantiles between 1985 and 1992, lead to the same convergence phenomena described earlier for the whole sample. The other two age groups (Tables 6 and 7) exhibit more similarities to the earlier results although the magnitude of the results seems to be dominated by the youngest group. That is, the more substantial increases in the returns to education are for the 18–34 age group.

Overall, the results indicate that the returns to education exhibit different patterns of changes over time, for the various education-experience groups. Moreover, in most cases there are significant differences (of a few percentage points) in the levels of the returns to education at the various quantiles. Changes in the mean returns to education are quite different, as well, across the various skill groups. Nevertheless, these changes do not capture the differential changes across the various quantiles induced by changes in the shapes of the corresponding conditional wage distribution.

C. Standard Error Estimates

In Section III several alternative estimators for the asymptotic covariance matrix were described. From this set of estimators, only the general kernel and the design matrix bootstrap estimators are valid under the general dependence structure, namely, $f_{u_\theta}(0|x) \neq f_{u_\theta}(0)$ for all x. An immediate practical question is: how different are the alternative estimates for a given sample. To answer this question I report ratios between alternative standard errors estimates and the design matrix bootstrap

Table 5

Return to Education for the 18–34 Age Group, by Education and Experience Groups

		1972						1979					
		Quantile						Quantile					
Ed	Ex	.10	.25	.50	.75	.90	LS	.10	.25	.50	.75	.90	LS
12	5	10.9	10.5	10.1	9.6	8.8	9.8	8.5	9.0	8.8	8.2	8.3	8.5
		(2.38)	(0.92)	(0.66)	(0.27)	(0.34)	(0.51)	(0.94)	(0.65)	(0.82)	(0.78)	(0.83)	(0.53)
12	15	7.6	6.5	5.6	5.7	6.2	6.2	5.9	6.1	5.9	4.9	5.1	5.5
		(2.70)	(1.20)	(0.86)	(0.48)	(0.29)	(0.69)	(1.13)	(0.76)	(0.95)	(0.95)	(1.01)	(0.64)
16	5	8.0	9.1	8.5	9.8	10.6	8.9	6.9	6.7	5.7	6.0	6.2	6.4
		(2.82)	(1.03)	(0.75)	(0.26)	(0.52)	(0.56)	(1.08)	(0.77)	(0.97)	(0.90)	(0.96)	(0.62)
16	15	4.7	5.1	4.0	5.8	8.1	5.2	4.2	3.9	2.8	2.7	3.0	3.4
		(3.12)	(1.30)	(0.95)	(0.47)	(0.41)	(0.74)	(1.27)	(0.87)	(1.10)	(1.07)	(1.14)	(0.72)
		1985						1992					
		Quantile						Quantile					
Ed	Ex	.10	.25	.50	.75	.90	LS	.10	.25	.50	.75	.90	LS
12	5	8.3	10.6	12.2	12.5	12.1	11.2	11.0	11.8	13.4	13.3	13.3	12.6
		(0.81)	(0.05)	(0.61)	(0.89)	(0.58)	(0.27)	(0.73)	(1.12)	(0.43)	(0.35)	(0.20)	(0.42)
12	15	10.9	10.6	9.0	8.3	7.3	9.1	11.3	12.7	12.3	11.5	10.2	11.6
		(1.00)	(0.05)	(0.76)	(1.16)	(0.95)	(0.36)	(0.76)	(1.17)	(0.39)	(0.28)	(0.34)	(0.39)
16	5	10.6	11.4	10.9	11.3	12.0	11.1	13.6	16.3	16.3	15.8	15.1	15.7
		(1.05)	(0.06)	(0.68)	(0.98)	(0.59)	(0.27)	(1.01)	(1.46)	(0.62)	(0.52)	(0.33)	(0.58)
16	15	13.2	11.4	7.7	7.0	7.2	9.0	13.9	17.2	15.2	14.0	12.1	14.7
		(1.23)	(0.06)	(0.83)	(1.26)	(0.96)	(0.37)	(1.03)	(1.51)	(0.57)	(0.44)	(0.34)	(0.55)

Note: See note from Table 4.

Table 6
Return to Education for the 35–49 Age Group, by Education and Experience Groups

		1972						1979					
		Quantile						Quantile					
Ed	Ex	.10	.25	.50	.75	.90	LS	.10	.25	.50	.75	.90	LS
12	15	7.1	6.6	8.5	7.8	8.4	7.7	7.1	7.4	7.5	6.9	7.5	7.2
		(1.34)	(0.12)	(0.27)	(1.14)	(2.82)	(0.21)	(1.76)	(0.86)	(0.87)	(0.09)	(0.31)	(0.53)
12	25	5.8	6.0	6.6	7.8	9.2	6.8	6.6	6.9	6.6	6.6	7.5	6.6
		(1.53)	(0.17)	(0.42)	(1.15)	(2.91)	(0.20)	(1.82)	(0.89)	(0.93)	(0.11)	(0.31)	(0.57)
16	15	6.5	7.6	9.7	11.9	14.2	9.6	4.6	5.7	5.8	7.5	9.3	6.2
		(1.43)	(0.14)	(0.23)	(1.47)	(3.51)	(0.33)	(2.04)	(0.98)	(0.98)	(0.08)	(0.44)	(0.59)
16	25	5.2	7.0	7.8	12.0	15.0	8.6	4.1	5.3	4.8	7.3	9.3	5.6
		(1.62)	(0.16)	(0.37)	(1.47)	(3.61)	(0.30)	(2.10)	(1.01)	(1.04)	(0.09)	(0.44)	(0.62)
		1985						1992					
		Quantile						Quantile					
Ed	Ex	.10	.25	.50	.75	.90	LS	.10	.25	.50	.75	.90	LS
12	15	10.5	9.9	10.4	9.9	10.6	10.3	10.4	9.4	9.3	10.7	11.6	10.1
		(1.55)	(0.70)	(1.23)	(0.89)	(0.21)	(0.71)	(0.72)	(0.77)	(0.48)	(0.21)	(0.47)	(0.27)
12	25	8.7	8.7	8.3	8.3	9.8	8.5	10.8	10.4	10.3	10.2	11.1	10.2
		(1.81)	(0.80)	(1.40)	(1.02)	(0.25)	(0.84)	(0.77)	(0.86)	(0.55)	(0.19)	(0.44)	(0.28)
16	15	10.0	9.5	8.9	9.2	12.1	9.8	12.6	11.9	11.3	12.6	14.3	12.2
		(1.62)	(0.73)	(1.35)	(0.94)	(0.29)	(0.74)	(1.01)	(0.99)	(0.63)	(0.34)	(0.70)	(0.40)
16	25	8.2	8.3	6.8	7.6	11.3	7.9	12.9	13.0	12.3	12.1	13.9	12.3
		(1.87)	(0.82)	(1.51)	(1.07)	(0.29)	(0.87)	(1.05)	(1.07)	(0.69)	(0.32)	(0.66)	(0.40)

Note: See note from Table 4.

Table 7

Return to Education for the 50–64 Age Group, by Education and Experience Groups

		1972						1979					
		Quantile						Quantile					
Ed	Ex	.10	.25	.50	.75	.90	LS	.10	.25	.50	.75	.90	LS
12	15	11.3	11.9	7.3	7.1	6.2	8.5	6.1	6.7	10.7	6.2	5.0	7.3
		(3.13)	(1.67)	(0.18)	(1.14)	(3.84)	(0.21)	(0.55)	(0.34)	(1.10)	(0.77)	(1.54)	(0.18)
12	25	8.3	8.9	6.8	7.3	7.7	7.6	5.5	6.0	8.4	6.1	5.6	6.4
		(3.70)	(2.04)	(0.18)	(1.16)	(4.07)	(0.24)	(0.66)	(0.42)	(1.36)	(0.76)	(1.60)	(0.25)
16	15	10.5	11.9	8.7	10.4	11.8	10.3	6.6	7.4	11.2	8.8	8.3	8.5
		(3.25)	(1.67)	(0.29)	(1.46)	(4.62)	(0.31)	(0.51)	(0.29)	(1.06)	(1.01)	(1.90)	(0.18)
16	25	7.5	8.9	8.2	10.6	13.4	9.4	5.9	6.7	8.8	8.7	8.8	7.6
		(3.82)	(2.04)	(0.27)	(1.47)	(4.85)	(0.30)	(0.62)	(0.37)	(1.33)	(1.00)	(1.96)	(0.22)

		1985						1992					
		Quantile						Quantile					
Ed	Ex	.10	.25	.50	.75	.90	LS	.10	.25	.50	.75	.90	LS
12	15	7.5	9.0	11.2	9.0	14.3	12.8	24.0	13.2	16.6	13.4	12.9	15.3
		(0.39)	(0.57)	(1.03)	(0.37)	(1.30)	(1.10)	(8.28)	(2.27)	(2.60)	(0.71)	(0.60)	(1.56)
12	25	6.8	8.0	9.3	8.2	11.6	10.3	16.5	10.8	13.0	11.5	11.4	12.1
		(0.43)	(0.71)	(1.26)	(0.33)	(1.77)	(1.36)	(10.01)	(2.72)	(3.13)	(0.97)	(0.63)	(1.93)
16	15	9.0	9.8	11.8	11.1	16.1	13.4	22.8	13.4	16.9	15.4	16.0	15.9
		(0.60)	(0.50)	(0.97)	(0.60)	(1.13)	(1.06)	(8.49)	(2.25)	(2.56)	(0.61)	(0.90)	(1.50)
16	25	8.3	8.8	10.0	10.3	13.5	11.0	15.4	11.0	13.2	13.5	14.5	12.7
		(0.55)	(0.63)	(1.20)	(0.53)	(1.58)	(1.32)	(10.22)	(2.70)	(3.10)	(0.85)	(0.84)	(1.87)

Note: See note from Table 4.

standard error estimates.[20] The results for a subset of the coefficients for the quantile regressions carried out for the entire sample are presented in Table 8.

Table 8 shows that there are considerable differences between the standard error estimates. In general, the order statistic and homoskedastic kernels provide standard errors which are much smaller than those provided by the bootstrap method. In some cases these standard errors are smaller by more than 20 percent and can lead to false inference if the general dependence structure applies. Note however that the general kernel estimator provides standard errors which are quite similar to those provided by the design matrix bootstrap. This is an encouraging result, since both methods are valid under any dependence structure.

Two main conclusions should be drawn from the experiment presented here. First, one must not use methods for estimating the covariance matrix which apply only under the independence structure if, in fact, tests indicate that this is not the case. Second, the two methods which are valid under the general dependence structure are equivalent and can always be used. Which of the two should be used is more of a practical question. Both methods require lengthy computation time. For the bootstrap estimate, the large number of bootstrap repetitions can be computationally very demanding. On the other hand, the choice of the optional bandwidth needed for the general kernel estimator can be just as lengthy if the sample size is large.

VIII. Censored Quantile Regression

When some of the observations are top coded (censored) an extension to the quantile regression was suggested by Powell (1984 and 1986).

A. The Estimator

The censored regression model (or "Tobit" model) can be written in the form

$$y_i = \min\{y_i^0, x_i'\beta_\theta + u_{\theta_i}\}, \quad i = 1, \ldots, n,$$

where y_i^0 is the top coding value of y_i in the sample. In this formulation y_i^0 is fixed and known, even for the observations which are not censored.[21]

This model can be written as a latent variable model,

$$y_i^* = x_i'\beta_\theta + u_{\theta_i},$$

where $\mathrm{Quant}_\theta(u_{\theta_i}|x_i) = 0$ and $y_i = y_i^* I(y_i^* \le y^0)$. It is easy to see that the θth quantile of the observed y_i is given by $\mathrm{Quant}_\theta(y_i|x_i, \beta_\theta) = \min\{y^0, x_i'\beta_\theta\}$.

20. The design matrix bootstrap standard errors are computed using the data dependent method suggested by Andrews and Buchinsky (1996) for determining the number of bootstrap repetitions. Typically, the number of repetitions required varied from about 300 (at the extreme quantiles) to about 400 (at the middle quantiles).

21. In the following I assume, for simplicity of presentation, that $y_i^0 = y^0$, namely, it is the same for all observations. Buchinsky and Hahn (1996) considered a more general case in which the censoring point is an unknown function of a known set of regressors. An estimator for the case in which the censoring point is random was proposed by Honoré and Powell (1993).

Table 8

Ratio Between Alternative Standard Error Estimates and the Design Matrix Bootstrap Standard Error Estimates

	Quantile					Quantile				
Coefficient	.10	.25	.50	.75	.90	.10	.25	.50	.75	.90
	1972					1985				
Order statistic										
Constant	0.87	0.99	0.83	0.67	0.69	0.91	1.08	0.88	0.78	0.78
Education	0.86	0.96	0.81	0.63	0.65	0.91	1.08	0.89	0.77	0.77
Education squared	0.84	0.92	0.80	0.60	0.61	0.90	1.02	0.88	0.78	0.78
Experience	0.85	1.03	0.86	0.74	0.70	0.90	0.92	0.87	0.86	0.76
Experience squared	0.80	1.05	0.86	0.77	0.69	0.92	0.86	0.85	1.03	0.79
Homoskedastic Kernel										
Constant	0.94	1.06	0.93	0.82	0.82	0.98	1.28	0.95	0.78	0.79
Education	0.94	1.04	0.91	0.77	0.76	0.97	1.28	0.96	0.77	0.78
Education squared	0.91	0.99	0.90	0.73	0.72	0.97	1.21	0.95	0.78	0.79
Experience	0.93	1.11	0.96	0.90	0.83	0.97	1.09	0.94	0.86	0.78
Experience squared	0.88	1.13	0.96	0.94	0.81	0.99	1.02	0.92	1.02	0.80
General Kernel										
Constant	1.02	1.08	1.07	1.06	1.07	0.97	1.15	0.99	0.95	1.05
Education	0.99	1.04	1.06	1.02	1.03	0.96	1.14	1.01	0.92	1.04
Education squared	0.96	0.99	1.05	0.98	0.99	0.93	1.08	1.01	0.91	1.04
Experience	1.02	1.18	1.10	1.09	1.01	1.06	1.07	0.99	0.98	0.92
Experience squared	1.03	1.22	1.06	1.06	0.94	1.10	1.06	0.99	1.09	0.92

Note: The standard errors are computed for the entire samples of 1972 and 1985. The results reported here are for a subset of the coefficients from the quantile regression (see the text for the description of all the coefficients included in the regressions).

The estimator $\hat{\beta}_\theta$ for β_θ, suggested by Powell, is defined as the solutions to

$$(31)\quad \min_\beta \frac{1}{n}\sum_{i=1}^{n} \rho_\theta(y_i - \min\{y^0, x_i'\beta\}),$$

where $\rho_\theta(\lambda)$ is a check function defined above. Note that in order to obtain a consistent estimator of β_θ it is necessary that $x_i'\beta_\theta < y^0$ for a positive fraction of the sample.

Rewriting (31) in a fashion similar to (3) gives that $\hat{\beta}_\theta$ solves

$$\min_\beta \frac{1}{n}\sum_{i=1}^{n} (\theta - 1/2 + 1/2\ \mathrm{sgn}(y_i - \min\{y^0, x_i'\beta\}))(y_i - \min\{y^0, x_i'\beta\}).$$

The F.O.C. to the problem in (31) is given by

$$\frac{1}{n}\sum_{i=1}^{n} I(x_i'\hat{\beta}_\theta < y^0)(\theta - 1/2 + 1/2\ \mathrm{sgn}(y_i - x_i'\hat{\beta}_\theta))x_i = o_p(n^{-1/2}).$$

Thus the moment function can be defined (dropping the i subscript for simplicity) by

$$\psi(x, y, \beta) = I(x'\beta < y^0)(\theta - 1/2 + 1/2\ \mathrm{sgn}(y - x'\beta))x.$$

Powell (1984 and 1986) showed that under certain regularity conditions $E[\psi(x, y, \beta_\theta)] = 0$. The GMM framework can be employed again to establish $\sqrt{n}$-consistency and asymptotic normality for $\hat{\beta}_\theta$. Specifically, it can be shown that under some regularity conditions (see Powell 1984, 1986 for details) $\sqrt{n}(\hat{\beta}_\theta - \beta_\theta) \xrightarrow{L} N(0, \Lambda_\theta^P)$, where[22]

$$\Lambda_\theta^P = \theta(1 - \theta)\Delta_f^{-1}\Delta\Delta_f^{-1},$$

$$\Delta_f = E[f_{u_\theta}(0|x)I(x'\beta_\theta \leq y^0)xx'],$$

and

$$\Delta = E[I(x'\beta_\theta < y^0)xx'].$$

Again, if $f_{u_\theta}(0|x) = f_{u_\theta}(0)$ with probability one, then Λ_θ^P simplifies to

$$\Lambda_\theta^P = \frac{\theta(1 - \theta)}{f_{u_\theta}^2(0)} E[I(x'\beta_\theta < y^0)xx'].$$

What is the intuitive rationale behind the estimation procedure of β_θ in the censored regression model? Recall that $x'\beta_\theta$ is the conditional quantile of y^* given x, namely, $x'\beta_\theta = F_{y^*}^{-1}(\theta|x)$. Now, two cases are possible: (i) $x_i'\beta_\theta < y^0$, and (ii) $x_i'\beta_\theta \geq y^0$. In case (i), the probability is $\geq\theta$ that $y^* \equiv x'\beta_\theta + u_\theta \leq y^0$ so that the conditional quantile of y^* given x can be exactly identified. On the other hand, in case (ii), the probability is $<\theta$ that $y^* \leq y^0$, namely, the conditional quantile is in the unobserved part of the distribution. Consequently, nothing can be done with that portion of the data; we know only that its conditional quantile is greater than y^0. The implication is that one has to drop that portion of the data that cannot be used. It follows that the calculated asymptotic covariance matrix has to be adjusted for the fact that the

22. This general form of the asymptotic covariance matrix is derived in Powell (1984) for censored median regression, and extended in Powell (1986) for the general censored quantile regression.

estimation is conditioned on the inclusion of only the observations for which $x'\hat{\beta}_\theta \leq y^0$.

It is important to note that, because of the nature of the censored regression model, if $x_i'\hat{\beta}_\theta \leq y^0$ for all observations, then the quantile regression and censored quantile regression estimates coincide. Although this is likely to happen for the low quantiles (θ close to 0), in the right-censoring case discussed here, the quality of the estimator decreases as θ increases, because fewer observations are used in the estimation of $\hat{\beta}_\theta$.

A considerable drawback of the censored quantile regression model is that it does not have the attractive linear programming representation of the quantile regression model, and hence linear programming algorithms cannot be directly used to estimate β_θ. Also, because the objective function defined in (31) is not convex in β, only a local minimizer $\hat{\beta}_\theta$ can be obtained.

1. Algorithms for Powell's Estimator

Several algorithms that have been suggested in the literature apply to Powell's estimator. Womersley (1986) suggested an algorithm which is directed toward solutions of Least Absolute Deviation (LAD) estimators for the censored regression model. It concentrates on the minimization of the objective function $Q_n(\beta) = \sum_{i=1}^{n}|y_i - \min\{y^0, x_i'\beta\}|$, which is a nonconvex piecewise linear function. The algorithm developed extends the reduced gradient algorithm for LP to provide an efficient (finite direct descent) method for calculating a local minimizer for $Q_n(\beta)$.

Buchinsky (1991) suggested an iterative LP algorithm (ILPA). The basic idea is that if one had known in advance the set of observations for which $x_i'\beta_\theta > y^0$, then these observations could have been excluded from the estimation. Because this set of observations is not known in advance, the algorithm solves for $\hat{\beta}_\theta$ in an iterative way. In any iteration the solution $\hat{\beta}_\theta^{(j)}$ defines the set of observations to be excluded from the next iteration. The $j + 1$ iteration is carried out (using an LP algorithm) on the set of observations such that $x_i'\hat{\beta}_\theta^{(j)} \leq y^0$. Convergence is achieved then when the set of excluded observations in two successive iterations are the same.

Koenker and Park (1996) suggested an interior point algorithm which is suited for a general non-linear quantile regression problem of the form $y = f(x; \beta_\theta) + u_\theta$, for some known function $f(\cdot; \beta_\theta)$. This method in the linear quantile regression case amounts to iterative reweighting of least-squares (IRLS) estimates. Although this method is not as efficient as simplex methods for the linear case, it has the advantage of providing an extremely useful tool for the nonlinear case, such as in the case of the censored regression model. Moreover, it is a natural extension to the linear quantile regression and it has the very desirable property (unlike many of other IRLS methods) that it is guaranteed to converge to the right solution.

B. Some Empirical Evidence Based on Pseudo-Data

In the data used in the empirical example, censoring is not an issue in any of the years analyzed. To evaluate the empirical implications associated with the estimation of a censored quantile regression, I generated pseudo-data from the CPS data for 1979 and 1985. In these two years I artificially censored the data, setting the censor-

Table 9
Return to Education Based on Censored Data for the 35–49 Age Group

		Censored Quantile Regression					Raw Quantile Regression				
		Quantile					Quantile				
Ed	Ex	.10	.25	.50	.75	.90	.10	.25	.50	.75	.90
						1979					
12	15	7.1	7.5	7.5	8.4	11.0	7.1	7.4	7.3	5.0	1.6
		(1.69)	(0.96)	(0.69)	(0.82)	(3.21)	(1.69)	(0.92)	(0.81)	(0.97)	(0.77)
		(1.81)	(0.98)	(0.71)	(1.02)	(2.86)	(1.81)	(0.92)	(0.86)	(2.06)	(3.13)
12	25	6.6	6.9	6.6	6.3	12.4	6.6	6.9	6.3	4.1	1.4
		(1.76)	(1.01)	(0.74)	(0.97)	(3.43)	(1.76)	(0.96)	(0.86)	(1.03)	(0.79)
		(1.87)	(1.02)	(0.76)	(1.18)	(3.06)	(1.87)	(0.95)	(0.91)	(2.14)	(3.16)
16	15	4.6	5.6	5.9	7.5	15.3	4.6	5.6	4.9	1.9	0.0
		(1.95)	(1.10)	(0.77)	(0.88)	(4.08)	(1.95)	(1.05)	(0.92)	(1.14)	(0.91)
		(2.10)	(1.12)	(0.80)	(1.10)	(3.64)	(2.10)	(1.05)	(0.98)	(2.46)	(3.89)
16	25	4.1	5.0	5.0	5.4	16.8	4.1	5.1	3.9	1.0	−0.2
		(2.02)	(1.14)	(0.82)	(1.02)	(4.30)	(2.02)	(1.09)	(0.98)	(1.19)	(0.93)
		(2.16)	(1.16)	(0.85)	(1.25)	(3.83)	(2.16)	(1.08)	(1.03)	(2.54)	(3.92)

						1985					
12	15	10.5	9.9	10.6	10.8	10.7	10.5	9.9	10.2	6.6	2.4
		(1.37)	(0.75)	(1.10)	(1.49)	(0.64)	(1.37)	(0.75)	(1.30)	(1.72)	(1.54)
		(1.69)	(0.68)	(1.08)	(1.76)	(0.90)	(1.69)	(0.68)	(1.36)	(4.11)	(9.39)
12	25	8.7	8.7	8.2	8.0	9.2	8.7	8.7	7.4	4.9	2.0
		(1.61)	(0.85)	(1.26)	(1.70)	(0.88)	(1.61)	(0.85)	(1.46)	(1.83)	(1.60)
		(1.96)	(0.78)	(1.23)	(1.99)	(1.21)	(1.96)	(0.78)	(1.52)	(4.30)	(9.56)
16	15	10.0	9.5	9.0	9.0	11.7	10.0	9.5	7.2	2.2	0.0
		(1.42)	(0.78)	(1.20)	(1.64)	(0.57)	(1.42)	(0.78)	(1.45)	(2.00)	(1.83)
		(1.77)	(0.71)	(1.18)	(1.95)	(0.83)	(1.77)	(0.71)	(1.53)	(4.90)	(11.57)
16	25	8.2	8.3	6.6	6.3	10.2	8.2	8.3	4.4	0.5	−0.4
		(1.66)	(0.88)	(1.36)	(1.85)	(0.79)	(1.66)	(0.88)	(1.61)	(2.11)	(1.88)
		(2.03)	(0.80)	(1.33)	(2.18)	(1.11)	(2.03)	(0.80)	(1.68)	(5.09)	(11.73)

Note: Ex denotes potential experience, Ed denotes education. The numbers in parentheses are standard errors. The first ones are the standard errors computed based on the homoskedastic kernel, while the second ones are computed based on the general kernel. The fractions of uncensored observations for 1979 are 1.0, .99, .92, .70 and .25 for the .10, .25, .50, .75 and .90 quantiles, respectively. The same fractions for 1985 are 1.0, 1.0, .93, .66 and .30, for the same five quantiles.

ing value of real weekly wage (defined above) at $800. Any observation for which the true reported value was above $800 was recorded in the psuedo-data as $800. I then treated the data set as if it were censored to begin with and considered two alternative estimation procedures.

In the first estimation procedure carried out, the quantile regression model was estimated at the same five quantiles as before (with exactly the same regressors as in the previous section) using Powell's censored quantile regression model. This procedure corrects for censoring and uses only the observations whose conditional quantiles are below the censoring point.[23]

In the second estimation procedure considered, the censoring problem was ignored. That is, all observations that were censored were treated as if the censoring value was their true value. We expect this procedure to yield estimates of the returns to education which are downward biased, especially at the quantiles most affected by the censoring.

We can compare these two sets of estimates with those obtained from the real data. The estimation was only carried out for the 35–49 age group. The resulting returns to education are reported in Table 9. Also provided in Table 9 are two alternative standard error estimates based on the: (a) homoskedastic kernel; and (b) general kernel. The estimates based on the order statistic was very similar to the first kernel estimate, while the design matrix bootstrap yielded standard errors very similar to those from the latter kernel method.

The results are very clear, using the above stringent censoring value has a tremendous effect on the estimated returns to education and their standard errors. The first two quantiles are not affected by the introduction of the artificial censoring point and the resulting returns are therefore identical to those reported in Table 6 which are based on the real data. For quantiles of .50 and higher, we see the effect of censoring. Approximately 8 percent and 7 percent of the .50 conditional quantiles are above the censoring value for 1979 and 1985, respectively. This problem becomes more severe at the .75 quantile where 30 percent and 34 percent (for 1979 and 1985 respectively) of the conditional quantiles are above the censoring point. At the .90 quantile the problem becomes almost unbearable; 75 percent and 70 percent of the conditional quantiles for the two years are above the censoring point.

When the censoring is ignored, the returns to education are severely downward biased as can be seen from the last five columns of the table. Moreover, in most cases the estimated standard errors are very large, so that in fact the estimated returns are not statistically different from 0. When the censored regression model is employed there are significant improvements across the board, even for the quantiles that are severely affected by censoring. The returns based on the censored regression model look more similar to those in Table 6 where there is no censoring. One major difference between the results reported in Table 6 and the first five columns of Table 9, is in the estimated standard errors. As one might expect, the standard errors in Table 9 are much larger than those in Table 6, since, essentially, only a small fraction of the data is used. Note, however, that in almost all cases reported in Table 9, the estimated returns to education from Table 6 are within approximately one standard error of the estimates reported in Table 9.

23. I use the ILPA algorithm described above to estimate this model.

This example shows that the appropriate use of the model, i.e., the introduction of the censored quantile regression model allows one to extend the estimation of the quantile regression into cases where other methods (such as the least-squares) are bound to fail.

References

Abdelmalek, N. N. 1971. "Linear l_1 Approximation for Discrete Point Set and l_1 Solutions of Overdetermined Linear Equations." *Journal of the American Statistical Association Computation* 18:41–47.

Andrews, D. W. K., and M. Buchinksy. 1996. "On the Number of Bootstrap Repetitions for Bootstrap Standard Error Estimates, Confidence Intervals, and Tests." Unpublished. Cowles Foundation for Research in Economics, Yale University.

Armstrong, R., E. Frome, and D. Kung. 1979. "A Revised Simplex Algorithm for the Absolute Deviation Curve-Fitting Problem." *Communications in Statistics B. Simulation Computation* 8:175–90.

Bai, J. 1995. "Least Absolute Estimation of a Shift." *Econometric Theory* 11:403–36.

Barrodale, I. 1968. "l_1 Approximation and the Analysis of Data." *Applied Statistics* 17: 51–57.

Barrodale, I., and F. Roberts. 1973. "An Improved Algorithm for Discrete l_1 Linear Approximation." *SIAM Journal of Numerical Analysis* 10:839–48.

Bickel, P., and D. Freedman. 1981. "Some Asymptotic Theory of the Bootstrap." *The Annals of Statistics* 9:1196–1217.

Buchinksy, M. 1995a. "Estimating the Asymptotic Covariance Matrix for Quantile Regression Models: A Monte Carlo Study." *Journal of Econometrics* 68:303–38.

———. 1995b. "Quantile Regression Box-Cox Transformation Model, and the U.S. Wage Structure, 1963–1987." *Journal of Econometrics* 65:109–54.

———. 1994. "Changes in the U.S. Wage Structure 1963–1987: Application of Quantile Regression." *Econometrica* 62:405–58.

———. 1991. "Methodological Issues of Quantile Regression." In: *The Theory and Practice of Quantile Regression*, Dissertation, Harvard University.

Buchinsky, M., and J. Hahn. 1996. "An Alternative Estimator for the Censored Quantile Regression Model." *Econometrica*. Forthcoming.

Chamberlain, G. 1994. "Quantile Regression, Censoring, and the Structure of Wage." In *Proceeding of the Sixth World Congress of the Econometric Society*, ed. C. Sims. New York: Cambridge University Press.

Efron, B. 1982. *The Jackknife, the Bootstrap and Other Resampling Plans*. Philadelphia: Society for Industrial and Applied Mathematics.

———. 1979. "Bootstrap Methods: Another Look at the Jackknife." *The Annals of Statistics* 7:1–26.

Evans, W., and R. Schwab. 1997. "Who Benefits from Catholic Education?" Unpublished. Department of Economics, University of Maryland.

Franklin, J. 1980. *Methods of Mathematical Economics* New York: Springer-Verlag.

Freedman, D. 1981. "Bootstrapping Regression Models." *The Annals of Statistics* 9:1218–28.

Hansen, L. P. 1982. "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica* 50:1029–54.

Honoré, B., and J. Powell. 1993. "Quantile Regression Under Random Censoring." Unpublished. Department of Economics, Princeton University.

Horowitz, J. L., and G. R. Neumann. 1987. "Semiparametric Estimation of Employment Duration Models." *Econometric Reviews* 6:5–40.
Huber, P. J. 1981. *Robust Statistics*. New York: John Wiley and Sons.
———. 1967. "The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions." *Proceedings of the Fifth Berkeley Symposium* Vol. 1. 221–33. Berkeley: University of California Press.
Koenker, R., and G. Bassett. 1982. "Robust Tests for Heteroscedasticity Based on Regression Quantiles." *Econometrica* 50:43–61.
———. 1978a. "The Asymptotic Distribution of the Least Absolute Error Estimator." *Journal of the American Statistical Association* 73:618–22.
———. 1978b. "Regression Quantiles." *Econometrica* 46:33–50.
Koenker, R., and V. D'Orey. 1987. "Computing Regression Quantiles." *Journal of the Royal Statistical Society, Applied Statistics* 36:383–93.
Koenker, R., and B. J. Park. 1996. "An Interior Point Algorithm for Nonlinear Quantile Regression." *Journal of Econometrics* 71:265–83.
Manski, C. 1988. *Analog Estimation Methods in Econometrics*. New York: Chapman and Hall.
Mincer, J. 1974. *Schooling, Experience and Earnings*. New York: National Bureau of Economic Research.
Mood, A., F. Graybill, and D. Boes. 1974. *Introduction to the Theory of Statistics*. New York: McGraw-Hill.
Newey, W., and J. Powell. 1990. "Efficient Estimation of Linear and Type I Censored Regression Models Under Conditional Quantile Restrictions." *Econometric Theory* 6:295–317.
———. 1987a. "Asymmetric Least Squares Estimation and Testing." *Econometrica* 55:819–47.
Portnoy, S., and R. Koenker. 1989. "Adaptive L-Estimation for Linear Models." *The Annals of Statistics* 17:362–81.
Poterba, J. M., and K. S. Rueben. 1994. "The Distribution of Public Sector Wage Premia: New Evidence Using Quantile Regression Methods." NBER Working Paper No. 4734.
Powell, J. 1986. "Censored Regression Quantiles." *Journal of Econometrics* 32:143–55.
———. 1984. "Least Absolute Deviation Estimation for the Censored Regression Model." *Journal of Econometrics* 25:303–25.
Robers, P. D., and A. Ben-Israel. 1969. "An Interval Programming Algorithm for Discrete l_1 Approximation Problems." *Journal of Approximation Theory* 2:323–36.
Robers, P. D., and S. S. Robers. 1970. "Discrete Linear l_1 Approximation by Interval Linear Programming." Unpublished.
Weiss, A. A. 1991. "Estimating Nonlinear Dynamic Models Using Least Absolute Error Estimation." *Econometric Theory* 7:46–68.
Womersley, R. S., 1986. "Censored Discrete Linear l_1 Approximation." *SIAM, Journal of Scientific and Statistical Computing* 7:105–22.