

NBER WORKING PAPER SERIES

QUANTILE REGRESSION UNDER MISSPECIFICATION,  
WITH AN APPLICATION TO THE U.S. WAGE STRUCTURE

Joshua Angrist  
Victor Chernozhukov  
Iván Fernández-Val

Working Paper 10428  
<http://www.nber.org/papers/w10428>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
April 2004

We thank David Autor, Gary Chamberlain, George Deltas, Jinyong Hahn, Jerry Hausman, Roger Koenker, and Art Lewbel for helpful discussions, and seminar participants at BYU, the University of Michigan, Michigan State University, the Harvard-MIT Econometrics Workshop, the University of Toronto, the University of Illinois at Urbana- Champaign, and the 2004 Winter Econometric Society Meetings for comments. The views expressed herein are those of the author(s) and not necessarily those of the National Bureau of Economic Research.

©2004 by Joshua Angrist, Victor Chernozhukov, and Iván Fernández-Val. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure  
Joshua Angrist, Victor Chernozhukov, and Iván Fernández-Val  
NBER Working Paper No. 10428  
April 2004  
JEL No. J31, C13, C14

### ABSTRACT

Quantile regression (QR) fits a linear model for conditional quantiles, just as ordinary least squares (OLS) fits a linear model for conditional means. An attractive feature of OLS is that it gives the minimum mean square error linear approximation to the conditional expectation function even when the linear model is misspecified. Empirical research using quantile regression with discrete covariates suggests that QR may have a similar property, but the exact nature of the linear approximation has remained elusive. In this paper, we show that QR can be interpreted as minimizing a weighted mean-squared error loss function for specification error. The weighting function is an average density of the dependent variable near the true conditional quantile. The weighted least squares interpretation of QR is used to derive an omitted variables bias formula and a partial quantile correlation concept, similar to the relationship between partial correlation and OLS. We also derive general asymptotic results for QR processes allowing for misspecification of the conditional quantile function, extending earlier results from a single quantile to the entire process. The approximation properties of QR are illustrated through an analysis of the wage structure and residual inequality in US Census data for 1980, 1990, and 2000. The results suggest continued residual inequality growth in the 1990s, primarily in the upper half of the wage distribution and for college graduates.

Joshua Angrist  
MIT  
Department of Economics  
50 Memorial Drive  
Cambridge, MA 02142-1347  
and NBER  
angrist@mit.edu

Victor Chernozhukov  
MIT  
Department of Economics  
50 Memorial Drive  
Cambridge, MA 02142-1347  
vchern@mit.edu

Iván Fernández-Val  
MIT  
Department of Economics  
50 Memorial Drive  
Cambridge, MA 02142-1347  
ifern12@mit.edu

# 1 Introduction

The Quantile Regression (QR) estimator, introduced by Koenker and Bassett (1978), is an increasingly important empirical tool, allowing researchers to fit parsimonious models to an entire conditional distribution. Part of the appeal of quantile regression derives from a natural parallel with conventional ordinary least squares (OLS) or mean regression. Just as OLS regression coefficients offer convenient summary statistics for conditional expectation functions (CEF), quantile regression coefficients can be used to make easily interpreted statements about conditional distributions. Moreover, unlike OLS coefficients, QR estimates capture changes in distribution shape and spread, as well as changes in location.

An especially attractive feature of OLS regression estimates is their robustness and interpretability under misspecification of the CEF. In addition to consistently estimating a linear CEF, OLS estimates provide the minimum mean square error (MMSE) linear approximation to a CEF of any shape. The MMSE interpretation of OLS is emphasized by Chamberlain (1984) and Goldberger (1991), while an average derivative interpretation of OLS features in Angrist and Krueger (1999). This robustness property – i.e., the fact that OLS provides a meaningful and well-understood summary statistic for multivariate conditional expectations under almost all circumstances – undoubtedly contributes to the primacy of OLS regression as an empirical tool. In view of the possibility of interpretation under misspecification, modern theoretical research on regression inference typically also allows for misspecification when deriving limiting distributions (see, e.g., White, 1980).

While QR estimates are as easy to compute as OLS regression coefficients, an important difference between OLS and QR is that most of the theoretical and applied work on QR postulates a true linear model for conditional quantiles. This raises the question of whether and how QR estimates can be interpreted when the linear model for conditional quantiles is misspecified (for example, QR estimates at different quantiles may imply conditional quantile functions that cross). One interpretation for QR under misspecification is that it provides the best linear predictor for a response variable under asymmetric loss. This interpretation is not very satisfying, however, since prediction under asymmetric loss is typically not the object of interest in empirical work (see, e.g., Koenker and Hallock, 2001).<sup>1</sup> Empirical research on quantile regression with discrete covariates suggests that QR may have an approximation property similar to that of OLS, but the exact nature of the linear approximation has remained an important unresolved

---

<sup>1</sup>An exception is the forecasting literature; see, e.g., Giacomini and Komunjer (2003).

question (cf. Chamberlain, 1994, p. 181).

The first contribution of this paper is to show that QR can be interpreted as the best linear predictor (BLP) for the conditional quantile function (CQF) using a weighted mean-squared error loss function, much as OLS regression provides a MMSE fit to the CEF. The implied QR weighting function can be used to understand which, if any, parts of the distribution of regressors contribute disproportionately to a particular set of QR estimates. We also show how the weighted mean-square error interpretation can be used to interpret QR coefficients as partial quantile correlation coefficients and to develop an omitted variable bias formulae for QR.

A second contribution is to develop a distribution theory for the entire QR process that applies under misspecification of the conditional quantile function. The approach developed here has two advantages over current practice. First, we do not assume that the true quantile function is linear. Second, some of the regularity conditions that would be required for a fully nonparametric approach, such as multiple differentiability of the quantile function in regressors and continuity of regressors, are not needed. Our analysis of the QR process extends the results of Chamberlain (1994) and Hahn (1997), who derived the basic variance formula for a particular quantile under misspecification. See also Koenker and Machado (1999), Gutenbrunner and Jureckova (1992), and Gutenbrunner, Jureckova, Koenker, Portnoy (1993), who develop inference procedures based on QR processes for the linear location shift model and linear Pitman deviations from this model.

An important consequence of our analysis is that the currently used inference tools on the QR process, such as those in Koenker and Machado (1999), are not robust to misspecification. This is because the limit distribution of the QR process is not distribution-free under misspecification. Moreover, Khmaladzatian techniques, as in Bai (1998) and Koenker and Xiao (2002), cannot restore the distribution-free nature of the limit theory in this case. We therefore suggest alternative methods that provide valid inference for the QR process under misspecification.

The approximation theorems and other theoretical ideas in the paper are illustrated with an analysis of wage data from the 1980, 1990, and 2000 U.S. censuses. The analysis here is motivated by similar studies in labor economics, where quantile regression has been widely used to model changes in the wage distribution (see, e.g., Buchinsky, 1994 and Autor, Katz, and Kearney, 2004 for the US; Gosling, Machin, and Meghir, 2000, for the UK; Abadie, 1997, for Spain, and Machado and Mata, 2003, for Portugal). In particular, we show that quantile regression, while an inexact model for conditional quantiles, gives a good account of the relevant stylized facts. An appealing feature of quantile regression in this context is that quantile regression coefficients

can be used directly to describe “residual inequality,” i.e. the spread in the wage distribution conditional on the variables included in the quantile regression model. Attempts to model residual wage inequality have been of major substantive importance to labor economists since Juhn, Murphy, and Pierce (1993).

The paper is organized as follows. Section 2 introduces assumptions and notation and presents the main approximation theorems, followed by an empirical illustration. Section 3 provides the inference theory for QR processes under misspecification. Section 4 presents additional empirical results on the evolution of residual inequality using data from the 1980, 1990, and 2000 censuses. Section 5 concludes with a brief summary.

## 2 Interpreting QR Under Misspecification

### 2.1 Notation and Framework

Given a continuous response variable  $Y$  and a  $d \times 1$  regressor vector  $X$ , we are interested in the (population) conditional quantile function (CQF) of  $Y$  given  $X$ . The conditional quantile function is defined as:

$$Q_\tau(Y|X) \equiv \inf \{y : F_Y(y|X) \geq \tau\}, \quad (1)$$

where  $F_Y(y|X)$  is the distribution function for  $Y$  conditional on  $X$ , with associated conditional density  $f_Y(y|X)$ . The CQF can also be defined as the solution to the following minimization problem (assuming integrability throughout where needed):

$$Q_\tau(Y|X) \equiv \arg \min_{q(X)} E [\rho_\tau(Y - q(X))], \quad (2)$$

where  $\rho_\tau(u) = (\tau - 1(u \leq 0))u$  and the minimum is over the set of measurable functions of  $X$ . This is a potentially infinite-dimensional problem if covariates are continuous, and can be very high-dimensional even with discrete  $X$ . It may nevertheless be possible to capture important features of the CQF using a linear model. This motivates linear quantile regression.

The Koenker and Bassett (1978) linear quantile regression (QR) estimator solves the following population minimization problem:

$$\beta(\tau) \equiv \arg \min_{\beta \in \mathbb{R}^d} E [\rho_\tau(Y - X'\beta)]. \quad (3)$$

If  $q(X)$  is in fact linear, the QR minimand will find it (just as if the CEF is linear, OLS regression will find it). More generally, QR provides the best linear predictor for  $Y$  under the asymmetric

loss function,  $\rho_\tau$ . As noted in the introduction, however, prediction under asymmetric loss is rarely the object of empirical work. Rather, the conditional quantile function is of intrinsic interest. For example, labor economists are often interested in comparisons of conditional deciles as a measure of how the spread of a wage distribution changes conditional on covariates, as in Katz and Murphy (1992) and Juhn, Murphy, and Pierce (1993). Thus, we would like to establish the nature of approximation that QR provides.

## 2.2 The QR Approximation Property

Our principal theoretical result is that the population QR vector minimizes a weighted sum of squared specification errors. This is easiest to show using notation for a quantile-specific specification error and for a quantile-specific residual. For a given quantile  $\tau$ , we define the QR specification error as:

$$\Delta_\tau(X, \beta) = X'\beta - Q_\tau(Y|X). \quad (4)$$

Similarly, let  $\epsilon_\tau$  be a quantile-specific residual, defined as the deviation of the response variable from the conditional quantile of interest:

$$\epsilon_\tau = Y - Q_\tau(Y|X), \quad (5)$$

with conditional density  $f_{\epsilon_\tau}(e|X)$  at  $\epsilon_\tau = e$ . The following theorem shows that QR is the weighted least squares approximation to the unknown CQF.

**Theorem 1 (Approximation Property)** *Suppose that (i) the conditional density  $f_Y(y|X)$  exists a.s., (ii)  $Q_\tau(Y|X)$  is uniquely defined by (2), and (iii)  $\beta(\tau)$  is uniquely defined by (3).*

*Then*

$$\beta(\tau) = \arg \min_{\beta \in \mathbb{R}^d} E [w_\tau(X, \beta) \cdot \Delta_\tau^2(X, \beta)], \quad (P1)$$

*where*

$$w_\tau(X, \beta) = \int_0^1 (1-u) f_{\epsilon_\tau}(u\Delta_\tau(X, \beta)|X) du \quad (6)$$

$$= \int_0^1 (1-u) \cdot f_Y(u \cdot X'\beta + (1-u) \cdot Q_\tau(Y|X)|X) du > 0. \quad (7)$$

This result says that the population QR coefficient vector  $\beta(\tau)$  minimizes the expected weighted mean squared approximation error, i.e. the square of the difference between the true CQF and the linear approximation, with weighting function  $w_\tau(X, \beta)$ . The weights involve an integral in either the conditional density of the quantile residual, or, by a change of variables using

$Y = Q_\tau(Y|X) + \epsilon_\tau$ , the conditional density of the response variable. The latter representation shows the weighting function to be given by the average density of the response variable over a line from the point of approximation,  $X'\beta$ , to the true conditional quantile,  $Q_\tau(Y|X)$ . Pre-multiplication by the term  $(1 - u)$  in the integral results in more weight being applied at points on the line closer to the true CQF.

We refer to the function  $w_\tau(X, \beta)$  as defining *importance weights*, since this function determines the importance the QR minimand gives to points in the support of  $X$  for a given distribution of  $X$ . In addition to the importance weights, the probability distribution of  $X$  also determines the ultimate weight given to different values of  $X$  in the least squares problem. To see this, note that we can also write the QR minimand as

$$\beta(\tau) = \arg \min_{\beta \in \mathbb{R}^d} \int w_\tau(x, \beta) \cdot \Delta_\tau^2(x, \beta) dP(x), \quad (8)$$

where  $P(x)$  is the CDF of  $X$  (with associated probability or density function  $p(x)$ ). Thus, the overall weight varies in the distribution of  $X$  according to

$$w_\tau(x, \beta) \cdot p(x). \quad (9)$$

The sense in which QR approximates a nonlinear CQF can be seen for an empirical wage equation in Figure 1. This figure plots an estimate of the CQF for log-earnings given education for the 0.10, 0.25, 0.50, 0.75 and 0.90 quantiles, using data for US-born black and white men aged 40-49 from the 1980 census (see Appendix for details concerning data). Here we take advantage of the discreteness of the schooling variable and the large census sample to compare QR estimates with the true (sample) CQF evaluated at each point in the support of  $X$ . In addition to the dots plotting  $Q_\tau(Y|X)$  against  $X$ , the figure also shows the (solid) QR regression line.

To compare the consequences of combined importance- and histogram-weighting, as in Theorem 1, to a weighting scheme using the  $X$  histogram only, the figure also shows a graphical representation of a minimum distance (MD) estimator suggested by Chamberlain (1994). The MD estimator is the sample analog of the vector  $\tilde{\beta}(\tau)$  solving

$$\tilde{\beta}(\tau) = \arg \min_{\beta \in \mathbb{R}^d} E [(Q_\tau(Y|X) - X'\beta)^2] = \arg \min_{\beta \in \mathbb{R}^d} E [\Delta_\tau^2(X, \beta)]. \quad (10)$$

In other words,  $\tilde{\beta}(\tau)$  is the slope of the linear regression of  $Q_\tau(Y|X)$  on  $X$ , weighted only by the probability distribution of  $X$ ,  $p(x)$ . The dashed line in the figure has the slope determined by Chamberlain's estimator. Note that unlike QR, the MD estimator relies on the ability to

nonparametrically estimate  $Q_\tau(Y|X)$  in a nonparametric first step. This is facilitated here by the discreteness of  $X$  and our large census samples, but would otherwise require additional restrictions and regularity conditions. Chamberlain (1994) observes that, in general, the MD estimator is likely to be attractive only when  $X$  is low dimensional and the sample size is large.

For every quantile, the QR and MD regression lines are remarkably close, supporting the conclusion reached in Theorem 1 – that QR is a weighted MD approximation to the unknown CQF – and suggesting the extra weighting by the importance weights  $w_\tau(x, \beta)$  does not induce big differences between MD and QR. In fact, for some quantiles, the MD and QR lines are not discernible different. Under either weighting scheme, the linear fits appear to describe the actual conditional quantiles reasonably well.

By way of comparison and to provide a visual standard for the goodness of fit of QR to the CQF, the figure also incorporates a panel illustrating the fit of an OLS regression line to the CEF. This panel (bottom, right position in the figure) shows points on the CEF plotted as dots, along with the dashed OLS regression line and the solid generalized least squares (GLS) regression line. To compute the GLS slope,  $E[Y|X]$  was regressed on  $X$ , weighted by the inverse of the conditional variance of  $Y$  given  $X$ . The OLS fit to the CEF is similar to the QR fit to the CQF at the median. The estimated median QR and OLS regression slopes are also similar, at 6.39 and 6.98 in percentage terms. Panel A of Table 1 reports the slopes of the lines plotted in each panel of Figure 1.

To further investigate the nature of the QR weighting function in the schooling example, Figure 2 plots  $w_\tau(x, \beta(\tau))p(x)$  against the regressor  $X$ . The solid line in the figure shows the product  $w_\tau(x, \beta(\tau))p(x)$ , along with the histogram of education,  $p(x)$ , the weights used in the Chamberlain MD estimator. The figure also shows normalized kernel density estimates of the importance weights,  $w_\tau(x, \beta(\tau))$ , plotted with a dashed line.<sup>2</sup> Consistent with the comparison of estimators in Figure 1, the importance weights are reasonably flat for the quantiles considered here, so that most of the variation in the overall weighting function comes from the  $X$  histogram. As in Figure 1, Figure 2 again includes an analogous panel for mean regression and the CEF. The CEF analog of the QR importance weights is the inverse of  $V[Y|X]$ , since the latter plays the role of importance-weighting in GLS estimation. Here too, the importance weighting function is reasonably flat.

---

<sup>2</sup>See Appendix B for a detailed description of the procedure used for kernel density estimation of the weights.



### 2.3 Conditional Density as Primary Determinant of Importance Weights

What features of the joint distribution of  $Y$  and  $X$  determine the theoretical shape of the importance weighting function for QR? Suppose initially that the linear model for conditional quantiles is correct, so the approximation error is zero and  $\Delta_\tau(X, \beta(\tau)) = 0$ . In this case, the weighting function when evaluated at  $\beta = \beta(\tau)$  simplifies to

$$w_\tau(X, \beta(\tau)) = 1/2 \cdot f_Y(Q_\tau(Y|X)|X), \quad (11)$$

i.e., the weights are proportional to the conditional density of the response variable at the relevant conditional quantile. More generally, for response data with a smooth conditional density around the relevant quantile, we have for  $\beta$  in the neighborhood of  $\beta(\tau)$ :

$$w_\tau(X, \beta) = 1/2 \cdot f_Y(Q_\tau(Y|X)|X) + r_\tau(X), \quad \text{where } |r_\tau(X)| \leq 1/6 \cdot |\Delta_\tau(X, \beta)| \cdot |\bar{f}'|. \quad (12)$$

Here,  $r_\tau(X)$  is a remainder term and the density  $f_Y(y|X)$  has a first derivative in  $y$  bounded a.s. by a constant, denoted  $\bar{f}'$ .<sup>3</sup> This argument demonstrates that we can in most cases think of the *density weights*  $1/2 \cdot f_Y(Q_\tau(Y|X)|X)$  as being the primary determinant of the importance weights.<sup>4</sup> This interpretation applies when the degree of misspecification is modest or the variability of conditional density  $f_Y(y|X)$  in  $y$  near the true CQF is not substantial.

For the empirical example considered in this section, the weighting function  $w_\tau(X, \beta(\tau))$  and the density-based approximation are remarkably close. This can be seen in Figure 3, which plots estimates of both importance and density weights constructed using a kernel method. The previous argument suggests there are two reasons for this: the approximation error  $\Delta_\tau(X, \beta(\tau))$  is mostly small and the conditional density  $f_Y(y|X)$  does not vary much in  $y$  near the true quantiles. The figure also shows that both the weighting function and its first order approximation are fairly stable, suggesting that the conditional density of  $y$  is stable across the levels of  $X$  at each quantile.

---

<sup>3</sup>The remainder term is  $|r_\tau(X)| = w_\tau(X, \beta) - \frac{1}{2} \cdot f_{\epsilon_\tau}(0|X) = \int_0^1 (1-u)(f_{\epsilon_\tau}(u \cdot \Delta_\tau(X, \beta)|X) - f_{\epsilon_\tau}(0|X)) du \leq |\Delta_\tau(X, \beta)| \cdot \bar{f}' \cdot \int_0^1 (1-u) \cdot u \cdot du = \frac{1}{6} \cdot |\Delta_\tau(X, \beta)| \cdot \bar{f}'$ .

<sup>4</sup>Powell (1994, p. 2473) notes that an efficient weighted QR estimator (in the sense of attaining the relevant semiparametric efficiency bound) is obtained by weighting the original Koenker and Bassett QR minimand by  $f_{\epsilon_\tau}(0|X)$ . Since the variance of the sample analog of  $Q_\tau(Y|X)$  is proportional to  $1/f_{\epsilon_\tau}^2(0|X)$ , Powell's estimator is equivalent to a GLS (efficient) estimator for conditional quantiles under correct specification. The first order asymptotic equivalence of the GLS fit and Powell's estimator under correct specification is noted by Knight (2002).

## 2.4 Partial Quantile Correlation

The least squares interpretation of QR has a practical payoff in that we can use it to develop a regression-decomposition scheme and an omitted variables bias formula for QR. The idea here is to express each QR coefficient as a coefficient in a bivariate LS projection of the unknown CQF on each regressor, after the effects of other regressors have been “partialled out.” Since these derivations rely on least-squares algebra, a pre-requisite for the development of this decomposition is a version of the LS approximation property with weights that are fixed in the optimization problem. This version of the QR minimand is given in the theorem below.

**Theorem 2 (Iterative Approximation Property)** *Under the conditions of Theorem 1, QR coefficients satisfy the equation*

$$\beta(\tau) = \arg \min_{\beta \in \mathbb{R}^d} E [\tilde{w}_\tau(X) \cdot \Delta_\tau^2(X, \beta)], \quad (\text{P2})$$

where

$$\tilde{w}_\tau(X) = \frac{1}{2} \int_0^1 f_{e_\tau}(u \cdot \Delta_\tau(X, \beta(\tau)) | X) du \quad (13)$$

$$= \frac{1}{2} \int_0^1 f_Y(u \cdot X' \beta(\tau) + (1 - u) Q_\tau(Y | X) | X) du. \quad (14)$$

Theorem 2 differs from Theorem 1 in that the weights are defined *ex post*, i.e., they are defined using the solution vector to the QR problem. Theorem 2 complements Theorem 1 in that it characterizes the QR coefficient as a fixed point to an iterated minimum distance approximation.<sup>5</sup> The relationship between the weighting functions in Theorems 1 and 2 is analogous to the relationship between the weights used to compute a continuously updated GMM Estimator and the corresponding iterated estimator (see Hansen, Heaton, and Yaron, 1996).

The weighting function  $\tilde{w}_\tau(X)$  is again related to the conditional density of the dependent variable. In particular, for a response variable with smooth conditional density around the relevant quantile, we have

$$\tilde{w}_\tau(X) = 1/2 \cdot f_Y(Q_\tau(Y | X) | X) + \tilde{r}_\tau(X), \quad \text{where } |\tilde{r}_\tau(X)| \leq 1/4 \cdot |\Delta_\tau(X, \beta(\tau))| |\tilde{f}'|, \quad (15)$$

---

<sup>5</sup>In other words, given weights defined in terms of  $\beta(\tau)$ , the solution to the weighted minimum distance approximation is  $\beta(\tau)$ . It is easy to show that this fixed point property defines  $\beta(\tau)$  uniquely whenever  $\beta(\tau)$  is the the unique solution to the original QR problem.

where  $\tilde{r}_\tau(X)$  is a remainder term and the density  $f_Y(y|X)$  has a first derivative in  $y$  bounded a.s. by a constant, denoted  $\bar{f}'$ .<sup>6</sup> When either  $\Delta_\tau(X, \beta(\tau))$  or  $\bar{f}'$  is small,

$$\tilde{w}_\tau(X) \approx w_\tau(X, \beta(\tau)) \approx \frac{1}{2} f_Y(Q_\tau(Y|X)|X), \quad (16)$$

so the approximate weighting function is the same as before when the QR coefficient vector is evaluated at its solution value.

Partial quantile correlation is defined with regard to a partition of the regression vector into a variable,  $X_1$ , and the remaining  $d - 1$  variables  $X_2$ , along with the corresponding partition of the QR coefficients. Thus,

$$X = [X_1, X_2']', \quad \beta(\tau) = (\beta_1(\tau), \beta_2(\tau)')'. \quad (17)$$

We can now decompose  $Q_\tau(Y|X)$  and  $X_1$  using orthogonal projections onto  $X_2$  weighted by  $\tilde{w}_\tau(X)$ , just as can be done for weighted least squares mean regression:

$$Q_\tau(Y|X) = X_2' \pi_Q + q_\tau(Y|X), \text{ such that } E[\tilde{w}_\tau(X) \cdot X_2 \cdot q_\tau(Y|X)] = 0, \quad (18)$$

$$X_1 = X_2' \pi_1 + V_1, \text{ such that } E[\tilde{w}_\tau(X) \cdot X_2 \cdot V_1] = 0. \quad (19)$$

In this decomposition,  $q_\tau(Y|X)$  and  $V_1$  are residuals created by a weighted linear projection of the CQF,  $Q_\tau(Y|X)$ , and  $X_1$  on  $X_2$ , respectively, using  $\tilde{w}_\tau(X)$  as weight.<sup>7</sup> Then, standard mathematics for least squares gives

$$\beta_1(\tau) = \arg \min_{\beta_1} E [\tilde{w}_\tau(X) (q_\tau(Y|X) - V_1 \beta_1)^2], \quad (20)$$

and also

$$\beta_1(\tau) = \arg \min_{\beta_1} E [\tilde{w}_\tau(X) (Q_\tau(Y|X) - V_1 \beta_1)^2]. \quad (21)$$

This shows that  $\beta_1(\tau)$  can be interpreted as the ‘‘partial quantile correlation coefficient’’ in the sense that it can be obtained from a regression of the CQF,  $Q_\tau(Y|X)$ , on  $X_1$ , once we have partialled out the effect of  $X_2$ . Both the partialling-out and second-step regressions are weighted by the QR weighting function.

Figure 4 shows partial quantile correlation plots for the effect of schooling on wages, adjusting for the effect of a quadratic function of potential experience.<sup>8</sup> For this example the sample age

<sup>6</sup>The remainder term is  $|\tilde{r}_\tau(X)| = \tilde{w}_\tau(X) - \frac{1}{2} \cdot f_{\epsilon_\tau}(0|X) = \frac{1}{2} \int_0^1 (f_{\epsilon_\tau}(u \cdot \Delta_\tau(X, \beta(\tau))|X) du - f_{\epsilon_\tau}(0|X)) du \leq \frac{1}{2} \cdot |\Delta_\tau(X, \beta(\tau))| \cdot \bar{f}' \cdot \int_0^1 u \cdot du = \frac{1}{4} \cdot |\Delta_\tau(X, \beta(\tau))| \cdot \bar{f}'$ .

<sup>7</sup>Thus,  $\pi_Q = E [\tilde{w}_\tau(X) X_2 X_2']^{-1} E [\tilde{w}_\tau(X) X_2 Q_\tau(Y|X)]$  and  $\pi_1 = E [\tilde{w}_\tau(X) X_2 X_2']^{-1} E [\tilde{w}_\tau(X) X_2 X_1]$ .

<sup>8</sup>Potential experience is defined in the standard way as age - years of schooling - 6.

range is extended to 30-54 to increase the range of variation of potential experience, and the sample is restricted to white men.<sup>9</sup> The points in the figure correspond to the scatterplot of the partial residuals of the CQF of log-earnings and schooling for the 0.10, 0.25, 0.50, 0.75 and 0.90 quantiles, i.e.  $q_\tau(Y|X)$  plotted against  $V_1$ ; while the solid line represents the partial QR slope. In this example, the partial CQF of log-earnings given schooling looks to be close to linear for every quantile. The dashed line is a counterfactual QR with the same slope as for schooling without controls. As for conventional least squares estimates (see bottom right panel), the omission of experience causes downward bias in the coefficient of schooling for every quantile, since experience and schooling are negatively correlated and experience raises wages.

## 2.5 Omitted Variables Bias

The previous discussion suggests we can use a reasoning process much like that for OLS when analyzing omitted variables bias in the context of QR. Here we use the least squares interpretation of QR to construct formal relationship between “long” QR coefficients and “short” QR coefficients. In particular, suppose we are interested in a quantile regression with explanatory variables  $X = [X'_1, X'_2]'$ , but  $X_2$  is not available, e.g. ability in the wage equation. We run QR on  $X_1$  only, obtaining the coefficient vector

$$\gamma_1(\tau) = \arg \min_{\gamma_1} E[\rho_\tau(Y - X'_1\gamma_1)]. \quad (22)$$

The long regression coefficient vectors are  $(\beta_1(\tau), \beta_2(\tau))$ , defined by

$$(\beta_1(\tau)', \beta_2(\tau)')' = \arg \min_{\beta_1, \beta_2} E[\rho_\tau(Y - X'_1\beta_1 - X'_2\beta_2)]. \quad (23)$$

Finally, it is useful to define a remainder term

$$R_\tau(X) = Q_\tau(Y|X) - X'_1\beta_1(\tau), \quad (24)$$

equal to the residual of the CQF, given both  $X_1$  and  $X_2$ , not explained by the linear function of  $X_1$  in the long QR. If the CQF is linear, then  $R_\tau(X) = X'_2\beta_2(\tau)$ .

The following theorem describes the relationship between  $\gamma_1(\tau)$  and  $\beta_1(\tau)$ .

**Theorem 3 (Long and Short Coefficients)** *Suppose that the conditions of Theorem 1 hold and  $\gamma_1(\tau)$  is uniquely defined by (22). Then, 1.*

$$\gamma_1(\tau) = \arg \min_{\gamma_1} E[\tilde{w}_\tau^*(X) \cdot \Delta_\tau^2(X, \gamma_1)], \quad (25)$$

---

<sup>9</sup>The inclusion of black men complicates estimation of the weights and CQF because of small cells.

where  $\Delta_\tau(X, \gamma_1) = X_1' \gamma_1 - Q_\tau(Y|X)$ ,  $\epsilon_\tau = Y - Q_\tau(Y|X)$ , and

$$\tilde{w}_\tau^*(X) = \frac{1}{2} \int_0^1 f_{\epsilon_\tau}(u \cdot \Delta_\tau(X, \gamma_1(\tau)) | X) du. \quad (26)$$

2. If  $E[\tilde{w}_\tau^*(X) \cdot X_1 X_1']$  is invertible,  $\gamma_1(\tau) = \beta_1(\tau) + B_1(\tau)$ , where

$$B_1(\tau) = E[\tilde{w}_\tau^*(X) \cdot X_1 X_1']^{-1} E[\tilde{w}_\tau^*(X) \cdot X_1 R_\tau(X)]. \quad (27)$$

As with OLS short and long calculations, the omitted variables formula in this case shows the short QR coefficients to be equal to the corresponding long QR coefficients plus the coefficients in a weighted projection of omitted effects on included variables. While the parallel with OLS seems clear, there are two complications in the QR case. First, the effect of omitted variables appears through the remainder term,  $R_\tau(X)$ . In practice, it seems reasonable to think of this as being approximated by the omitted linear part,  $X_2' \beta_2(\tau)$ . Second, the regression of omitted variables on included variables is weighted by  $\tilde{w}_\tau^*(X)$ , while for OLS it is unweighted.<sup>10</sup>

### 3 Large Sample Properties of QR and Robust Inference Under Misspecification

In this section, we study the consequences of misspecification for large sample inference on the quantile regression process

$$\hat{\beta}(\tau) = \arg \min_{\beta \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^n \rho_\tau(Y_i - X_i' \beta), \quad \tau \in \mathcal{T} \equiv \text{closed subinterval of } (0, 1). \quad (28)$$

The QR process  $\hat{\beta}(\bullet)$ , viewed as a function of the probability index  $\tau$ , is a regression generalization of the quantile processes and quantile-quantile plots used in univariate and two-sample treatment control problems, cf. Doksum (1974). To see this, suppose the regressor is a dummy for receiving a treatment, denoted  $D$ , so we have  $X = (1, D)'$ . Then, the components of the quantile regression process  $\hat{\beta}(\bullet) = (\hat{\beta}_1(\bullet), \hat{\beta}_2(\bullet))'$  measure easily interpreted quantities. In particular, the intercept  $\hat{\beta}_1(\bullet)$  measures the quantile function in the control group, and the slope  $\hat{\beta}_2(\bullet)$  measures the quantile treatment effect (as a function of the probability  $\tau$ ). When the regressors are continuous,  $\hat{\beta}_2(\bullet)$  measures the quantile treatment effect as a response to a unit

---

<sup>10</sup>Note that the omitted variables bias formula derived here can be used to determine the bias from measurement error in regressors, by identifying the error as the omitted variable. For example, classical measurement error is likely to generate an attenuation bias in QR as well as OLS estimates. We thank Arthur Lewbel for pointing this out.

change in the treatment. Under misspecification, the QR slope process,  $\hat{\beta}_2(\bullet)$ , should be interpreted as approximating the quantile treatment effect, while  $\hat{\beta}_1(\bullet)$  approximates the quantile function in the control group, in the sense stated in Theorem 1.

Previous studies of the QR process  $\hat{\beta}(\bullet)$  focused on the linear location or scale shift models, or Pitman deviations from these models. See especially Koenker and Machado (1999), Gutenbrunner and Jureckova (1992), Gutenbrunner, Jureckova, Koenker, Portnoy (1993). The first purpose of this section is to extend previous limit theory for the QR process to allow for misspecification of any type. The second purpose is to analyze the consequences of misspecification for currently used inference tools, and derive inference procedures that remain valid under misspecification.

### 3.1 Basic Large Sample Properties

The following conditions are used to insure consistency:

**A.1**  $(Y_i, X_i, i \leq n)$  are iid on the probability space  $(\Omega, \mathcal{F}, P)$  for each  $n$ .

**A.2** The conditional density  $f_Y(y|X = x)$  exists  $P$ -a.s.

**A.3**  $E \|X\| < \infty$ , and for all  $\tau \in \mathcal{T}$ ,  $\beta(\tau)$  defined to solve

$$E [(\tau - 1\{Y \leq X'\beta(\tau)\})X] = 0 \tag{29}$$

is the unique solution in  $\mathbb{R}^d$ .

**Theorem 4 (Consistency of QR Process)** *Under conditions A.1-A.3,*

$$\sup_{\tau \in \mathcal{T}} \|\hat{\beta}(\tau) - \beta(\tau)\| = o_p(1). \tag{30}$$

The following additional conditions are imposed to obtain asymptotic normality:

**A.4** The conditional density  $f_Y(y|X = x)$  is bounded and uniformly continuous in  $y$ , uniformly in  $x$  over the support of  $(Y, X)$ .

**A.5**  $J(\tau) \equiv E f_Y(X'\beta(\tau)|X)XX'$  is positive definite and finite for all  $\tau$ , and  $E \|X\|^{2+\epsilon} < \infty$  for some  $\epsilon > 0$ .

**Theorem 5 (Gaussianity of QR Process)** *Under A.1-A.5, we have that*

$$J(\bullet)\sqrt{n}\left(\hat{\beta}(\bullet) - \beta(\bullet)\right) = -\frac{1}{\sqrt{n}}\sum_{i=1}^n (\bullet - 1\{Y_i \leq X_i'\beta(\bullet)\}) X_i + o_p(1)$$

*converges weakly to a tight zero mean Gaussian process  $z(\bullet)$ , in the space of bounded function  $\ell^\infty(\mathcal{T})$ , where  $z(\bullet)$  is defined by its covariance function  $\Sigma(\tau, \tau') \equiv E\{z(\tau)z(\tau)'\}$ , where*

$$\Sigma(\tau, \tau') = E[(\tau - 1\{Y < X'\beta(\tau)\})(\tau' - 1\{Y < X'\beta(\tau')\})XX']. \quad (31)$$

*When the model is correctly specified, i.e.  $Q_\tau(Y|X) = X'\beta(\tau)$ , then*

$$\Sigma(\tau, \tau') = \Sigma_0(\tau, \tau') \equiv [\min(\tau, \tau') - \tau\tau'] \cdot E[XX']. \quad (32)$$

*In general,  $\Sigma(\cdot, \cdot) \neq \Sigma_0(\cdot, \cdot)$ .*

The proof of this result (in the appendix) is of independent interest, since it does not rely on either convexity arguments, which are not applicable for the process case, or explicit chaining arguments, which are case-specific and therefore difficult to establish for all QR problems (see e.g. Portnoy, 1991). In contrast, the proof relies primarily on the fact that the functional class  $\{1\{Y \leq X'\beta\}, \beta \in \mathbb{R}^d\}$  is Donsker. Thus, the theorem easily extends to a wide range of cases where a uniform central limit theorem holds for this functional class. In particular, extensions to strong, uniformly mixing, and various Markovian data are immediate.

Theorem 5 allows for misspecification and imposes little structure on the underlying conditional quantile function  $Q_\tau(Y|X)$ . For example, smoothness of  $Q_\tau(Y|X)$  in  $X$ , which is needed to pursue the fully nonparametric estimation approach, is not needed. Theorem 5 also has important consequences for general inference on the QR process, since it implies that  $(EXX')^{-1}\Sigma(\tau, \tau')$  is not proportional to the covariance function of the standard  $d$ -dimensional Brownian bridge  $[\min(\tau, \tau') - \tau\tau'] \cdot I$ , unlike in the correctly specified case, where

$$(EXX')^{-1}\Sigma_0(\tau, \tau') = [\min(\tau, \tau') - \tau\tau'] \cdot I, \quad (33)$$

which in turn implies that the conventional inference methods developed in Koenker and Machado (1999) do not apply under misspecification. Moreover, the problem of a nonstandard covariance function cannot be alleviated by the Khmaladization techniques implemented in Koenker and Xiao (2002). We therefore rely on Theorem 5 to develop general inference methods on the QR process that are robust to misspecification.

An important though previously known corollary of Theorem 5 is that the conventional standard errors used for basic pointwise inference are not robust to misspecification. This follows

from the fact that the covariance kernel  $\Sigma(\tau, \tau')$  generally differs from  $\Sigma_0(\tau, \tau')$ . In particular, we have:

**Corollary 1 (Finite-Dimensional Limit Theory)** *Under A.1-A.5, for a finite collection  $\tau_k \in \mathcal{T}$ ,  $k = 1, \dots, K$ , the regression quantile statistics  $\sqrt{n}(\hat{\beta}(\tau_k) - \beta(\tau_k))$  are asymptotically jointly normal, with asymptotic variance given by  $J(\tau_k)^{-1}\Sigma(\tau_k, \tau_k)J(\tau_k)^{-1}$  and asymptotic covariance between the  $k$ -th and  $l$ -th subsets equal to  $J(\tau_k)^{-1}\Sigma(\tau_k, \tau_l)J(\tau_l)^{-1}$ . Under correct specification  $\Sigma(\cdot, \cdot)$  is replaced with  $\Sigma_0(\cdot, \cdot)$  in these expressions.*

Chamberlain (1994) and Hahn (1997) give this result for a single quantile, that is for a given  $\tau \in \mathcal{T}$ ,  $\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau)) \xrightarrow{d} N(0, V(\tau) \equiv J^{-1}(\tau)\Sigma(\tau, \tau)J^{-1}(\tau))$ .<sup>11</sup> Under correct specification the variance formula simplifies to  $V_0(\tau) \equiv J^{-1}(\tau)\tau(1 - \tau)E[XX']J^{-1}(\tau)$ . Hence commonly reported estimates of  $V_0(\tau)$  are inconsistent for  $V(\tau)$  under misspecification except for the median, i.e.  $\tau = 0.5$ . (In this case, the two formulae coincide because  $[\tau - 1\{Y \leq X'\beta(\tau)\}]^2 = 1/4 = \tau(1 - \tau)$  for  $\tau = 0.5$ ). Also, since the difference between  $V_0(\tau)$  and  $V(\tau)$  is

$$(1 - 2\tau) \cdot J(\tau)^{-1} \cdot E((1\{Y \leq X'\beta(\tau) - 1\{Y \leq Q_\tau(Y|X)\}) \cdot XX') \cdot J(\tau)^{-1}, \quad (34)$$

we have that, for the same degree of misspecification, the difference grows as we move away from the median and it can be positive or negative depending on the sign of specification error and its correlation with the elements of  $XX'$ . For example, if  $X$  is one-dimensional and  $Y$  is positive, then for  $\tau < 1/2$  the difference between  $V(\tau)$  and  $V_0(\tau)$  will be positive if the corresponding conditional quantile is lower than the linear approximation for higher absolute values of the regressor, and negative otherwise, i.e. if the conditional quantile is above the linear approximation for these values.

Table 1 illustrates these basic implications by reporting estimates of the schooling coefficients and their asymptotic standard errors, using the two alternative formulae  $V_0(\tau)$  and  $V(\tau)$ , for the empirical example considered in the previous section. Panel A reports QR and OLS coefficients from regressions of log-earnings on schooling for the 1980, 1990 and 2000 census samples, while Panel B presents the same schooling coefficients from a model that also controls for race and a quadratic function of potential experience. The standard errors were estimated using equations (44)-(46), below. The alternative estimates of the standard errors are fairly close, with the biggest differences for tail quantiles (0.10 and 0.90). Here, the commonly reported standard error is biased downwards since, for the high levels of schooling where misspecification is more severe, the conditional quantile is below the linear approximation for the 0.10 quantile, while it is above the linear approximation for the 0.90 quantile.

---

<sup>11</sup>See also Kim and White (2002).



### 3.2 Simultaneous (Uniform) Inference

An alternative to pointwise inference is Kolmogorov-type uniform inference on the QR process. Uniform inference provides a parsimonious strategy for the study of changes in an entire response distribution. Here we derive robust uniform confidence regions that allow us to simultaneously test, in the Scheffé sense, a variety of potentially multi-faceted hypotheses about conditional distributions without compromising significance levels. Examples include specification tests (omission of variables), stochastic dominance, constant treatment effects, and changes in distribution.

Of course, a finite number of quantile regression coefficients are always estimated in practice. Nevertheless, it is still convenient to treat the quantile-specific estimates as realizations of a stochastic process rather than as a large vector of parameters. To see this, consider the construction of joint confidence intervals for, say,  $d = 2$  of the coefficients from a quantile regression, estimated at  $K = 20$  different quantiles (i.e., increments of .05). The number of variance and covariance terms to be estimated is  $dK(dK + 1)/2 = 820$ . The functional limit result in Theorem 5 allow us to avoid this high-dimensional estimation problem.<sup>12</sup> This approach also leads to a convenient graphical inference procedure, illustrated below.

The simplest use of the quantile regression process is to test linear hypotheses of the form:

$$H_0 : R(\tau)' \beta(\tau) = r(\tau) \text{ for all } \tau \in \mathcal{T}. \quad (35)$$

For example, we might want to test whether the coefficient corresponding to the variable  $j$  is zero over the whole quantile process, i.e. whether

$$\beta_j(\tau) = 0 \text{ for all } \tau \in \mathcal{T}. \quad (36)$$

This corresponds to  $R(\tau) = [0, \dots, 1, \dots, 0]'$  with 1 in the  $j$ -th position and  $r(\tau) = 0$ . Similarly, we may want to construct uniform or simultaneous confidence intervals for parameters or for linear functions of parameters of the form

$$R(\tau)' \beta(\tau) - r(\tau) \quad \text{for all } \tau \in \mathcal{T}. \quad (37)$$

The following corollaries facilitate both hypotheses testing and the construction of confidence intervals in this framework:

---

<sup>12</sup>Formally, this is because the empirical quantile regression process  $\sqrt{n}(\hat{\beta}(\bullet) - \beta(\bullet))$  asymptotically behaves continuously, so that  $\sqrt{n}(\hat{\beta}(\bullet) - \beta(\bullet))$  is approximately equivalent to a large finite collection of regression quantiles  $\sqrt{n}(\beta(\tau_k) - \beta(\tau_k))$ ,  $k = 1, \dots, K$ , for a suitably fine grid of quantile indices  $\mathcal{T}_K = \{\tau_k, k = 1, \dots, K\} \subset \mathcal{T}$ .

**Corollary 2 (Kolmogorov Statistic)** *Under the conditions of Theorem 5, and (35), for any  $\hat{V}(\tau) = V(\tau) + o_p(1)$  uniformly in  $\tau \in \mathcal{T}$ ,*

$$\mathcal{K}_n = \sup_{\tau \in \mathcal{T}} \left| \left[ R(\tau)' \hat{V}(\tau) R(\tau) \right]^{-1/2} \sqrt{n} \left( R(\tau)' \hat{\beta}(\tau) - r(\tau) \right) \right| \rightarrow_d \mathcal{K}, \quad (38)$$

where  $|x|$  denotes the sup norm of a vector, i.e.  $|x| = \max_j |x_j|$ , and  $\mathcal{K}$  is a random variable with an absolutely continuous distribution, defined as

$$\mathcal{K} \equiv \sup_{\tau \in \mathcal{T}} \left| \left[ R(\tau)' V(\tau) R(\tau) \right]^{-1/2} R(\tau)' J(\tau)^{-1} z(\tau) \right|. \quad (39)$$

**Corollary 3 (General Uniform Inference)** *Then, for  $\kappa(\alpha)$  denoting the  $\alpha$ -quantile of  $\mathcal{K}$  and  $\hat{\kappa}(\alpha)$  any consistent estimate of it,*

$$\lim_{n \rightarrow \infty} P \left\{ \sqrt{n} \left( R(\tau)' \beta(\tau) - r(\tau) \right) \in \hat{I}_n(\tau), \text{ for all } \tau \in \mathcal{T} \right\} = 1 - \alpha, \quad (40)$$

where

$$\hat{I}_n(\tau) = \left[ u(\tau) : \left| \left[ R(\tau)' \hat{V}(\tau) R(\tau)' \right]^{-1/2} \sqrt{n} \left( R(\tau)' \hat{\beta}(\tau) - r(\tau) - u(\tau) \right) \right| \leq \hat{\kappa}(1 - \alpha) \right]. \quad (41)$$

For example, when  $R(\tau)' \beta(\tau) - r(\tau)$  is scalar, we have

$$\hat{I}_n(\tau) = \left[ R(\tau)' \hat{\beta}(\tau) - r(\tau) \pm \frac{\hat{\kappa}(1 - \alpha) \left[ R(\tau)' \hat{V}(\tau) R(\tau)' \right]^{1/2}}{\sqrt{n}} \right]. \quad (42)$$

The critical values  $\hat{\kappa}(\alpha)$  can easily be obtained by subsampling in cross-sectional applications of the sort considered here. Let  $I_1, \dots, I_B$  be  $B$  randomly chosen subsamples of  $(Y_i, X_i, i \leq n)$  of size  $b$ , where  $b \rightarrow \infty$ ,  $b/n \rightarrow 0$ ,  $B \rightarrow \infty$  as  $n \rightarrow \infty$ . First, compute the test statistic for each subsample

$$K_{I_j, b} = \sup_{\tau \in \mathcal{T}} \left| \left[ R(\tau)' \hat{V}(\tau) R(\tau) \right]^{-1/2} \sqrt{b} R(\tau)' \left( \hat{\beta}_{I_j, b}(\tau) - \hat{\beta}(\tau) \right) \right|, \quad (43)$$

where  $\hat{\beta}_{I_j, b}(\tau)$  is the QR estimate using subsample  $I_j$ . Then, define  $\hat{\kappa}(\alpha)$  as the  $\alpha$ -quantile of the subsampling sequence  $\{K_{I_1, b}, \dots, K_{I_B, b}\}$ . If recomputation of quantiles is not desirable, one can replace  $\sqrt{b}(\hat{\beta}_{I_j, b}(\tau) - \hat{\beta}(\tau))$  by its first order approximation, which is a re-centered one-step estimator:  $\hat{A}_{I_j, b}(\tau) = -\hat{J}(\tau)^{-1} \frac{1}{\sqrt{b}} \sum_{i \in I_j} (\tau - 1(Y_i \leq X_i' \hat{\beta}(\tau))) X_i$ .

**Corollary 4 (Consistent  $\hat{\kappa}(\alpha)$ )** *The estimator  $\hat{\kappa}(\alpha)$ , described above, is consistent for  $\kappa(\alpha)$ .*

As noted above, in practice we replace the continuum of quantile indices  $\mathcal{T}$  by a finite-grid  $\mathcal{T}_{K_n}$ , where the distance between adjacent grid points goes to zero as  $n \rightarrow \infty$ . Since the

inference processes considered are stochastically equicontinuous, this replacement does not affect the asymptotic theory.

To make previous inference methods operational, we also need uniformly consistent estimators for the components of the variance formulae:

$$\hat{\Sigma}(\tau, \tau') = \frac{1}{n} \sum_{i=1}^n (\tau - 1\{Y_i \leq X_i' \hat{\beta}(\tau)\})(\tau' - 1\{Y_i \leq X_i' \hat{\beta}(\tau')\}) \cdot X_i X_i', \quad (44)$$

$$\hat{\Sigma}_0(\tau, \tau') = [\min(\tau, \tau') - \tau \tau'] \cdot \frac{1}{n} \sum_{i=1}^n X_i X_i', \quad (45)$$

$$\hat{J}(\tau) = \frac{1}{2nh_n} \sum_{i=1}^n 1\{|Y_i - X_i' \hat{\beta}(\tau)| \leq h_n\} \cdot X_i X_i', \quad (46)$$

where  $h_n$  is such that  $h_n \rightarrow 0$  and  $h_n^2 n \rightarrow \infty$ .<sup>13</sup> The next result establishes the uniform consistency of these estimators.

**Corollary 5** *The estimators shown in equations (44)-(46) are uniformly consistent in  $(\tau, \tau')$  if  $E\|X\|^4 < \infty$ .*

Figure 5 illustrates uniform inference in our empirical example. The figure shows robust pointwise and uniform 95% confidence intervals for the schooling coefficient  $\hat{\beta}(\bullet)$  from quantile regressions of log-earnings on schooling, race and a quadratic function of experience, using data from the 1980, 1990 and 2000 censuses. The horizontal lines indicate the corresponding OLS estimates. The uniform bands were obtained by subsampling using 200 repetitions ( $B = 200$ ) with subsample size  $b = 5n^{2/5}$ , and a grid of quantiles  $\mathcal{T}_{K_n} = \{.1, .15, \dots, .9\}$ .<sup>14</sup>

The figure suggests the returns to schooling were low and essentially flat across quantiles in 1980, (except for  $\tau > .85$ , where they shift up), a finding similar to Buchinsky's (1994) using Current Population Surveys (CPS) for this period. On the other hand, the returns increased sharply and became more heterogeneous in 1990 and especially in 2000, a result we also confirmed in the CPS. Since the uniform confidence bands do not contain a horizontal line, we can reject the hypothesis of homogeneous returns to schooling for 1990 and 2000. Moreover, the uniform band for 1990 does not overlap with the 1980 band, suggesting a marked and statistically significant change in the relationship between schooling and the conditional wage distribution in this period.<sup>15</sup> A variety of other hypotheses regarding the returns to schooling can similarly be tested using Figure 5. Note also that the uniform bands are not much wider than

<sup>13</sup>Following Koenker (1994), we use Hall and Sheather's (1988) rule setting  $h_n = c \cdot n^{-1/3}$ .

<sup>14</sup>Chernozhukov (2002) discusses subsampling for QR inference in greater detail.

<sup>15</sup>Using Bonferroni bounds, our graphical test that looks for overlap in two 95% confidence bands has a sig-

the corresponding pointwise bands due to the high correlation between individual coefficients in the QR process.

## 4 Estimates of Changing Residual Inequality

One of the most significant and widely-studied developments in the American economy in the last three decades is the changing wage structure. The broad pattern has been one of increasing inequality, as measured by either the variance or the gap between upper and lower quantiles of the wage distribution. For example, Katz and Autor (1999) note that the 90-10 ratio (i.e. the ratio of the .9 and .1 quantiles) increased by 25 percent from 1979 to 1995. Wage inequality appears to have continued to increase since 1995, though the recent inequality trend is less clear-cut due in part to changes in the way US wage data are collected (Lemieux, 2003).

The increase in wage inequality is typically described as arising in two ways: increasing wage differentials associated with observed worker characteristics such as education and experience, and increased dispersion conditional on these characteristics. The first, known as “between-group inequality,” has increased as a consequence of changes in the distribution of characteristics, and especially changes in the economic returns to these characteristics. For example, increases in the economic return to schooling have been an important factor working to increase overall wage dispersion. The second, known as within-group or “residual inequality,” is – by definition – not directly linked to changes in the distribution of covariates or their returns, though increases in residual inequality are sometimes said to reflect increasing returns to “unobserved skills” (as in Juhn, Murphy, and Pierce, 1993).

An appealing feature of quantile regression as a tool for understanding wage inequality is that QR coefficients can easily be used to construct a measure of within-group or residual inequality. To see this, note that if we approximate  $Q_\tau(Y|X)$  by  $X'\beta(\tau)$ , with log wages as the dependent variable, then the within-group  $\tau$  to  $\tau'$  ratio is provided by  $X'[\beta(\tau) - \beta(\tau')]$ . This fact highlights a key difference between quantile regression and mean regression: a *ceteris paribus* increase in an OLS regression coefficient increases a variance-based measure of between-group inequality, without changing within-group inequality as measured by the residual variance. In contrast, a *ceteris paribus* increase in any non-central quantile,  $\tau$ , increases within-group inequality as measured by the spread from the  $\tau$  to  $1 - \tau$  quantiles.

---

nificance level of approximately 10% ( $1 - .95^2$ ). A test with exactly 5% size can be obtained by constructing confidence bands for the difference in estimated quantiles across years, again using the procedure outlined in section 3.2.

## 4.1 The QR Summary Picture

Our goal in this brief empirical section is to use linear QR to measure changing residual inequality in the 1980, 1990, and 2000 censuses. To mitigate the impact of changes in labor force participation, we continue to focus on a prime-age sample consisting of US-born white and black men aged 40-49.

Figure 6 provides a compact QR-generated summary of the evolution of residual inequality from 1980 through 2000. The figure plots the averaged (across covariates) conditional quantiles of earnings, as predicted from a QR model controlling for schooling, race, and a quadratic function of potential experience. The leftmost panel shows the unconditional quantiles, i.e., the marginal earnings distribution; the middle panel conditions on covariate means for each year; the third panel fixes the covariate means at their 1980 values.<sup>16</sup> Each panel shows quantiles for the three census years, plotted using a line-width determined by the uniform inference bands for fitted values derived from our QR estimates of the quantile process. To facilitate a comparison of inequality while holding location fixed, the line for each year is centered at median earnings for that year.

The largest shift in unconditional distributions occurred between 1980 and 1990, primarily in the lower half of the earnings distribution. This shift is statistically significant, as can be seen from the fact that the bands for these two years do not overlap. A comparison of Panels B and C with Panel A shows the residual distribution shifting more smoothly than the marginal distribution. This is because conditioning smooths out some of the heaping commonly found in survey-based earnings data. Panel B shows a clear increase in residual inequality from 1980 to 1990, with a continuing increase from 1990 to 2000. An interesting feature of the latter increase, however, is that it appears to have occurred only in the upper half of the wage distribution. Below the median, the conditional quantiles for 1990 and 2000 overlap. Panel C shows a similar pattern when the covariate distribution is held fixed. Autor, Katz, and Kearney (2004) report a similar asymmetry in their analysis of CPS data, with virtually all inequality growth in the 1990s in the upper half of the wage distribution.

## 4.2 Accuracy of the QR Picture

While Figure 6 provides a useful distillation of the QR results, we are especially interested in whether the linear QR model accurately captures key features of changing residual inequality in this period, both overall and for specific groups. The large census data sets allow us to compare

---

<sup>16</sup>Panel C uses a slightly different schooling recode to maximize comparability; see the appendix for details.

QR estimates with the corresponding non-parametric estimates of the CQF. Paralleling the analysis of 1980 census data in the previous section, we begin our analysis of changing residual inequality by assessing the quality of the QR fit to the CQF for 1990 and 2000 census data. Figures 7 and 8 show the QR fit to the CQF in both census data sets, for a model where the sole regressor is years of schooling. As for 1980, the fit is reasonably good at all quantiles, though somewhat worse at the .75 and .9 quantiles than lower down, especially for 2000. Again, the corresponding QR coefficient estimates are reported in Panel A of Table 1.

The figures also compare the QR regression line to the Chamberlain MD line, obtained from a histogram-weighted fit of the linear model to the CQF. Again, as for 1980, the MD and QR lines are almost indistinguishable, suggesting the importance weights are flat and/or the true CQF is not too far from linear. More evidence on the nature of the weighting function can be seen in Figures 9 and 10, which plot importance weights and histogram weights, and Figures 11 and 12, which plot the importance weights and density weights. These figures establish that the conditional density of  $Y$  given  $X$ , and hence the QR importance weights, are indeed fairly flat at all quantiles and in both years.

To assess the performance of QR as a tool for measuring residual inequality, Table 2 reports alternative inter-quantile spreads constructed from the CQF and QR. Panel A reports estimates for the whole sample, averaged using the sample distribution of the covariates. This panel shows an important overall increase in wage inequality, which cannot be totally explained by changes in the distribution of and returns to the covariates. The QR 90-10 spread tracks the CQF 90-10 spread remarkably well; the latter runs from 1.20 to 1.43, while QR implies a 90-10 spread ranging from 1.19 to 1.45 in the model that controls for schooling, race and experience. Results are equally good for the inter-quartile range and the two-half-spreads, and for the model that only controls for schooling. The asymmetry of residual inequality growth since 1990 can be seen by comparing the change in the 90-50 and 50-10 spreads.

The evolution of residual inequality for specific schooling groups provides a more stringent test of the QR approach. Panels B and C of Table 2 report results from a model that includes schooling with and without potential experience and race, evaluated for specific schooling groups. The 90-10 spread based on the CQF for high school graduates (12 years of schooling) moves from 1.09 in 1980 to 1.26 in 1990 to 1.29 in 2000, when race and experience are included. QR fitted values similarly show an increase from 1.17 in 1980 to 1.31 in 1990 and 1.32 in 2000. Thus, like the CQF for high school graduates, QR shows an increase in residual inequality of around .14 in the first decade, with essentially no change in the second. The results are similar without

controlling for race and experience. The evolution of the inter-quartile range appears to have been broadly similar to that of the 90-10 spread for the high school group.

While residual inequality grew little for high school graduates in the 1990s, college graduates (16 years of schooling) saw a substantial increase in wage dispersion. This echoes Autor, Katz, and Kearney’s (2004) comparison of college and high school graduates using the CPS. Again, QR captures the essential features of this pattern remarkably well. The 90-10 spread estimated from the CQF for college graduates increased from 1.26 to 1.44 in the 1980s and then to 1.55 in the 1990s. The corresponding QR estimates imply an increase from 1.19 to 1.38 in the 1980s, and then to 1.57 in the 1990’s. The QR estimates also capture about two-thirds of the growth in residual inequality over the entire period for the other spreads considered. The ability of QR to track these changes seems especially impressive given the changes (detailed in the data appendix) in the underlying schooling variable across censuses.

### 4.3 QR-based measures of inequality

As with variance-based measures of dispersion, we can use quantile spreads and their QR approximations to provide convenient summary measures of residual inequality. A natural measure already discussed is the inter-quantile range:  $IQR_{\tau,\tau'}[Y|X] \approx X'\beta(\tau) - X'\beta(\tau') = X'[\beta(\tau) - \beta(\tau')]$ , where  $\tau$  is some high index, for example 90%, and  $\tau'$  is some low index, for example 10%. A summary or typical measure of residual inequality is the median IQR,

$$RI_{\tau,\tau'} = Med\{IQR_{\tau,\tau'}[Y|X]\} \cong Med\{X'\beta(\tau) - X'\beta(\tau')\}. \quad (47)$$

On the other hand, a reasonable measure of between-group inequality can be given by the inter-quantile range of the conditional median:

$$BI_{\tau,\tau'} = IQR_{\tau,\tau'}\{Med[Y|X]\} \cong IQR_{\tau,\tau'}\{X'\beta(1/2)\}, \quad (48)$$

which measures the variation in the central location of the conditional distribution.

To grade the relative importance of within and between-group inequality, we can define the following “residual-to-total” ratio and its QR approximation:

$$RTR_{\tau,\tau'} = \frac{[Med\{IQR_{\tau,\tau'}[Y|X]\}]^2}{[Med\{IQR_{\tau,\tau'}[Y|X]\}]^2 + [IQR_{\tau,\tau'}\{Med[Y|X]\}]^2} \quad (49)$$

$$\cong \frac{[Med\{X'\beta(\tau) - X'\beta(\tau')\}]^2}{[Med\{X'\beta(\tau) - X'\beta(\tau')\}]^2 + [IQR_{\tau,\tau'}\{X'\beta(1/2)\}]^2}. \quad (50)$$

The RTR measure can be motivated by an analogy to traditional analysis of variance models, where the ratio of residual to total variance is “ $1 - R^2$ ”, i.e.

$$\frac{E\{V[Y|X]\}}{E\{V[Y|X]\} + V\{E[Y|X]\}} \quad (51)$$

RTR replaces standard deviation with inter-quantile range as a measure of dispersion and means with medians as a measure of location. In fact, in the classical normal location-shift model,  $Y = X'\beta + U$ , the two measures coincide (this is the reason why the squares are present in the definition of RTR), but they would be different in general.<sup>17</sup> RTR is nonnegative by construction and satisfies the natural restrictions

$$0 \leq \text{RTR}_{\tau, \tau'} \leq 1. \quad (52)$$

RTR necessarily equals 1 if  $Y$  is independent of  $X$  (no between-group inequality) and equals 0 when conditional dispersion is zero (no within-group inequality).

Table 3 compares ANOVA and quantile-based estimates of between-group inequality, within-group inequality, and the relative importance of within-group inequality for both a non-parametric and a linear model of log-earnings that includes schooling, race and potential experience as covariates. QR and CQ-based measures are generally closer for within-group inequality than for between-group inequality. Both QR and CQ-based measures suggest a sharp increase in within-group and between-group inequality, especially in the upper tail. For example,  $\text{RI}_{90,50}$  and  $\text{BI}_{90,50}$  grew much faster than  $\text{RI}_{50,10}$  and  $\text{BI}_{50,10}$ . On the other hand, there is no clear trend in the relative importance of within-group inequality. For example, the QR-based  $\text{RTR}_{90,10}$  go from 80% to 81% between 1980 and 1990, and then back to 78% in 2000. Some of these general trends are also captured by the standard ANOVA-based measures, but the latter does not capture the asymmetric changes in the upper and lower tails.

## 5 Summary and conclusions

We have shown how linear quantile regression provides a weighted least squares approximation to an unknown and potentially nonlinear conditional quantile function, much as OLS provides a least squares approximation to a nonlinear CEF. The QR approximation property leads to partial quantile plots and an omitted variables bias formula, analogous to standard specification tools for OLS.

---

<sup>17</sup>An alternative choice for the denominator is the marginal interquantile range  $Q_{\tau}(Y) - Q_{\tau'}(Y)$ . However, this leads to a relative measure that can exceed 1.



A natural question raised by the relationships explored here is the sensitivity of QR to changes in sample design. Unlike a postulated-as-true linear model, the nature of the QR approximation changes in stratified samples. Of course, an OLS regression line has this feature as well. Like the OLS approximation to a nonlinear CEF, the nature of the weights underlying the QR approximation to a nonlinear CQF change as the histogram of  $X$  changes (though not otherwise, since the importance weights are a function of  $X$ ). The role played by the QR weighting scheme seems like an empirical, application-specific question. In practice, it may be of interest to use stratification weights to improve the linear QR fit for subpopulations of special interest. This is a topic we plan to explore in future work.

While misspecification of the CQF functional form does not affect the usefulness of QR, it does have implications for inference. We have presented a misspecification-robust distribution theory for the QR process. This provides a foundation for uniform confidence intervals and a basis for global tests of hypotheses about distribution. The interpretation of such tests is more subtle, however, when the assumption of correct specification is dropped. The results of a global test may change as the nature of the QR approximation changes.

Finally, we used the tools here to describe the wage distribution in three censuses, proposing summary measures of between and within-group inequality. For the most part, linear QR captures the evolution of the conditional wage distribution remarkably well. Of particular interest is the finding that the growth of within-group inequality between 1990 and 2000 is largely due to an expansion of the upper half of the conditional wage distribution and the growing inequality in the wage distribution of college graduates. Traditional regression-based inequality measures miss these developments.

# A Appendix: Proofs

## A.1 Proof of Theorem 1.

We have that

$$\beta(\tau) = \arg \min_{\beta \in \mathbb{R}^d} E[\rho_\tau(Y - X'\beta)]. \quad (53)$$

Then, we can subtract  $E[\rho_\tau(Y - Q_\tau(Y|X))]$ , without affecting the optimization, because it does not depend on  $\beta$  and is finite by condition (ii):

$$\beta(\tau) = \arg \min_{\beta \in \mathbb{R}^d} \{E[\rho_\tau(Y - X'\beta)] - E[\rho_\tau(Y - Q_\tau(Y|X))]\}. \quad (54)$$

Write

$$\begin{aligned} & E[\rho_\tau(\epsilon_\tau - \Delta_\tau(X, \beta))] - E[\rho_\tau(\epsilon_\tau)] \\ &= E[(\tau - 1\{\epsilon_\tau < \Delta_\tau(X, \beta)\})(\epsilon_\tau - \Delta_\tau(X, \beta))] - E[(\tau - 1\{\epsilon_\tau < 0\})\epsilon_\tau] \\ &= \underbrace{E[(1\{\epsilon_\tau < \Delta_\tau(X, \beta)\} - \tau)\Delta_\tau(X, \beta)]}_I - \underbrace{E[(1\{\epsilon_\tau < \Delta_\tau(X, \beta)\} - 1\{\epsilon_\tau < 0\})\epsilon_\tau]}_{II}. \end{aligned} \quad (55)$$

Now, write

$$\begin{aligned} I &= E[(1\{\epsilon_\tau < \Delta_\tau(X, \beta)\} - \tau)\Delta_\tau(X, \beta)] \\ &\stackrel{(a)}{=} E[E[(1\{\epsilon_\tau < \Delta_\tau(X, \beta)\} - \tau)|X]\Delta_\tau(X, \beta)] \\ &= E[[F_{\epsilon_\tau}(\Delta_\tau(X, \beta)|X) - F_{\epsilon_\tau}(0|X)]\Delta_\tau(X, \beta)] \\ &\stackrel{(b)}{=} E\left[\left(\int_0^1 f_{\epsilon_\tau}(u\Delta_\tau(X, \beta)|X)\Delta_\tau(X, \beta)du\right)\Delta_\tau(X, \beta)\right] \\ &\stackrel{(c)}{=} E\left[\left(\int_0^1 f_{\epsilon_\tau}(u\Delta_\tau(X, \beta)|X)du\right)\Delta_\tau^2(X, \beta)\right] \end{aligned} \quad (56)$$

where (a) is by the law of iterated expectations, (b) is by condition (i) (a.s. existence of conditional density), and (c) is by linearity of the integral. Similarly,

$$\begin{aligned} II &= E[1\{\epsilon_\tau \in [0, \Delta_\tau(X, \beta)]\} \cdot |\epsilon_\tau|] + E[1\{\epsilon_\tau \in [\Delta_\tau(X, \beta), 0]\} \cdot |\epsilon_\tau|] \\ &= E[1\{u_\tau \in [0, 1]\} \cdot u_\tau \cdot |\Delta_\tau(X, \beta)|] \end{aligned} \quad (57)$$

where

$$\begin{aligned} u_\tau &\equiv \epsilon_\tau / \Delta_\tau(X, \beta) && \text{if } \Delta_\tau(X, \beta) \neq 0, \\ u_\tau &\equiv 1 && \text{if } \Delta_\tau(X, \beta) = 0. \end{aligned} \quad (58)$$

Next, note that for the case  $\Delta_\tau(X, \beta) \neq 0$

$$f_{u_\tau}(u|X) \cdot du = f_{\epsilon_\tau}(u\Delta_\tau(X, \beta)|X) \cdot |\Delta_\tau(X, \beta)| \cdot du, \text{ so} \quad (59)$$

$$E[1\{u_\tau \in [0, 1]\}u_\tau|X] \cdot |\Delta_\tau(X, \beta)| = \left[ \int_0^1 u f_{u_\tau}(u|X) du \right] \cdot |\Delta_\tau(X, \beta)| \quad (60)$$

$$= \left[ \int_0^1 u f_{\epsilon_\tau}(u\Delta_\tau(X, \beta)|X) du \right] \cdot \Delta_\tau^2(X, \beta). \quad (61)$$

For cases when  $\Delta_\tau(X, \beta) = 0$

$$E[1\{u_\tau \in [0, 1]\}u_\tau|X] \cdot |\Delta_\tau(X, \beta)| = 0. \quad (62)$$

Thus, it follows that

$$\begin{aligned} II &= E[1\{u_\tau \in [0, 1]\}u_\tau |\Delta_\tau(X, \beta)|] = E[E[1\{u_\tau \in [0, 1]\}u_\tau|X] |\Delta_\tau(X, \beta)|] \\ &= E\left[\left[\int_0^1 u f_{\epsilon_\tau}(u\Delta_\tau(X, \beta)|X) du\right] \cdot \Delta_\tau^2(X, \beta)\right]. \blacksquare \end{aligned} \quad (63)$$

## A.2 Proof of Theorem 2.

We have to prove that  $\beta(\tau)$  that solves

$$\beta(\tau) = \arg \min_{\beta \in \mathbb{R}^d} E[\rho_\tau(Y - X'\beta)], \quad (P1)$$

is equal to  $\beta^*(\tau)$  that solves

$$\beta^*(\tau) = \arg \min_{\beta \in \mathbb{R}^d} E[\tilde{w}_\tau(X) \cdot \Delta_\tau^2(X, \beta)]. \quad (P2)$$

The FOC for program (P2) is given by

$$2 E[\tilde{w}_\tau(X) \Delta_\tau(X, \beta^*(\tau)) X] = 0, \quad (64)$$

where

$$\tilde{w}_\tau(X) = \frac{1}{2} \int_0^1 f_{\epsilon_\tau}(u \cdot \Delta_\tau(X, \beta(\tau))|X) du. \quad (65)$$

The FOC for program (P1) is given by

$$I' = E[(1\{\epsilon_\tau < \Delta_\tau(X, \beta(\tau))\} - \tau) X] = 0, \quad (66)$$

which by calculations similar to those in (56) can be written as

$$\begin{aligned} I' &\stackrel{(a)}{=} E[E[(1\{\epsilon_\tau < \Delta_\tau(X, \beta(\tau))\} - \tau) |X] \cdot X] \\ &= E[(F_{\epsilon_\tau}(\Delta_\tau(X, \beta(\tau))|X) - F_{\epsilon_\tau}(0|X)) \cdot X] \\ &\stackrel{(b)}{=} E\left[\left(\int_0^1 f_{\epsilon_\tau}(u\Delta_\tau(X, \beta(\tau))|X) \Delta_\tau(X, \beta(\tau)) du\right) \cdot X\right] \\ &\stackrel{(c)}{=} E\left[\left(\int_0^1 f_{\epsilon_\tau}(u\Delta_\tau(X, \beta(\tau))|X) du\right) \Delta_\tau(X, \beta(\tau)) \cdot X\right] \end{aligned} \quad (67)$$

where (a) is by the law of iterated expectations, (b) is by a.s. existence of conditional density, and (c) by linearity of the integral. By the definition of  $\tilde{w}_\tau(X)$

$$I' = 2 E[\tilde{w}_\tau(X) \Delta_\tau(X, \beta(\tau)) X] = 0. \quad (68)$$

Finally, note that this is precisely the FOC for program (P2).

Both program (P1) and program (P2) are convex. (P1) has unique solution  $\beta(\tau)$  by assumption, which means it uniquely solves the FOC. Hence since the (P1) and (P2) have the same first order condition, it follows that  $\beta^*(\tau) = \beta(\tau)$  uniquely solves the FOC for both programs.  $\blacksquare$ .

### A.3 Proof of Theorem 3.

Taking claim 1 as given, claim 2 is immediate:

$$\gamma_1(\tau) = E [\tilde{w}_\tau^*(X)X_1X_1']^{-1}E [\tilde{w}_\tau^*(X)X_1(X_1'\beta_1(\tau) + R_\tau(X))] \quad (69)$$

$$= \beta_1(\tau) + E [\tilde{w}_\tau^*(X)X_1X_1']^{-1}E [\tilde{w}_\tau^*(X_1) X_1R_\tau(X)]. \quad (70)$$

It remains to prove claim 1. The first order condition of the quantile regression of  $Y$  on  $X_1$  in the population is given by

$$E [(1\{Y \leq X_1'\gamma_1(\tau)\} - \tau) X_1] = 0, \quad (71)$$

or for  $\epsilon_\tau = Y - Q_\tau(Y|X)$  and  $\Delta_\tau(X, \gamma_1(\tau)) = X_1'\gamma_1(\tau) - Q_\tau(Y|X)$ ,

$$E [(1\{\epsilon_\tau \leq \Delta_\tau(X, \gamma_1(\tau))\} - \tau) X_1] = 0. \quad (72)$$

This can be rewritten as

$$E [E[1\{\epsilon_\tau \leq \Delta_\tau(X, \gamma_1(\tau))\} - 1\{\epsilon_\tau \leq 0\}|X_1] X_1] = 0, \quad (73)$$

since  $P\{\epsilon_\tau \leq 0|X_1\} = E[P\{\epsilon_\tau \leq 0|X_1, X_2\} |X_1] = E[\tau|X_1] = \tau$ . Write

$$\begin{aligned} E[1\{\epsilon_\tau \leq \Delta_\tau(X, \gamma_1(\tau))\} - 1\{\epsilon_\tau \leq 0\}|X_1] &\stackrel{(a)}{=} E[E[1\{\epsilon_\tau \leq \Delta_\tau(X, \gamma_1(\tau))\} - 1\{\epsilon_\tau \leq 0\}|X]|X_1] \\ &= E[[F_{\epsilon_\tau}(\Delta_\tau(X, \gamma_1(\tau))|X) - F_{\epsilon_\tau}(0|X)]|X_1] \\ &\stackrel{(b)}{=} E \left[ \left( \int_0^1 f_{\epsilon_\tau}(u\Delta_\tau(X, \gamma_1(\tau))|X) \Delta_\tau(X, \gamma_1(\tau)) du \right) \middle| X_1 \right] \\ &\stackrel{(c)}{=} E \left[ \left( \int_0^1 f_{\epsilon_\tau}(u\Delta_\tau(X, \gamma_1(\tau))|X) du \right) \Delta_\tau(X, \gamma_1(\tau)) \middle| X_1 \right], \end{aligned} \quad (74)$$

where (a) is by the law of iterated expectations, (b) is by a.s. existence of conditional density, and (c) by linearity of the integral. Defining  $\tilde{w}_\tau^*(X) = \frac{1}{2} \int_0^1 f_{\epsilon_\tau}(u\Delta_\tau(X, \gamma_1(\tau))|X) du$ , we can rewrite the previous first order condition as

$$2 E [E [\tilde{w}_\tau^*(X) \Delta_\tau(X, \gamma_1(\tau))|X_1] \cdot X_1] = 0, \quad (75)$$

or, by the law of iterated expectations

$$2 E [\tilde{w}_\tau^*(X) \Delta_\tau(X, \gamma_1(\tau)) \cdot X_1] = 0. \quad (76)$$

Finally, note that this is precisely the first order condition for the program

$$\gamma_1(\tau) = \arg \min_{\gamma_1 \in \mathbb{R}^{d_1}} E[\tilde{w}_\tau^*(X) \cdot (X_1'\gamma_1 - Q_\tau(Y|X))^2]. \blacksquare \quad (77)$$

### A.4 Notation for Proofs of Theorems 4 and 5

We use the following empirical processes in the sequel, for  $W \equiv (Y, X)$

$$f \mapsto \mathbb{E}_n [f(W)] \equiv \frac{1}{n} \sum_{i=1}^n f(W_i), \quad f \mapsto \mathbb{G}_n [f(W)] \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n (f(W_i) - E [f(W_i)]). \quad (78)$$

If  $\widehat{f}$  is an estimated function,  $\mathbb{G}_n [\widehat{f}(W)]$  denotes  $\frac{1}{\sqrt{n}} \sum_{i=1}^n (f(W_i) - E[f(W_i)])_{f=\widehat{f}}$ . Other basic notation and stochastic convergence concepts, such as weak convergence in the space of bounded functions, stochastic equicontinuity, Donsker classes, and Vapnik-Červonenkis (VC) classes, are used and defined as in van der Vaart (1998).

## A.5 Proof of Theorem 4

Observe that, for each  $\tau$  in  $\mathcal{T}$ ,  $\widehat{\beta}(\tau)$  minimizes

$$Q_n(\tau, \beta) \equiv \mathbb{E}_n [\rho_\tau(Y - X'\beta) - \rho_\tau(Y - X'\widehat{\beta}(\tau))]. \quad (79)$$

Define

$$Q_\infty(\tau, \beta) \equiv E [\rho_\tau(Y - X'\beta) - \rho_\tau(Y - X'\beta(\tau))]. \quad (80)$$

By A.2 and A.3,  $Q_\infty(\tau, \beta)$  is uniquely minimized at  $\beta(\tau)$  for each  $\tau$  in  $\mathcal{T}$ .

Since by Knight's identity  $\rho_\tau(u - v) - \rho_\tau(u) = -(\tau - 1\{u < 0\})v + \int_0^v [1\{u \leq s\} - 1\{u \leq 0\}]ds$ , we have, by setting  $u = Y - X'\beta(\tau)$  and  $v = X'(\beta - \beta(\tau))$ , that

$$\begin{aligned} \rho_\tau(Y - X'\beta) - \rho_\tau(Y - X'\beta(\tau)) &= -(\tau - 1\{Y \leq X'\beta(\tau)\})X'(\beta - \beta(\tau)) \\ &\quad + \int_0^{X'(\beta - \beta(\tau))} [1\{Y \leq X'\beta(\tau) + s\} - 1\{Y \leq X'\beta(\tau)\}]ds. \end{aligned} \quad (81)$$

Thus, it follows that for any  $\beta \in \mathbb{R}^d$

$$|Q_\infty(\tau, \beta)| \leq 2 \cdot E|X'(\beta - \beta(\tau))| \leq 2 \cdot E\|X\| \cdot \|\beta - \beta(\tau)\| < \infty. \quad (82)$$

We can also show that for any compact set  $\mathcal{B}$

$$Q_n(\tau, \beta) = Q_\infty(\tau, \beta) + o_{p^*}(1), \text{ uniformly in } (\tau, \beta) \in \mathcal{T} \times \mathcal{B}. \quad (83)$$

This statement is true pointwise by the Khinchin LLN. The uniform convergence follows because

$$|Q_n(\tau', \beta') - Q_n(\tau'', \beta'')| \leq C_1 \cdot |\tau' - \tau''| + C_2 \cdot \|\beta' - \beta''\|, \quad (84)$$

where

$$C_1 = 2 \cdot E\|X\| \cdot \sup_{\beta \in \mathcal{B}} \|\beta\| < \infty \text{ and } C_2 = 2 \cdot E\|X\| < \infty. \quad (85)$$

Hence the empirical process  $(\tau, \beta) \mapsto Q_n(\tau, \beta)$  is stochastically equicontinuous, which implies the uniform convergence.

Consider a collection of closed balls  $B_M(\beta(\tau))$  of radius  $M$  and center  $\beta(\tau)$ , and let  $\beta_M(\tau) = \beta(\tau) + \delta_M(\tau) \cdot v(\tau)$ , where  $v(\tau) = (v_1(\tau), \dots, v_d(\tau))'$  is a direction vector with unity norm  $\|v(\tau)\| = 1$  and  $\delta_M(\tau)$  is a positive scalar such that  $\delta_M(\tau) \geq M$ . Then uniformly in  $\tau \in \mathcal{T}$ ,

$$\begin{aligned} \frac{M}{\delta_M(\tau)} (Q_n(\tau, \beta_M(\tau)) - Q_n(\tau, \beta(\tau))) &\stackrel{(a)}{\geq} Q_n(\tau, \beta_M^*(\tau)) - Q_n(\tau, \beta(\tau)) \\ &\stackrel{(b)}{\geq} Q_\infty(\tau, \beta_M^*(\tau)) - Q_\infty(\tau, \beta(\tau)) + o_{p^*}(1) \\ &\stackrel{(c)}{>} \epsilon_M + o_{p^*}(1), \end{aligned} \quad (86)$$

for some  $\epsilon_M > 0$ , where (a) follows by convexity in  $\beta$ , for  $\beta_M^*(\tau)$  on the line connecting  $\beta_M(\tau)$  and  $\beta(\tau)$ , (b) follows by the uniform convergence established in (83), (c) follows by the assumption that  $\beta(\tau)$  is the unique minimizer of  $Q_\infty(\beta, \tau)$  uniformly in  $\tau \in \mathcal{T}$ . Hence for any  $M > 0$ , the minimizer  $\hat{\beta}(\tau)$  must be within  $M$  from  $\beta(\tau)$  uniformly for all  $\tau \in \mathcal{T}$ , with probability approaching to one. That is, we have that for any  $M > 0$ ,  $\|\hat{\beta}(\tau) - \beta(\tau)\| \leq M$  uniformly for all  $\tau \in \mathcal{T}$  with probability approaching to one. ■

## A.6 Proof of Theorem 5

First, by the computational properties of  $\hat{\beta}(\tau)$ , for all  $\tau \in \mathcal{T}$ , cf. Theorem 3.3 in Koenker and Bassett (1978):

$$\left\| \mathbb{E}_n \left[ \varphi_\tau(Y - X' \hat{\beta}(\tau)) X \right] \right\| \leq \text{const} \cdot \left( \frac{\sup_{i \leq n} \|X_i\|}{n} \right), \quad (87)$$

where  $\varphi_\tau(u) = \tau - 1\{u < 0\}$ . Note that  $E\|X_i\|^{2+\epsilon} < \infty$  implies  $\sup_{i \leq n} \|X_i\| = o_{p^*}(n^{1/2})$ , since

$$P \left( \sup_{i \leq n} \|X_i\| > n^{1/2} \right) \leq nP(\|X_i\| > n^{1/2}) \leq nE\|X_i\|^{2+\epsilon}/n^{\frac{2+\epsilon}{2}} = o(1). \quad (88)$$

Hence uniformly in  $\tau \in \mathcal{T}$ ,

$$\mathbb{E}_n \left[ \varphi_\tau(Y - X' \hat{\beta}_n(\tau)) X \right] = o_{p^*} \left( n^{-1/2} \right). \quad (89)$$

Second,  $(\tau, \beta) \mapsto \mathbb{G}_n[\varphi_\tau(Y - X'\beta)X]$  is stochastically equicontinuous over  $\mathcal{B} \times \mathcal{T}$ , where  $\mathcal{B}$  is any compact set, with respect to the  $L_2(P)$  pseudometric

$$\rho((\tau', \beta'), (\tau'', \beta'')) \equiv \max_{j \in 1, \dots, d} \sqrt{E \left[ (\varphi_{\tau'}(Y - X'\beta') X_j - \varphi_{\tau''}(Y - X'\beta'') X_j)^2 \right]}, \quad (90)$$

for  $j \in 1, \dots, d$  indexing the components of the vector  $X$ . This is because the functional class  $\mathcal{F} = \{1\{Y \leq X'\beta\}, \beta \in \mathcal{B}\}$  is a VC subgraph class and hence also Donsker class, with envelope 2. Hence the functional class  $\mathcal{T} - \mathcal{F}$  is also Donsker with envelope equal 2, by Theorem 2.10.6 in Van der Vaart and Wellner (1996). The product of  $\mathcal{T} - \mathcal{F}$  with  $X$  also forms a Donsker class with a square integrable envelope  $2 \cdot \max_{j \in 1, \dots, d} |X|_j$ , by Theorem 2.10.6 in Van der Vaart and Wellner (1996). The stochastic equicontinuity then is a part of being Donsker.

The uniform consistency  $\sup_{\tau \in \mathcal{T}} \|\hat{\beta}(\tau) - \beta(\tau)\| = o_{p^*}(1)$  implies

$$\sup_{\tau \in \mathcal{T}} \rho \left( (\tau, b(\tau)), (\tau, \beta(\tau)) \right) \Big|_{b(\tau) = \hat{\beta}(\tau)} = o_{p^*}(1), \quad (91)$$

and therefore by stochastic equicontinuity of  $(\tau, \beta) \mapsto \mathbb{G}_n[\varphi_\tau(Y - X'\beta)X]$  we have that

$$\mathbb{G}_n \left[ \varphi_\tau(Y - X' \hat{\beta}(\tau)) X \right] = \mathbb{G}_n [\varphi_\tau(Y - X' \beta(\tau)) X] + o_{p^*}(1), \text{ uniformly in } \tau. \quad (92)$$

In order to show (91) note that for  $\bar{f}$  denoting the upper bound on  $f_Y(y|X=x)$ , application of the

Hölder's inequality, a Taylor expansion and Cauchy-Schwarz inequality, give the series of inequalities:

$$\begin{aligned}
& \sup_{\tau \in \mathcal{T}} \rho\left((\tau, b(\tau)), (\tau, \beta(\tau))\right) \\
&= \sup_{\tau \in \mathcal{T}} \max_{j \in 1, \dots, d} \sqrt{E \left[ (\varphi_\tau(Y - X'b(\tau)) X_j - \varphi_\tau(Y - X'\beta(\tau)) X_j)^2 \right]} \\
&\leq \sup_{\tau \in \mathcal{T}} \max_{j \in 1, \dots, d} \left( \left( E \left[ |\varphi_\tau(Y - X'b(\tau)) - \varphi_\tau(Y - X'\beta(\tau))|^{\frac{2(2+\epsilon)}{\epsilon}} \right] \right)^\epsilon \cdot \left( E \left[ |X_j|^{2+\epsilon} \right] \right)^2 \right)^{\frac{1}{2(2+\epsilon)}} \\
&\leq \sup_{\tau \in \mathcal{T}} \max_{j \in 1, \dots, d} \left( E \left[ 1\{Y \leq X'b(\tau)\} - 1\{Y \leq X'\beta(\tau)\} \right]^{\frac{\epsilon}{2(2+\epsilon)}} \cdot \left( E |X_j|^{2+\epsilon} \right)^{\frac{1}{2+\epsilon}} \right) \\
&\leq \sup_{\tau \in \mathcal{T}} \left( E \left[ \bar{f} \cdot X'(b(\tau) - \beta(\tau)) \right]^{\frac{\epsilon}{2(2+\epsilon)}} \cdot (E \|X\|^{2+\epsilon})^{\frac{1}{2+\epsilon}} \right) \\
&\leq \text{const} \cdot \sup_{\tau \in \mathcal{T}} \left( \bar{f} \cdot (E \|X\|^2)^{1/2} \cdot \|b(\tau) - \beta(\tau)\| \right)^{\frac{\epsilon}{2(2+\epsilon)}},
\end{aligned} \tag{93}$$

where the second inequality follows by binomiality of  $|\varphi_\tau(Y - X'b(\tau)) - \varphi_\tau(Y - X'\beta(\tau))|$ . Then, evaluating at  $b(\tau) = \hat{\beta}(\tau)$

$$\sup_{\tau \in \mathcal{T}} \rho\left((\tau, b(\tau)), (\tau, \beta(\tau))\right) \Big|_{b(\tau)=\hat{\beta}(\tau)} \leq \text{const} \cdot \sup_{\tau \in \mathcal{T}} \|\hat{\beta}(\tau) - \beta(\tau)\|^{\frac{\epsilon}{2(2+\epsilon)}} = o_{p^*}(1), \tag{94}$$

by uniform convergence and  $\epsilon > 0$ .

Third, by a Taylor expansion, uniformly in  $\tau \in \mathcal{T}$

$$E [\varphi_\tau(Y - X'\beta)X] \Big|_{\beta=\hat{\beta}(\tau)} = E [f_Y(X'b(\tau)|X)XX'] \Big|_{b(\tau)=\beta^*(\tau)} (\hat{\beta}(\tau) - \beta(\tau)), \tag{95}$$

where  $\beta^*(\tau)$  is on the line connecting  $\hat{\beta}(\tau)$  and  $\beta(\tau)$  for each  $\tau$ .  $\hat{\beta}(\tau)$  is uniformly consistent by Theorem 4, hence  $\beta^*(\tau)$  is also uniformly consistent. Thus by **A5**, i.e. the uniform continuity and boundedness of the mapping  $y \mapsto f_Y(y|x)$ , uniformly in  $x$  over the support of  $X$ , it follows that

$$E [f_Y(X'b(\bullet)|X)XX'] \Big|_{b(\bullet)=\beta^*(\bullet)} = \underbrace{E [f_Y(X'\beta(\bullet)|X)XX']}_{J(\bullet)} + o(1) \text{ in } \ell^\infty(\mathcal{T}). \tag{96}$$

Indeed, by **A5** for any compact  $K$ ,  $E [f_Y(X'b(\bullet)|X)XX'1\{X \in K\}] \Big|_{b(\bullet)=\beta^*(\bullet)} = E [f_Y(X'\beta(\bullet)|X)XX'1\{X \in K\}] + o(1)$ . Then, for  $K^c = \mathbb{R}^d \setminus K$ ,  $E [f_Y(X'b(\bullet)|X)XX'1\{X \in K^c\}] \Big|_{b(\bullet)=\beta^*(\bullet)}$  and  $E [f_Y(X'\beta(\bullet)|X)XX'1\{X \in K^c\}]$  can be made arbitrarily small in large samples. This follows by setting the set  $K$  sufficiently large and using  $E \|XX'\| < \infty$  and  $f_Y(X'\beta(\bullet)|X) < \bar{f}$  a.s.

Fourth, since

$$\text{the left hand side (lhs) of (89) = lhs of (95) + } n^{-1/2} \text{ lhs of (92),} \tag{97}$$

we have by using (96)

$$\begin{aligned}
o_{p^*} \left( n^{-1/2} \right) &= J(\bullet)(\hat{\beta}(\bullet) - \beta(\bullet)) + o_{p^*} \left( \sup_{\tau \in \mathcal{T}} \|\hat{\beta}(\tau) - \beta(\tau)\| \right) \\
&\quad + n^{-1/2} \mathbb{G}_n[\varphi_\bullet(Y - X'\beta(\bullet))X] + o_{p^*}(n^{-1/2}) \text{ in } \ell^\infty(\mathcal{T}).
\end{aligned} \tag{98}$$

Since  $\text{mineig}(J(\tau)) > \lambda > 0$ , uniformly in  $\tau \in \mathcal{T}$

$$\begin{aligned} & \sup_{\tau \in \mathcal{T}} \left\| n^{-1/2} \mathbb{G}_n [\varphi_\tau(Y - X'\beta(\tau))X] + o_{p^*}(n^{-1/2}) \right\| \\ &= \sup_{\tau \in \mathcal{T}} \left\| J(\tau)(\hat{\beta}(\tau) - \beta(\tau)) + o_{p^*} \left( \sup_{\tau \in \mathcal{T}} \|\hat{\beta}(\tau) - \beta(\tau)\| \right) \right\| \\ &\geq (\lambda + o_{p^*}(1)) \cdot \sup_{\tau \in \mathcal{T}} \|\hat{\beta}(\tau) - \beta(\tau)\|. \end{aligned} \quad (99)$$

Fifth, by the stated assumptions, the mapping  $\tau \mapsto \beta(\tau)$  is continuous. In fact, it is continuously differentiable, since by the implicit function theorem, for  $\beta(\tau)$  defined as solution to  $E[(\tau - 1\{Y \leq X'\beta\})X] = 0$ , we have that  $d\beta(\tau)/d\tau = J(\tau)^{-1}E[X]$ . Hence  $\tau \mapsto \mathbb{G}_n[\varphi_\tau(Y - X'\beta(\tau))X]$  is stochastically equicontinuous over  $\mathcal{T}$  by continuity of the mapping  $\tau \mapsto \beta(\tau)$  for the pseudo-metric given by  $\rho(\tau', \tau'') \equiv \rho((\tau', \beta(\tau')), (\tau'', \beta(\tau'')))$ . Then, stochastic equicontinuity of  $\tau \mapsto \mathbb{G}_n[\varphi_\tau(Y - X'\beta(\tau))X]$  and ordinary CLT imply that

$$\mathbb{G}_n[\varphi_\bullet(Y - X'\beta(\bullet))X] \Rightarrow z(\bullet) \text{ in } \ell^\infty(\mathcal{T}), \quad (100)$$

where  $z(\bullet)$  is a Gaussian process with covariance function  $\Sigma(\bullet, \bullet)$  specified in the statement of Theorem 5. Therefore, the lhs of (99) is  $O_p(n^{-1/2})$ , implying

$$\sup_{\tau \in \mathcal{T}} \|\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau))\| = O_{p^*}(1). \quad (101)$$

Finally, by (99)-(101)

$$\begin{aligned} \sqrt{n}(\hat{\beta}(\bullet) - \beta(\bullet)) &= -J^{-1}(\bullet)\mathbb{G}_n[\varphi_\bullet(Y - X'\beta(\bullet))] + o_{p^*}(1) \text{ in } \ell^\infty(\mathcal{T}) \\ &\Rightarrow J^{-1}(\bullet) \cdot z(\bullet) \text{ in } \ell^\infty(\mathcal{T}). \quad \blacksquare \end{aligned} \quad (102)$$

## A.7 Proof of Corollaries

**Proof of Corollary 1.** The result is immediate from the definition of weak convergence in  $\ell^\infty(\mathcal{T})$ .  $\blacksquare$

**Proof of Corollary 2.** The result follows by the continuous mapping theorem in  $\ell^\infty(\mathcal{T})$ .  $\blacksquare$

**Proof of Corollary 3.** The result is immediate from Corollary 2.  $\blacksquare$

**Proof of Corollary 4.** The result is immediate from Politis, Romano and Wolf (1999), Theorem 2.2.1 and Corollary 2.4.1, for the case when the rescaling matrices are known. For the case when the matrices are consistently estimated the proof follows by an argument similar to the proof of Theorem 2.5.1 in Politis, Romano and Wolf (1999). Finally, we also need that  $\mathcal{K}$  has an absolutely continuous distribution. This result follows from Theorem 11.1 in Davydov, Lifshits, and Smorodina (1998).  $\blacksquare$

**Proof of Corollary 5.** Note that this corollary is not covered by the results in Powell (1986) or Buchinsky and Hahn (1998) for consistency of  $\hat{J}(\tau)$ , because their proofs apply only pointwise in  $\tau$ , whereas we require a uniform result.

First, recall that

$$\hat{J}(\tau) = \frac{1}{2h_n} \mathbb{E}_n \left[ 1\{|Y_i - X_i'\hat{\beta}(\tau)| \leq h_n\} \cdot X_i X_i' \right]. \quad (103)$$

We will show that

$$\hat{J}(\tau) - J(\tau) = o_{p^*}(1) \text{ uniformly in } \tau \in \mathcal{T}. \quad (104)$$



Note that  $2h_n\hat{J}(\tau) = \mathbb{E}_n [f_i(\hat{\beta}(\tau), h_n)]$ , where  $f_i(\beta, h) = 1\{|Y_i - X_i'\beta| \leq h\} \cdot X_i X_i'$ . Next, for any compact set  $B$  and positive constant  $H$ , the functional class  $\{f_i(\beta, h), \beta \in B, h \in (0, H]\}$  is a Donsker class with square integrable envelope by Theorem 2.10.6 in Van der Vaart and Wellner (1996), since this is a product of a VC class  $\{1\{|Y_i - X_i'\beta| \leq h\}, \beta \in B, h \in (0, H]\}$  and a square integrable random matrix  $X_i X_i'$  (recall  $E\|X_i\|^4 < \infty$  by assumption). Therefore,  $(\beta, h) \mapsto \mathbb{G}_n [f_i(\beta, h)]$  converges to a Gaussian process in  $\ell^\infty(B \times (0, H))$ , which implies that

$$\sup_{\beta \in B, 0 < h \leq H} \left\| \mathbb{E}_n [f_i(\beta, h)] - E [f_i(\beta, h)] \right\| = O_{p^*}(n^{-1/2}). \quad (105)$$

Letting  $B$  be any compact set that covers  $\cup_{\tau \in \mathcal{T}} \beta(\tau)$ , this implies

$$\sup_{\tau \in \mathcal{T}} \left\| \mathbb{E}_n [f_i(\hat{\beta}(\tau), h_n)] - E [f_i(\beta, h)] \Big|_{\beta=\hat{\beta}(\tau), h=h_n} \right\| = O_{p^*}(n^{-1/2}). \quad (106)$$

Hence (104) follows by using that  $2h_n\hat{J}(\tau) = \mathbb{E}_n [f_i(\hat{\beta}(\tau), h_n)]$  and noting that  $1/2h_n \cdot E [f_i(\beta, h)] \Big|_{\hat{\beta}(\tau), h=h_n} = J(\tau) + o_p(1)$  by an argument similar to that used in (96) and the assumption  $h_n^2 n \rightarrow \infty$ .

Second, we can write

$$\hat{\Sigma}(\tau, \tau') = \mathbb{E}_n [g_i(\hat{\beta}(\tau), \hat{\beta}(\tau'), \tau, \tau') X_i X_i'], \quad (107)$$

where  $g_i(\beta', \beta'', \tau', \tau'') = (\tau - 1\{Y_i \leq X_i'\beta'\})(\tau' - 1\{Y_i \leq X_i'\beta''\}) \cdot X_i X_i'$ . We will show that

$$\hat{\Sigma}(\tau, \tau') - \Sigma(\tau, \tau') = o_{p^*}(1) \text{ uniformly in } (\tau, \tau') \in \mathcal{T} \times \mathcal{T}. \quad (108)$$

Note that  $\{g_i(\beta', \beta'', \tau', \tau''), (\beta', \beta'', \tau', \tau'') \in B \times B \times \mathcal{T} \times \mathcal{T}\}$  is Donsker and hence a Glivenko-Cantelli class, for any bounded set  $B$ . Indeed,  $\mathcal{F}_\beta = \{1\{Y_i \leq X_i'\beta\}, \beta \in B\}$  is a VC class, and hence is Donsker. Then,  $\mathcal{T} - \mathcal{F}_\beta$  is also a bounded Donsker class with envelope 2, by Theorem 2.10.6 in Van der Vaart and Wellner (1996). Next, the product of two bounded classes  $(\mathcal{T} - \mathcal{F}_\beta) \times (\mathcal{T} - \mathcal{F}_\beta)$  is a bounded Donsker class with envelope 4, by Theorem 2.10.6 in Van der Vaart and Wellner (1996). Last, the product of a bounded Donsker class with a square integrable random matrix  $X X_i'$  gives a Donsker class, by Theorem 2.10.6 in Van der Vaart and Wellner (1996).

This implies that uniformly in  $(\beta', \beta'', \tau', \tau'') \in (B \times B \times \mathcal{T} \times \mathcal{T})$

$$\mathbb{E}_n [g_i(\beta', \beta'', \tau', \tau'') X_i X_i'] - E [g_i(\beta', \beta'', \tau', \tau'') X_i X_i'] = o_{p^*}(1). \quad (109)$$

By inspection,  $E [g_i(\beta', \beta'', \tau', \tau'') X_i X_i']$  is continuous in  $(\beta', \beta'', \tau', \tau'')$  over  $(B \times B \times \mathcal{T} \times \mathcal{T})$ . Letting  $B$  cover  $\cup_{\tau} \beta(\tau)$ , continuity and (109) imply (108).

A similar argument applies to  $\hat{\Sigma}_0(\tau, \tau')$ . ■

## B Appendix: Estimating the QR Weighting Function

We calculate the importance weights using equation (7). The integral was estimated with a grid of 101 points between the non-parametric estimates of the CQF ( $\hat{Q}_\tau(Y|X)$ ) and the QR approximation ( $X'\hat{\beta}(\tau)$ ),

for each cell of the covariates  $X$ . This gives rise to the following discrete approximation formula for the importance weights

$$\widehat{w}_\tau(x, \widehat{\beta}(\tau)) = \frac{1}{101} \sum_{u=1}^{101} \left(1 - \frac{u-1}{100}\right) \cdot \widehat{f}_Y \left( \frac{u-1}{100} \cdot x' \widehat{\beta}(\tau) + \left(1 - \frac{u-1}{100}\right) \cdot \widehat{Q}_\tau(Y|X=x)|X=x \right). \quad (110)$$

We used kernel density estimates of  $f_Y(y|X=x)$  with a Gaussian kernel and bandwidth ( $h$ ) determined by

$$m = \min \left[ \sqrt{\text{Var}[Y - \widehat{Q}_\tau(Y|X=x)|X=x]}, \frac{IQR_{0.25,0.75}[Y - \widehat{Q}_\tau(Y|X=x)|X=x]}{1.349} \right], \quad (111)$$

$$h = \frac{0.9 \cdot m}{n^{1/5}}. \quad (112)$$

This bandwidth choice is optimal in the sense that it minimizes mean integrated square error with Gaussian data and a Gaussian kernel (Silverman, 1986). The density weights were calculated similarly. Sampling weights were used in the estimation of conditional densities for the 2000 census sample.

To calculate weights for partial quantile correlation,  $\widetilde{w}_\tau(X)$ , we also use a discrete approximation of the average density of the response variable representation. In particular, we have

$$\widehat{w}_\tau(x) = \frac{1}{101} \sum_{u=1}^{101} \frac{1}{2} \cdot \widehat{f}_Y \left( \frac{u-1}{100} \cdot x' \widehat{\beta}(\tau) + \left(1 - \frac{u-1}{100}\right) \cdot \widehat{Q}_\tau(Y|X=x)|X=x \right), \quad (113)$$

where the conditional densities are estimates using the same kernel method as for the importance weights.

## C Appendix: Sampling Weights

In order to take into account the weighted structure of the census 2000 sample, the estimators for the components of the variance formulae in Table 1 were modified as follows

$$\widehat{\Sigma}(\tau, \tau') = \frac{1}{n} \sum_{i=1}^n w_i^2 \cdot (\tau - 1(Y_i \leq X_i' \widehat{\beta}(\tau))) (\tau' - 1(Y_i \leq X_i' \widehat{\beta}(\tau'))) \cdot X_i X_i', \quad (114)$$

$$\widehat{\Sigma}_0(\tau, \tau) = [\min(\tau, \tau') - \tau \tau'] \cdot \frac{1}{n} \sum_{i=1}^n w_i^2 \cdot X_i X_i', \quad (115)$$

$$\widehat{J}(\tau) = \frac{1}{2nh_n} \sum_{i=1}^n w_i \cdot 1(|Y_i - X_i' \widehat{\beta}(\tau)| \leq h_n) \cdot X_i X_i'. \quad (116)$$

where  $w_i$  are the sampling weights (normalized to add to  $n$ ). Other calculations involving the 2000 sample use sampling weights in the standard way.

## D Appendix: Data

The data were drawn from the 1% self-weighting 1980 and 1990 samples, and the 1% weighted 2000 sample, all from the IPUMS website (Ruggles *et al.*, 2003). The sample for most of the calculations

consists of US-born black and white men with age 40-49 with at least 5 years of education, with positive annual earnings and hours worked in the year preceding the census, and with nonzero sampling weight. Individuals with imputed values for age, education, earnings or weeks worked were also excluded from the sample. After this selection process, the final sample sizes were 65,023, 86,785 and 97,397 for 1980, 1990 and 2000.

The log-earnings variable is the average log weekly wage and was calculated as the log of the reported annual income from work divided by weeks worked in the previous year. Annual income is expressed in 1989 dollars using the Personal Consumption Expenditures Price Index.

The education variable for 1980 corresponds to the highest grade of school completed, coded as follows:

Years of schooling	Highest grade of school completed
5	5th grade of Elementary School
6	6th grade of Elementary School
7	7th grade of Elementary School
8	8th grade of Elementary School
9	9th grade of High School
10	10th grade of High School
11	11th grade of High School
12	12th grade of High School
13	1st year of College
14	2nd year of College
15	3rd year of College
16	4th year of College
17	5th year of College
18	6th year of College
19	7th year of College
20	8th or more year of College

For the purposes of Figure 5 and most of the empirical work, years of schooling for 1990 and 2000 censuses

were imputed from categorical schooling variables as follows:

Years of schooling	Educational attainment
8	5th, 6th, 7th, or 8th grade
9	9th grade
10	10th grade
11	11th or 12th grade, no diploma
12	High school graduate, diploma or GED
13	Some college, but no degree
14	Completed associate degree in college, occupational program
15	Completed associate degree in college, academic program
16	Completed bachelor's degree, not attending school
17	Completed bachelor's degree, but now enrolled
18	Completed master's degree
19	Completed professional degree
20	Completed doctorate

For the purposes of Panel C in Figure 6, we modify this slightly, coding 5th-8th grade as 8 and 2-3 years in college as 14 in 1980, and coding the categories associate college degree, occupational program, and associate degree, academic program, as 14 in 1990. These changes generate schooling variables with the same range and points of support in all 3 years.

## References

- [1] Abadie, A. (1997): "Changes in Spanish Labor Income Structure during the 1980's: A Quantile Regression Approach," *Investigaciones Economicas XXI*(2), pp. 253-272.
- [2] Andrews, D. and D. Pollard (1994): "An Introduction to Functional Central Limit Theorems for Dependent Stochastic Processes," *International Statistical Review* 62(1), pp. 119-132.
- [3] Angrist, J. and A. Krueger (1999): "Empirical Strategies in Labor Economics," in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, Volume 3. Amsterdam. Elsevier Science.
- [4] Autor, D., L.F. Katz and M.S. Kearney (2004): "Inequality in the 1990s: Revising the Revisionists," MIT Department of Economics, mimeo, March 2004.
- [5] Bai, J. (1998): "Testing Parametric Conditional Distributions of Dynamic Models," preprint.
- [6] Buchinsky, M. (1994): "Changes in the US Wage Structure 1963-1987: Application of Quantile Regression," *Econometrica* 62, pp. 405-458.
- [7] Buchinsky, M. and J. Hahn (1998): "An Alternative Estimator for the Censored Quantile Regression Model," *Econometrica* 66, no. 3, pp. 653-671.

- [8] Chamberlain, G. (1984): "Panel Data," in Z. Griliches and M. Intriligator (eds.), *Handbook of Econometrics*, Volume 2. North-Holland. Amsterdam.
- [9] Chamberlain, G. (1994): "Quantile Regression, Censoring, and the Structure of Wages," in C. A. Sims (ed.), *Advances in Econometrics, Sixth World Congress*, Volume 1. Cambridge University Press. Cambridge.
- [10] Chernozhukov, V. (2002): "Inference on the Quantile Regression Process, An Alternative," Unpublished Working Paper 02-12 (February), MIT ([www.ssrn.com](http://www.ssrn.com)).
- [11] Doksum, K. (1974): "Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two-Sample Case," *Annals of Statistics* 2, pp. 267-277.
- [12] Davydov, Yu. A., M. A. Lifshits and N. V. Smorodina (1998): *Local properties of distributions of stochastic functionals*. Translated from the 1995 Russian original by V. E. Nazaikinskii and M. A. Shishkova. Translations of Mathematical Monographs, 173. American Mathematical Society, Providence, RI.
- [13] Giacomini, R. and Komunjer, I. (2003): "Evaluation and Combination of Conditional Quantile Forecasts," Working Paper 571, Boston College and California Institute of Technology, 06/2003.
- [14] Goldberger, A. S. (1991): *A Course in Econometrics*. Harvard University Press. Cambridge, MA.
- [15] Gosling, A., S. Machin and C. Meghir (2000): "The Changing Distribution of Male Wages in the U.K.," *Review of Economic Studies* 67, pp. 635-666.
- [16] Gutenbrunner, C. and J. Jurečková (1992): "Regression Quantile and Regression Rank Score Process in the Linear Model and Derived Statistics," *Annals of Statistics* 20, pp. 305-330.
- [17] Gutenbrunner, C., J. Jurečková, R. Koenker, and S. Portnoy (1993): "Test of Linear Hypotheses Based on Regression Rank Scores," *Journal of Nonparametric Statistics* 2, pp. 307-331.
- [18] Hahn, J. (1997): "Bayesian Bootstrap of the Quantile Regression Estimator: A Large Sample Study," *International Economic Review* 38(4), pp. 795-808.
- [19] Hall, P., and S. Sheather (1988): "On the Distribution of a Studentized Quantile," *Journal of the Royal Statistics Society B* 50, pp. 381-391.
- [20] Hansen, L.-P. , J. Heaton and A. Yaron (1996): "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics* 14(3), pp. 262-280.
- [21] Juhn, C., K. Murphy and B. Pierce (1993): "Wage Inequality and the Rise in Return to Skill," *Journal of Political Economy* 101, pp. 410-422.
- [22] Katz, L. and D. Autor (1999): "Changes in the Wage Structure and Earnings Inequality," in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, Volume 3A. Elsevier Science. Amsterdam.

- [23] Katz, L. and K. Murphy (1992): “Changes in the Relative Wages, 1963-1987: Supply and Demand Factors,” *Quarterly Journal of Economics* 107, pp. 35-78.
- [24] Kim T.H., and H. White (2002): “Estimation, Inference, and Specification Testing for Possibly Misspecified Quantile Regressions,” in *Advances in Econometrics*, forthcoming.
- [25] Knight, K. (2002): “Comparing Conditional Quantile Estimators: First and Second Order Considerations,” Mimeo. University of Toronto.
- [26] Koenker, R. (1994): “Confidence Intervals for Regression Quantiles,” in M.P. and M. Hušková (eds.), *Asymptotic Statistics: Proceeding of the 5th Prague Symposium of Asymptotic Statistics*. Physica-Verlag, Heidelberg.
- [27] Koenker, R. and G. Bassett (1978): “Regression Quantiles,” *Econometrica* 46, pp. 33-50.
- [28] Koenker, R. and K. Hallock (2001): “Quantile Regression,” *Journal of Economic Perspectives* 15(4), pp. 143-156.
- [29] Koenker, R., and J. A. Machado (1999): “Goodness of Fit and Related Inference Processes for Quantile Regression,” *Journal of the American Statistical Association* 94(448), pp. 1296-1310.
- [30] Koenker, R. and Z. Xiao (2002): “Inference on the quantile regression process,” *Econometrica* 70, no. 4, pp. 1583–1612.
- [31] Lemieux, T. (2003): “Residual Wage Inequality: A Re-examination,” Mimeo. University of British Columbia, June.
- [32] Machado, J. A. and J. Mata (2003): “Counterfactual Decomposition of Changes in Wage Distributions using Quantile Regression,” Mimeo. Universidade Nova de Lisboa.
- [33] Politis, D. N., J. P. Romano and M. Wolf (1999): *Subsampling*. Springer-Verlag. New York.
- [34] Portnoy, S. (1991): “Asymptotic behavior of regression quantiles in nonstationary, dependent cases” *Journal of Multivariate Analysis* 38, no. 1, pp. 100-113.
- [35] Powell, J. (1986): “Symmetrically Trimmed Least Squares Estimation for Tobit Models,” *Econometrica* 54, no. 6, pp. 1435-1460.
- [36] Powell, J. (1994): “Estimation of Semiparametric Models,” in R.F. Engle and D.L. McFadden (eds.), *Handbook of Econometrics*, Volume IV. Amsterdam. Elsevier Science.
- [37] Ruggles, S. and M. Sobek *et al.* (2003): *Integrated Public Use Microdata Series: Version 3.0*. Minneapolis: Historical Census Project. University of Minnesota.
- [38] Silverman, B. W. (1986): *Density Estimation for Statistics and Data Analysis*. Chapman & Hall. London.
- [39] Van der Vaart, A. W. (1998): *Asymptotic Statistics*. Cambridge University Press. Cambridge, UK.

- [40] Van der Vaart, A. W. and J. A. Wellner (1996): *Weak Convergence and Empirical Processes. With Applications to Statistics*. Springer Series in Statistics. Springer-Verlag. New York.
- [41] White, H. (1980): "Using Least Squares to Approximate Unknown Regression Functions," *International Economic Review* 21(1), pp. 149-170.

**Table 1:** Human capital earnings function:  
Estimates of schooling coefficients and standard errors (%)

Census	Obs.	Desc. Stats.		Quantile Regression Estimates					OLS Estimates	
		Mean	SD	0.1	0.25	0.5	0.75	0.9	Coeff.	Root MSE
A. Without controls										
1980	65,023	6.40	0.67	7.48 (0.223) [0.239]	7.13 (0.078) [0.081]	6.39 (0.067) [0.067]	6.56 (0.069) [0.070]	7.42 (0.100) [0.110]	6.98 (0.080) [0.087]	0.63
1990	86,785	6.46	0.69	10.04 (0.130) [0.135]	9.57 (0.119) [0.121]	8.93 (0.075) [0.075]	9.23 (0.108) [0.107]	11.59 (0.169) [0.178]	9.78 (0.082) [0.087]	0.64
2000	97,397	6.50	0.75	9.80 (0.201) [0.208]	10.57 (0.129) [0.133]	11.05 (0.109) [0.109]	11.89 (0.115) [0.115]	15.51 (0.624) [0.669]	11.71 (0.092) [0.113]	0.69
B. Controlling for race and quadratic function of potential experience										
1980	65,023	6.40	0.67	7.35 (0.190) [0.199]	7.35 (0.120) [0.123]	6.83 (0.099) [0.099]	7.01 (0.104) [0.106]	7.91 (0.145) [0.153]	7.20 (0.120) [0.127]	0.63
1990	86,785	6.46	0.69	11.15 (0.274) [0.285]	10.96 (0.123) [0.126]	10.62 (0.104) [0.104]	11.08 (0.149) [0.148]	13.69 (0.252) [0.263]	11.36 (0.117) [0.122]	0.64
2000	97,397	6.50	0.75	9.16 (0.195) [0.204]	10.49 (0.120) [0.122]	11.13 (0.126) [0.126]	11.95 (0.134) [0.134]	15.73 (0.385) [0.401]	11.44 (0.117) [0.141]	0.69

Notes: US-born white and black men aged 40-49. Standard Errors in parentheses. Standard Errors robust to misspecification in brackets. Sampling weights used for 2000 Census.



**Table 2:** Comparison of CQF and QR-based Interquantile Spreads

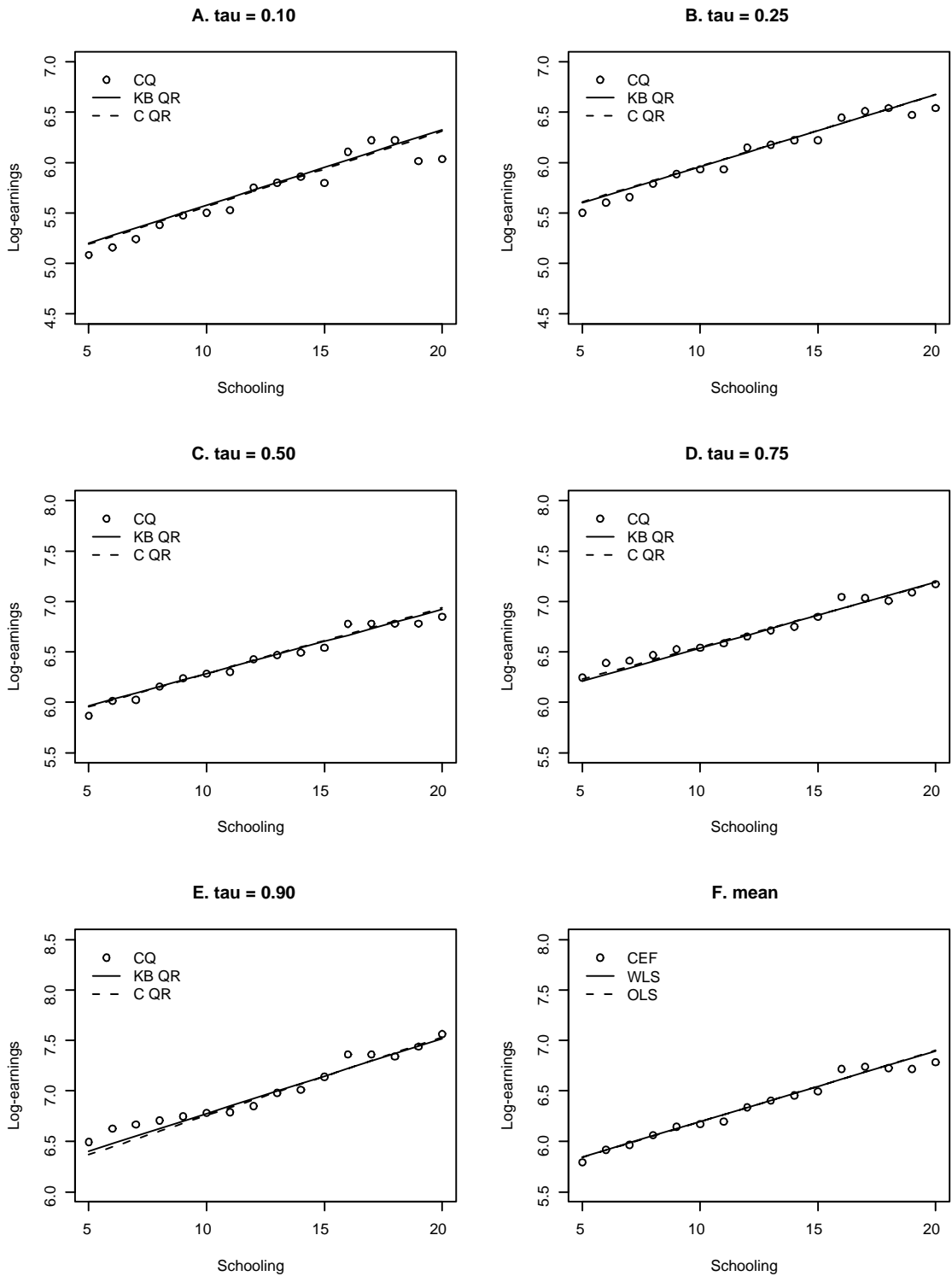
Census	Obs.	Controls	Interquantile Spread							
			90-10		75-25		90-50		50-10	
			CQ	QR	CQ	QR	CQ	QR	CQ	QR
A. Overall										
1980	65,023	No	1.20	1.20	0.56	0.56	0.51	0.52	0.69	0.68
		Yes	1.20	1.19	0.56	0.55	0.52	0.51	0.68	0.67
1990	86,785	No	1.37	1.36	0.65	0.65	0.61	0.61	0.76	0.75
		Yes	1.35	1.35	0.64	0.64	0.60	0.61	0.75	0.74
2000	97,397	No	1.45	1.45	0.71	0.70	0.68	0.69	0.77	0.76
		Yes	1.43	1.45	0.70	0.68	0.67	0.70	0.76	0.75
B. High School Graduates										
1980	25,020	No	1.10	1.20	0.51	0.57	0.42	0.51	0.67	0.69
		Yes	1.09	1.17	0.52	0.55	0.44	0.50	0.65	0.67
1990	22,837	No	1.27	1.33	0.64	0.66	0.51	0.56	0.76	0.77
		Yes	1.26	1.31	0.63	0.64	0.52	0.55	0.74	0.76
2000	25,963	No	1.32	1.34	0.68	0.67	0.60	0.61	0.72	0.73
		Yes	1.29	1.32	0.66	0.66	0.59	0.60	0.70	0.72
C. College Graduates										
1980	7,158	No	1.25	1.19	0.60	0.54	0.58	0.55	0.67	0.64
		Yes	1.26	1.19	0.59	0.53	0.61	0.54	0.65	0.64
1990	15,517	No	1.49	1.40	0.68	0.64	0.69	0.67	0.77	0.73
		Yes	1.44	1.38	0.66	0.63	0.70	0.66	0.74	0.72
2000	19,388	No	1.57	1.57	0.73	0.72	0.75	0.78	0.82	0.78
		Yes	1.55	1.57	0.74	0.71	0.75	0.80	0.80	0.78

Notes: US-born white and black men aged 40-49. Average measures calculated using the distribution of the covariates in each year. The covariates are schooling (controls = No) or schooling, race and a quadratic function of experience (controls = Yes). Sampling weights used for 2000 Census.

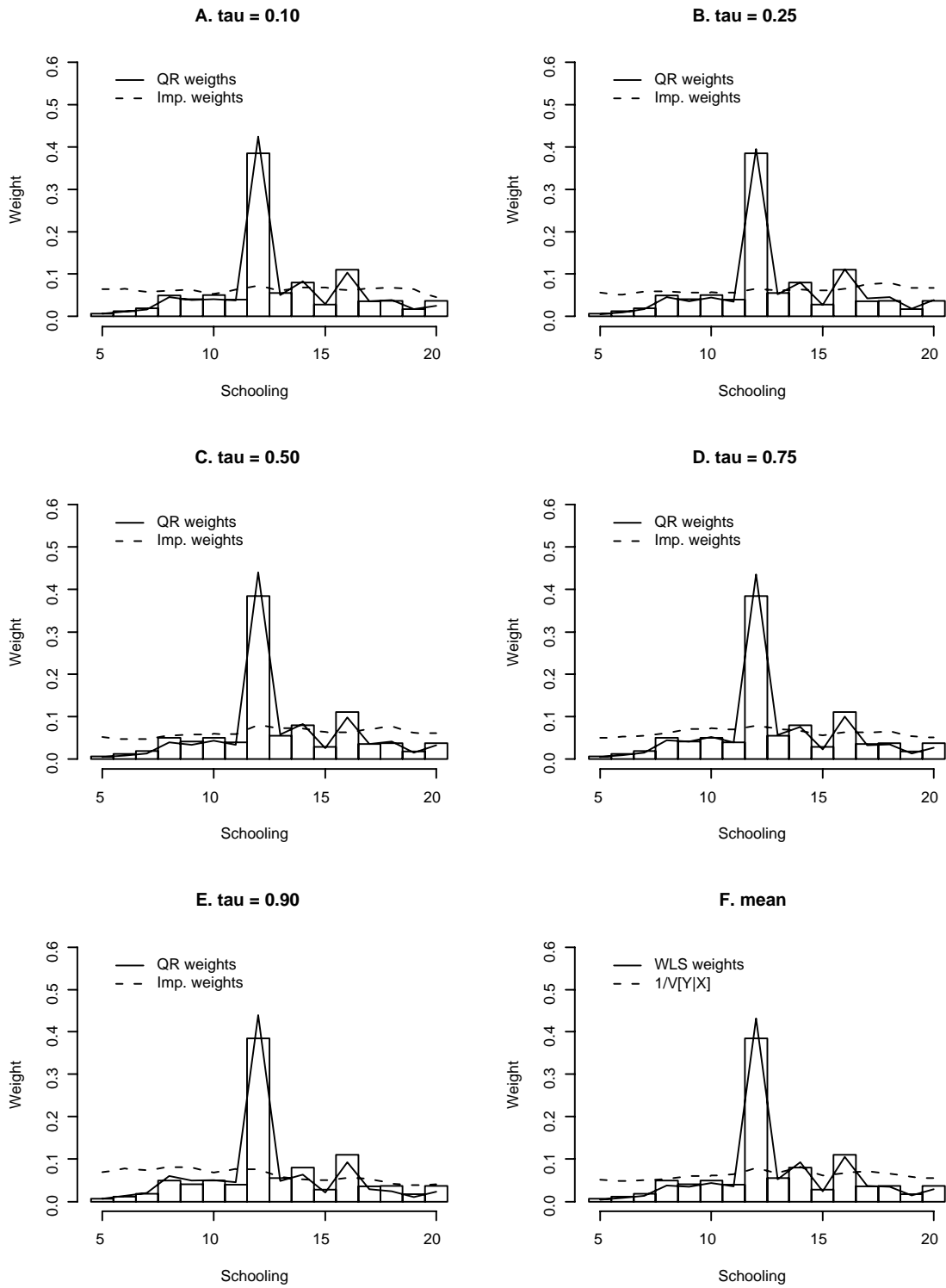
**Table 3:** Measures of Between-group (Model) and Within-group (Residual) Inequality and Linear (Quantile) Regression Approximations

Census	Obs.	Quantile-based Measures								ANOVA	
		90-10		75-25		90-50		50-10		Cond. Mean	OLS Fit
A. Between-group Inequality											
1980	65,023	0.60	0.59	0.15	0.23	0.35	0.32	0.25	0.27	0.24	0.23
1990	86,785	0.63	0.65	0.33	0.35	0.37	0.41	0.27	0.24	0.28	0.27
2000	97,397	0.66	0.75	0.51	0.43	0.42	0.53	0.24	0.22	0.30	0.29
B. Within-group Inequality											
1980	65,023	1.14	1.17	0.52	0.54	0.49	0.51	0.65	0.66	0.63	0.63
1990	86,785	1.32	1.35	0.62	0.63	0.57	0.59	0.73	0.75	0.63	0.64
2000	97,397	1.38	1.41	0.67	0.67	0.64	0.66	0.73	0.75	0.68	0.69
C. Relative Importance of Within-group Inequality (RTR and 1-R <sup>2</sup> )											
1980	65,023	78	80	93	85	65	72	87	86	87	88
1990	86,785	81	81	78	76	71	68	88	90	84	85
2000	97,397	82	78	63	71	70	61	90	92	84	85

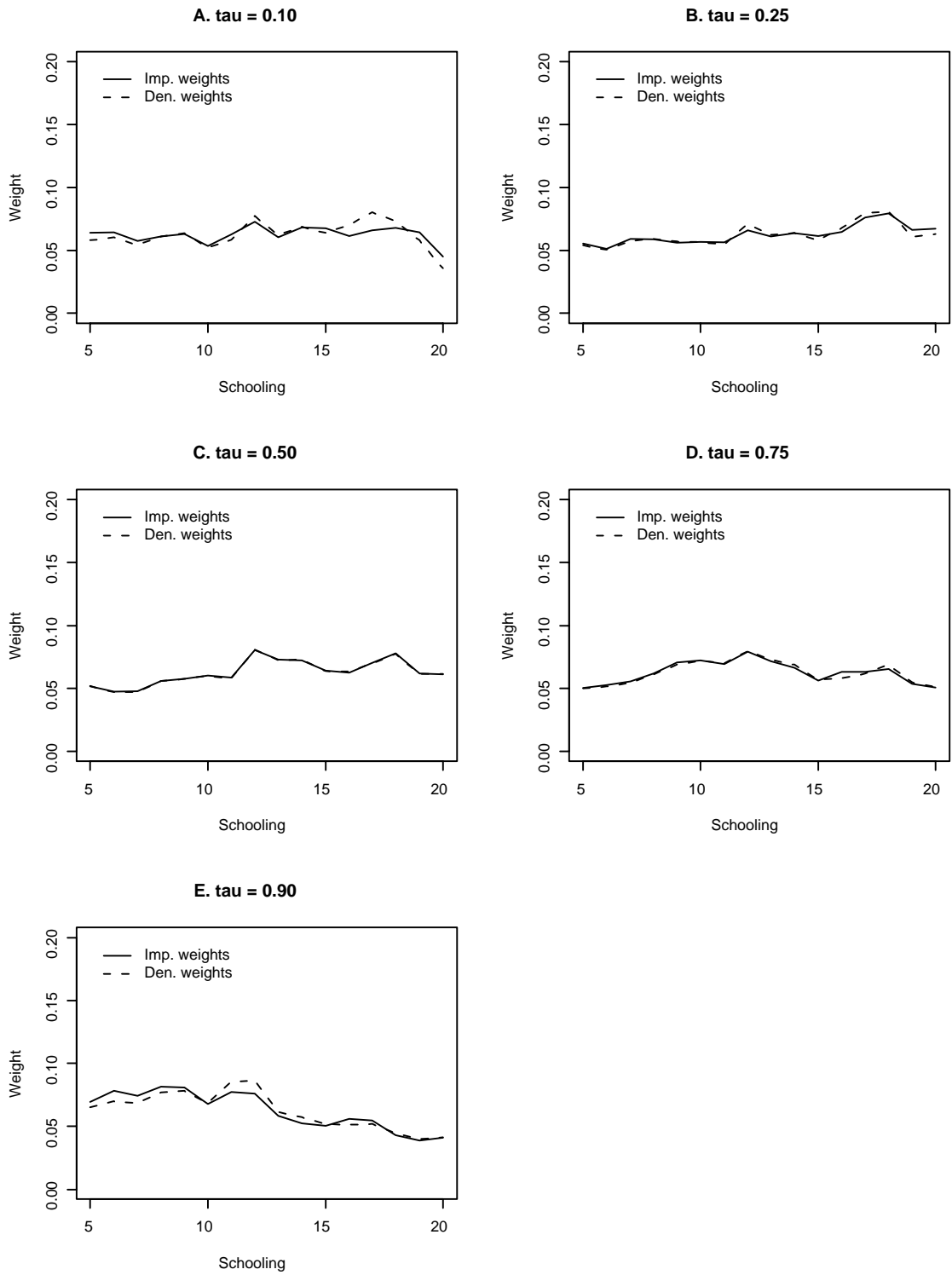
Notes: US-born white and black men aged 40-49. Measures calculated in a model that includes schooling, race and experience. Relative measures calculated as the square of Panel B divided by the sum of the square of Panel A and the square of Panel B. Sampling weights used for 2000 census.



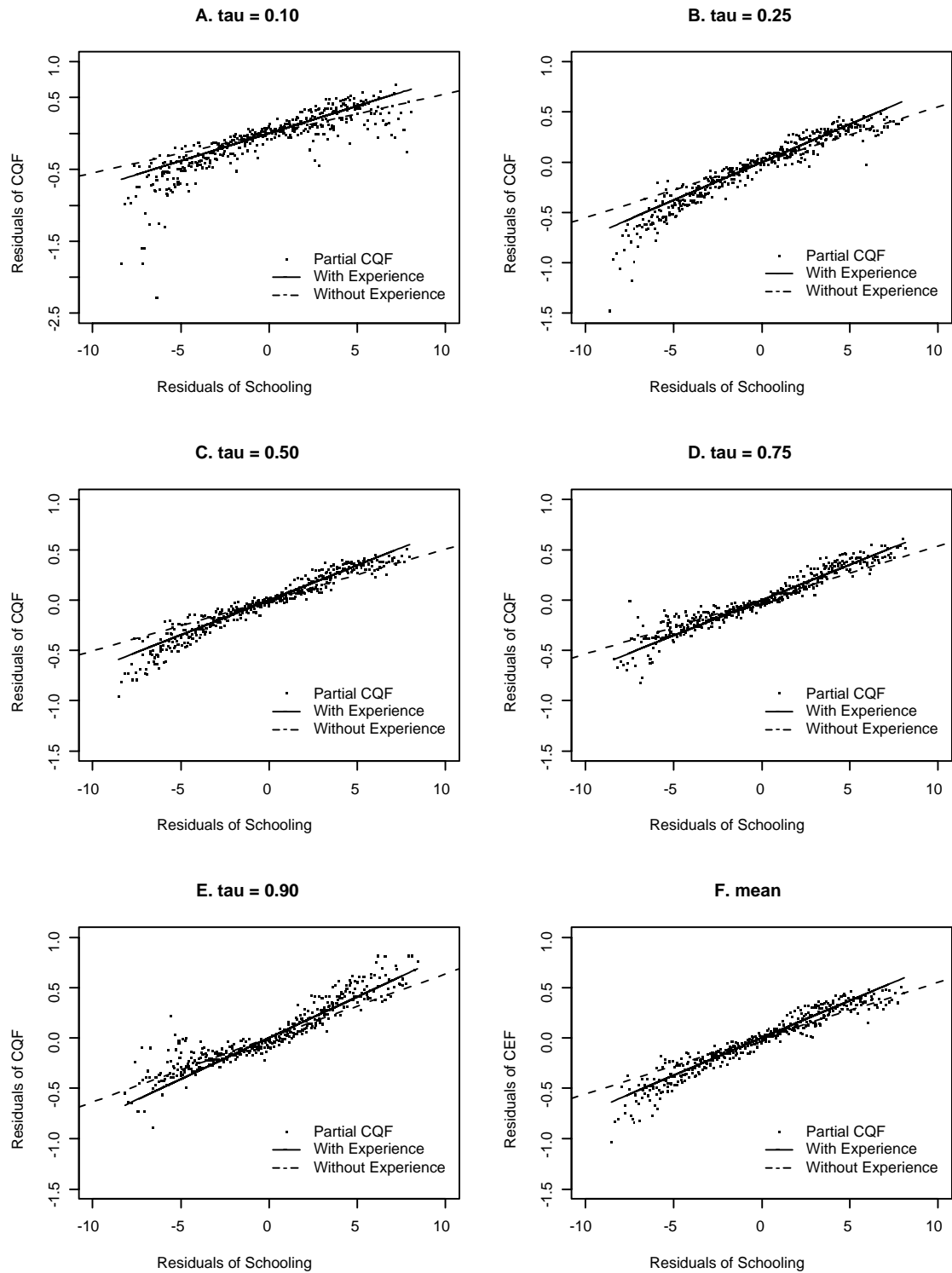
**Figure 1:** CQF and CEF in 1980 Census (US-born white and black men aged 40-49). Panels A - E plot the Conditional Quantile Function, Koenker and Basset's Quantile Regression fit and Chamberlain's Minimum Distance fit for weekly log-earnings given years of schooling. Panel F plots the Conditional Expectation Function (CEF), Weighted LS fit and OLS fit for weekly log-earnings given years of schooling.



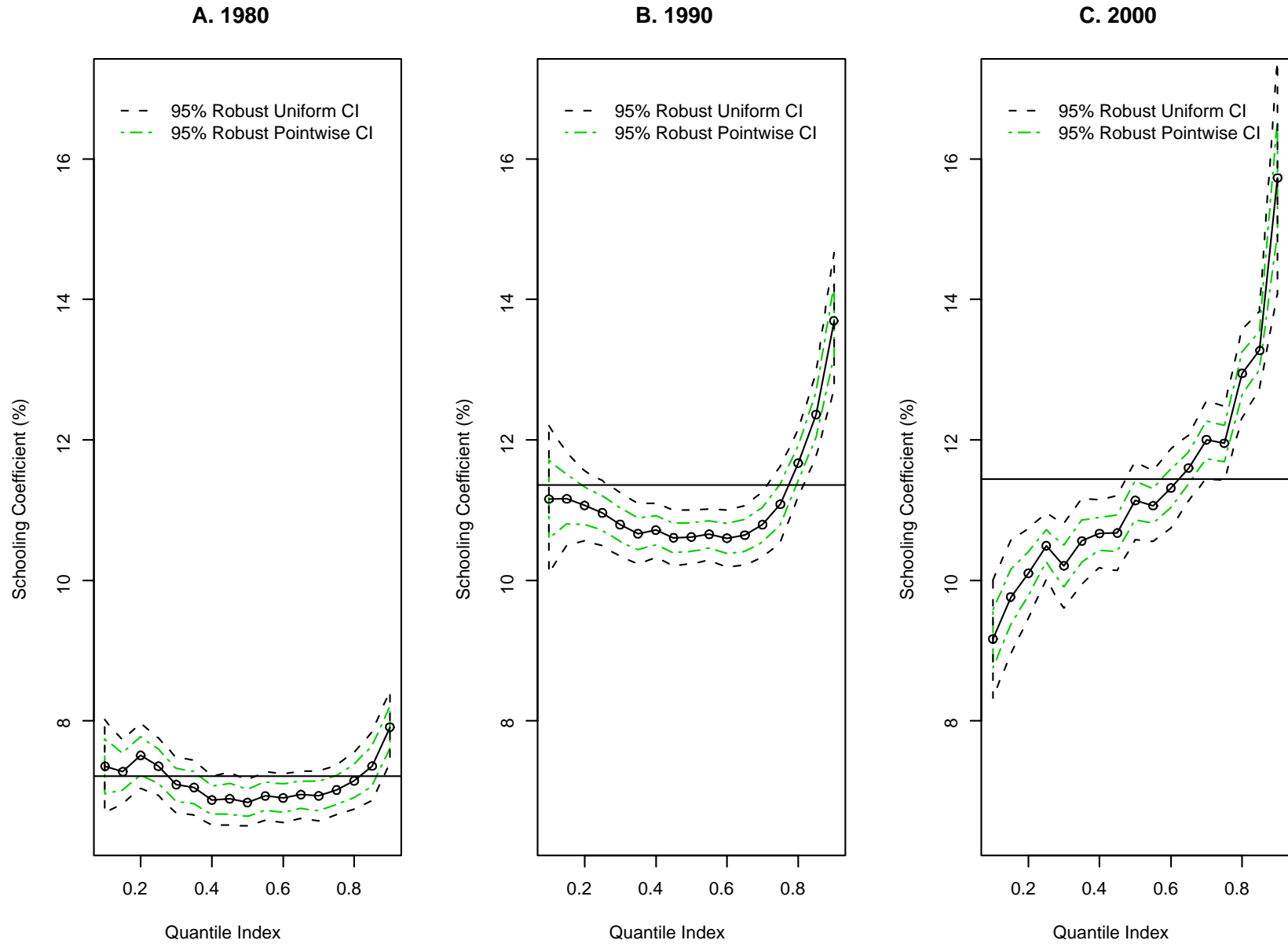
**Figure 2:** Weighting Functions in 1980 Census (US-born white and black men aged 40-49). Panels A-E plot the histogram of years of schooling, QR weighting function and importance weighting function for QR's of log-earnings on years of schooling. Panel F plots the histogram of years of schooling, WLS weighting function and inverse of the conditional variance for the linear regression of log-earnings on years of schooling.



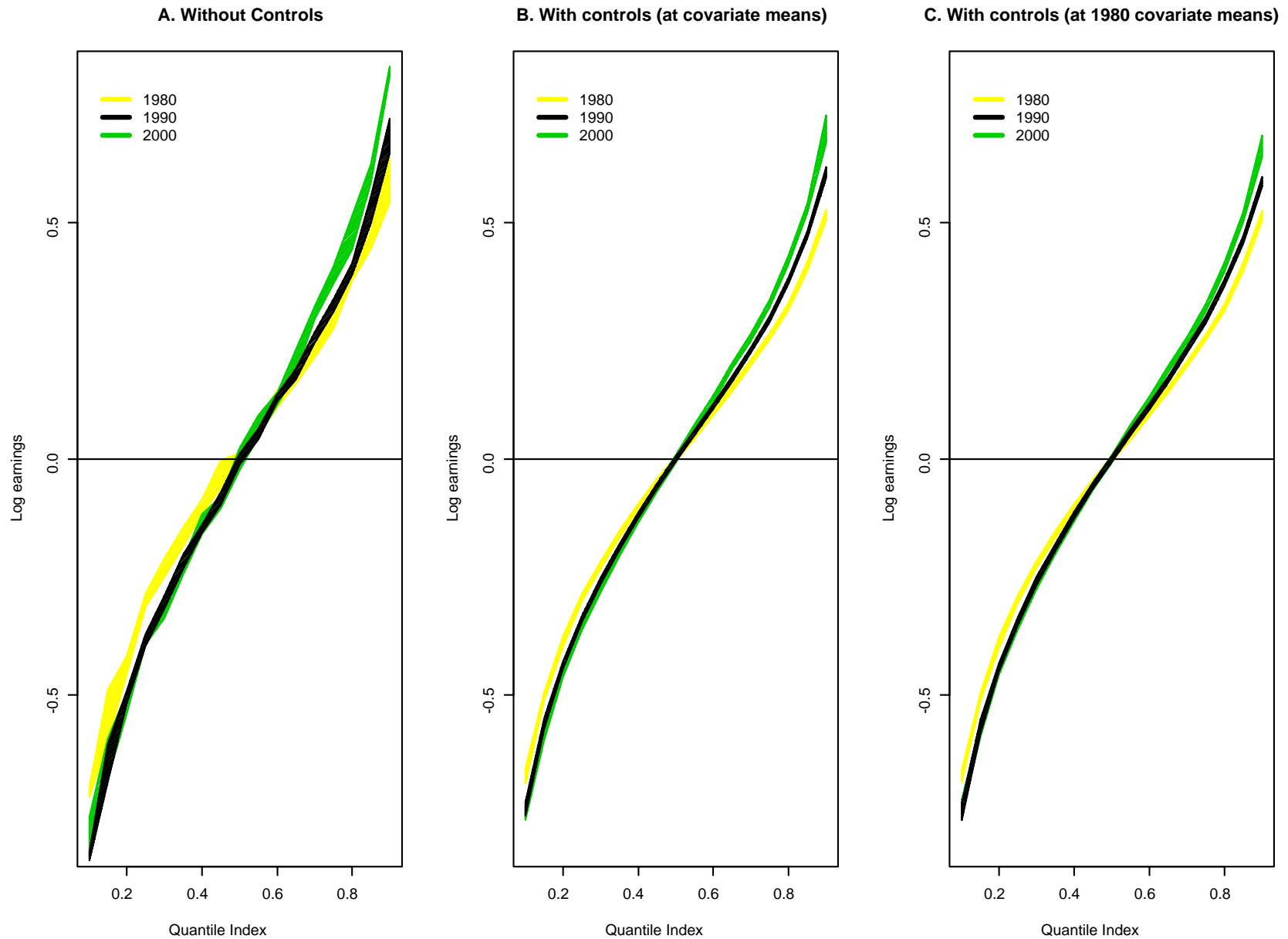
**Figure 3:** Importance and Density Weights in 1980 Census (US-born white and black men aged 40-49).



**Figure 4:** Partial Quantile Correlation Plots in 1980 Census (US-born white men aged 30-54). Panels A-E plot the Partial Conditional Quantile Function and Partial QR fit of log-earnings on years of schooling, controlling for a quadratic function of experience. The dashed line has the same slope as a QR line of log-earnings on years of schooling without controlling for experience. Panel F plots the Partial Conditional Expectation Function and Partial OLS fit of log-earnings on years of schooling, controlling for a quadratic function of experience. The dashed line has the same slope as a OLS line of log-earnings on years of schooling without controlling for experience.

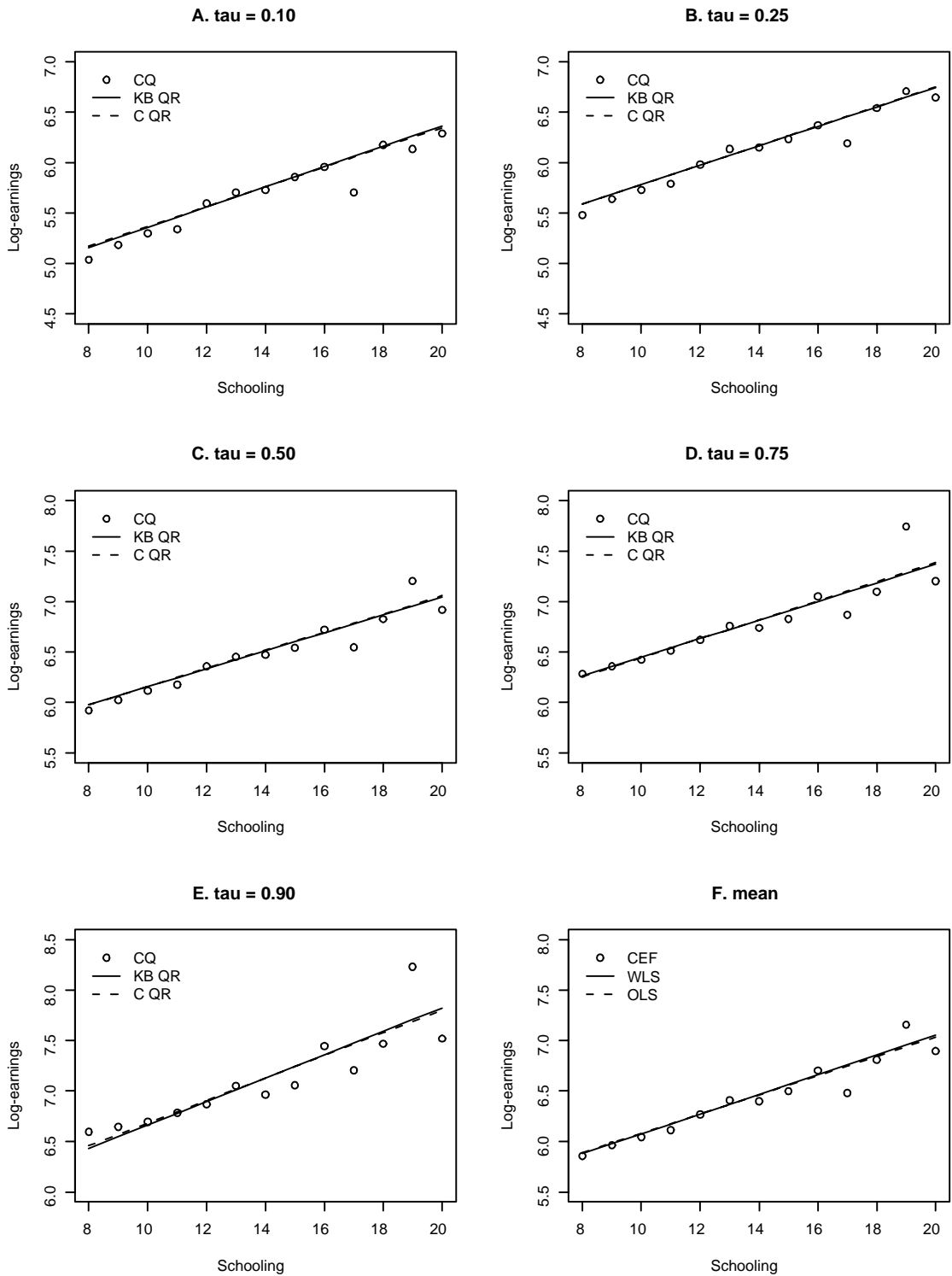


**Figure 5:** Schooling coefficients in 1980, 1990 and 2000 censuses (US-born white and black men aged 40-49). Panels A, B and C plot the Quantile Process for the coefficient of schooling in the QR of log-earnings on years of schooling, race and a quadratic function of experience, and 95% robust pointwise and uniform confidence intervals for 1980, 1990 and 2000, respectively. Uniform bands obtained by subsampling using 200 repetitions with subsample size  $b = 5n^{2/5}$ . The horizontal lines correspond to the OLS estimates of the schooling coefficient.

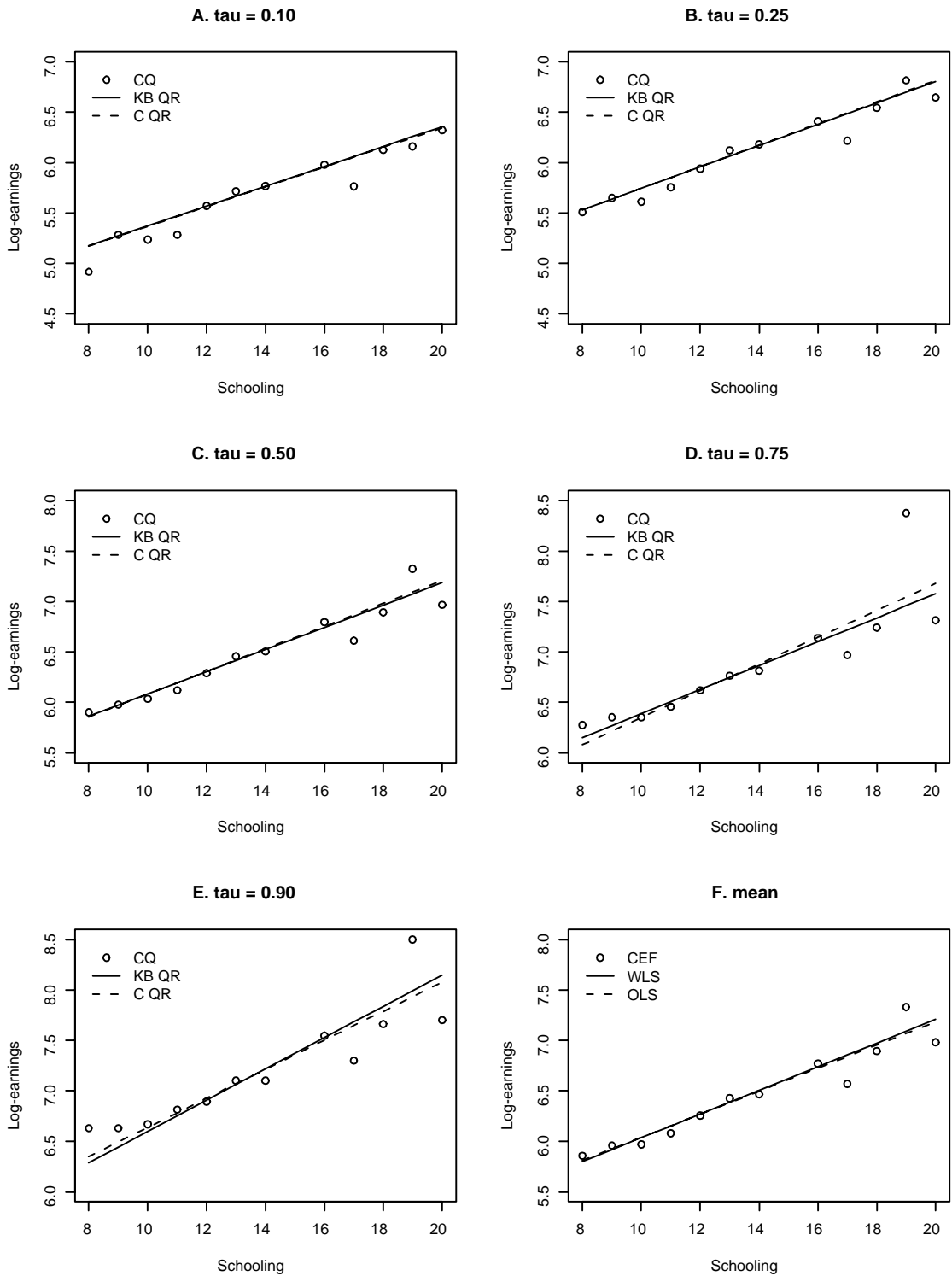


**Figure 6:** Unconditional quantiles and conditional quantiles of log-earnings in 1980, 1990 and 2000 censuses (US-born white and black men aged 40-49). Distributions are centered at median earnings for each year. Panel A plots 95 % uniform bands for unconditional quantiles. Panels B and C plot 95% uniform bands for the QR approximation to the conditional quantile function given schooling, race and a quadratic function of experience. In Panel B covariates are evaluated at sample mean values for each year and schooling is coded using IPUMS categories for each year. In Panel C covariates are evaluated at sample mean values for 1980 and schooling is recoded using 2000 IPUMS categories.

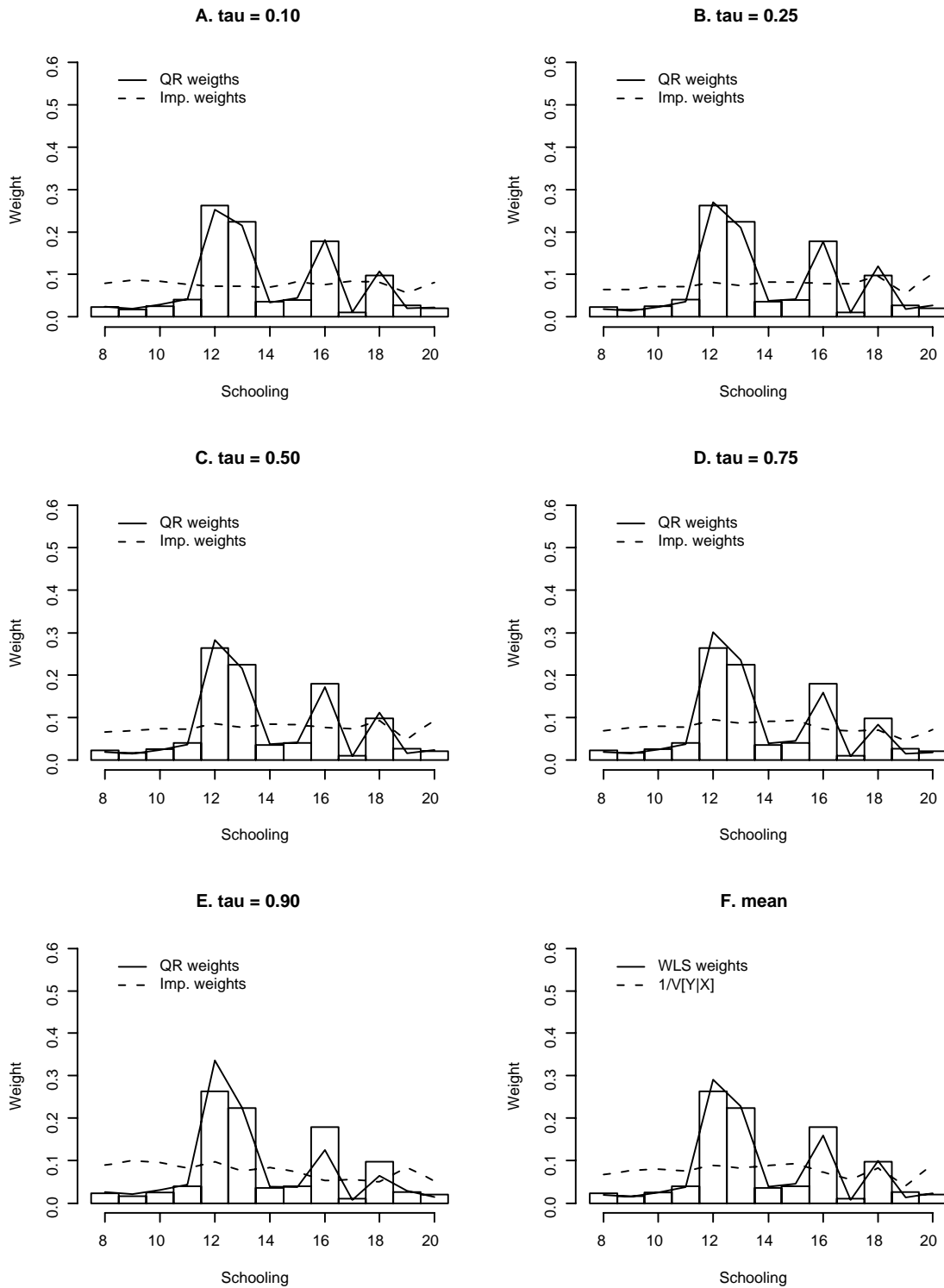




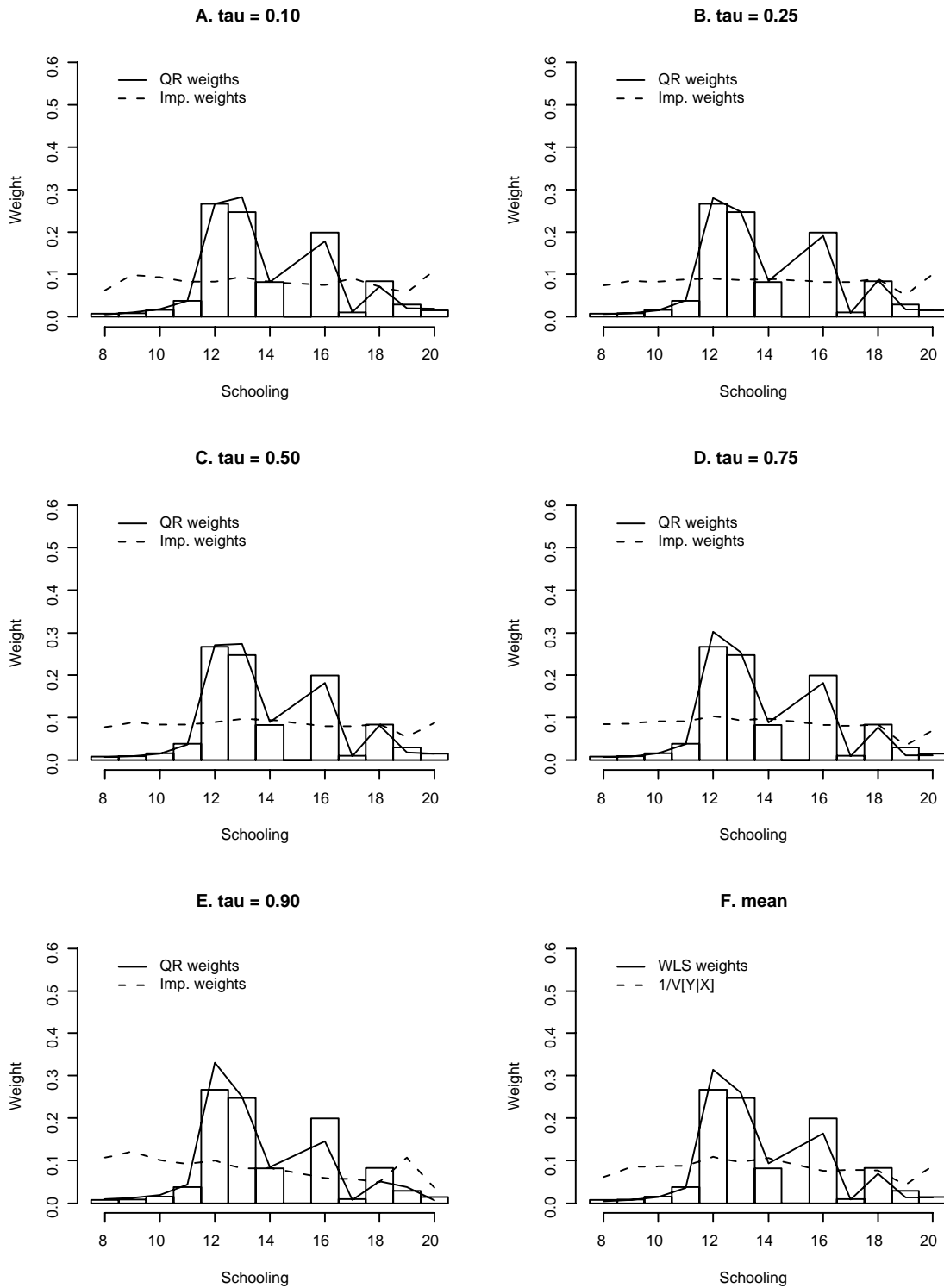
**Figure 7:** CQF and CEF in 1990 Census (US-born white and black men aged 40-49). Panels A - E plot the Conditional Quantile Function, Koenker and Basset's Quantile Regression fit and Chamberlain's Minimum Distance fit for weekly log-earnings given years of schooling. Panel F plots the Conditional Expectation Function (CEF), Weighted LS fit and OLS fit for weekly log-earnings given years of schooling.



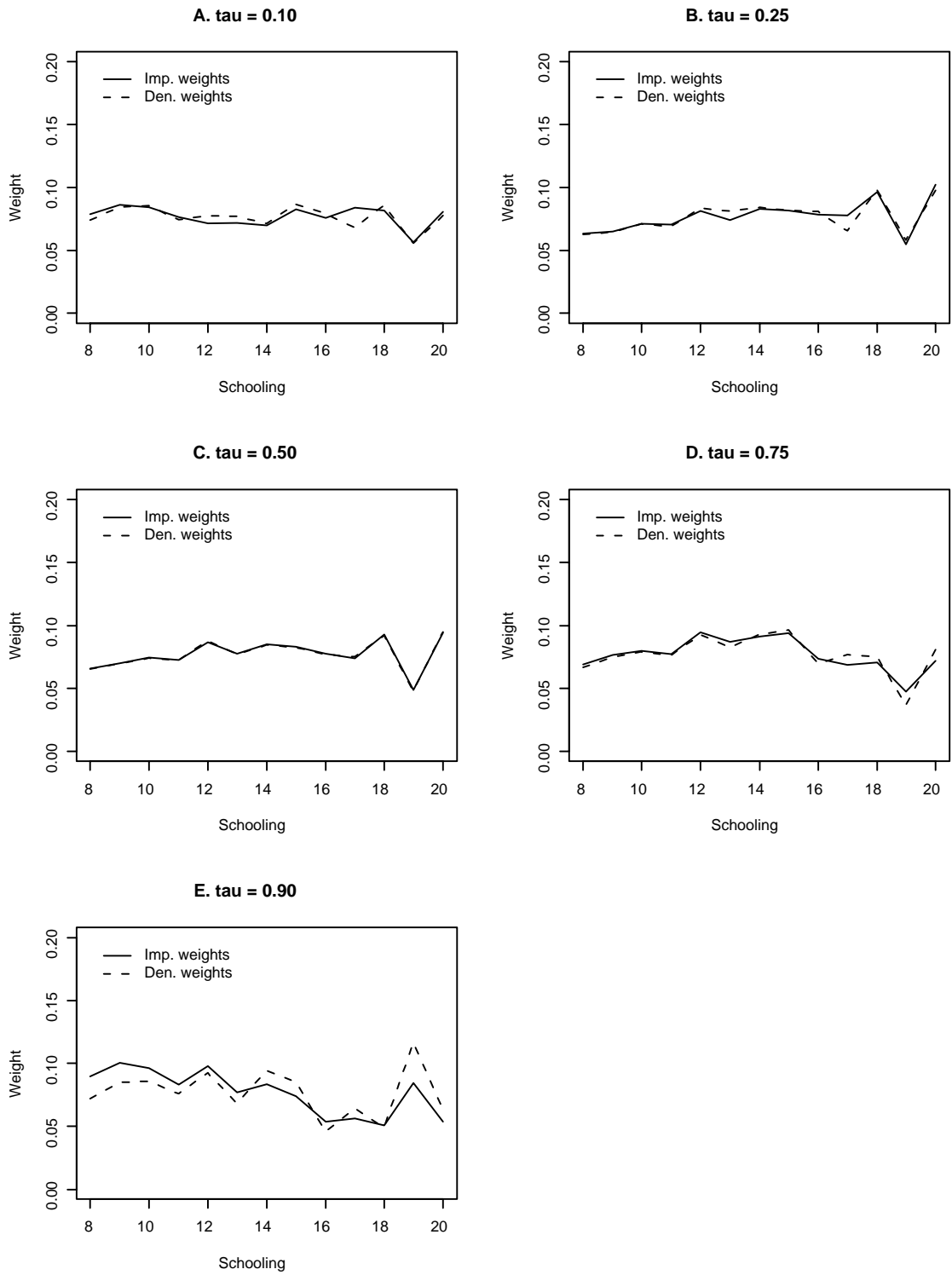
**Figure 8:** CQF and CEF in 2000 Census (US-born white and black men aged 40-49). Panels A - E plot the Conditional Quantile Function, Koenker and Basset's Quantile Regression fit and Chamberlain's Minimum Distance fit for weekly log-earnings given years of schooling. Panel F plots the Conditional Expectation Function (CEF), Weighted LS fit and OLS fit for weekly log-earnings given years of schooling.



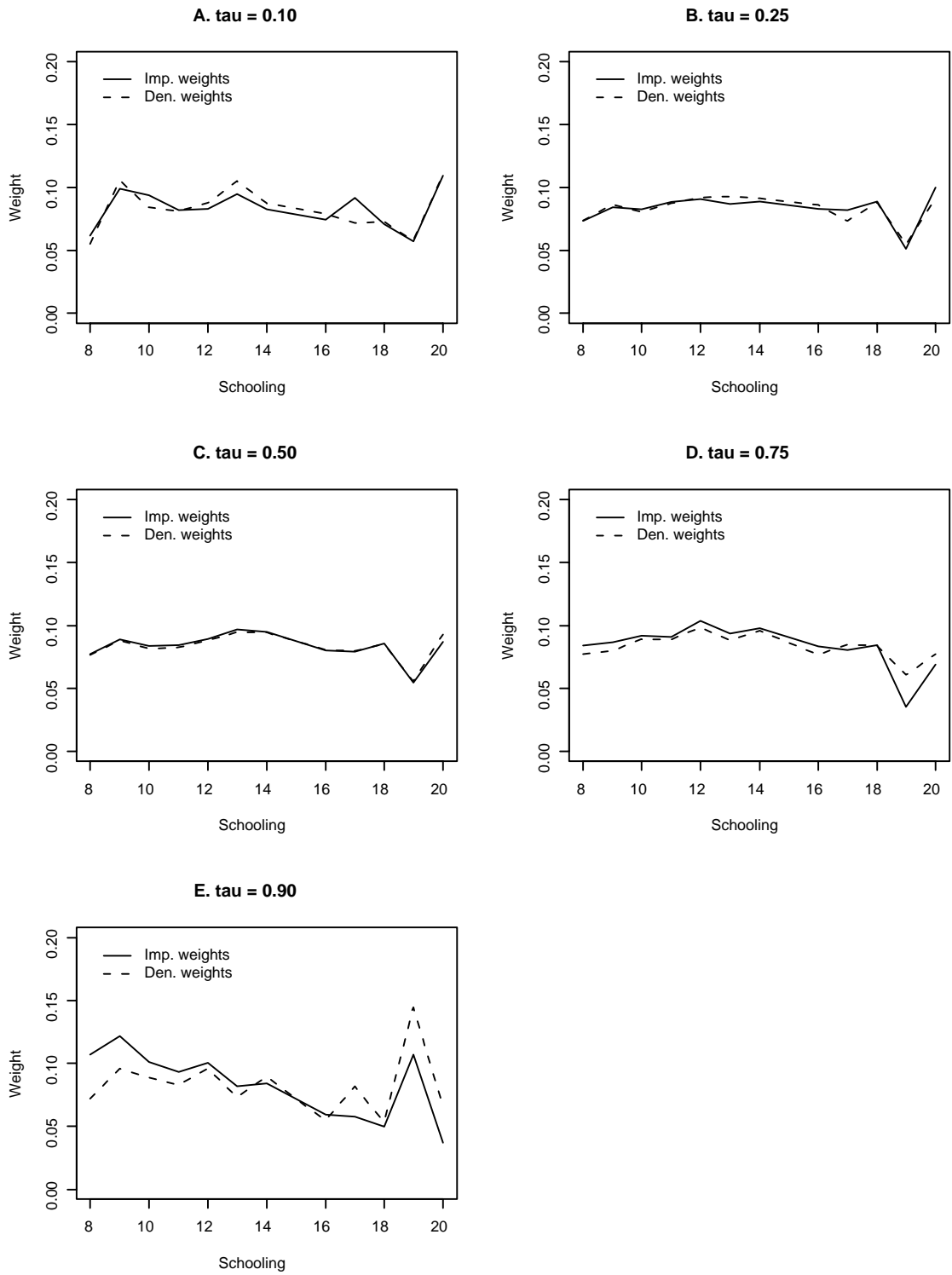
**Figure 9:** Weighting Functions in 1990 Census (US-born white and black men aged 40-49). Panels A-E plot the histogram of years of schooling, QR weighting function and importance weighting function for QR's of log-earnings on years of schooling. Panel F plots the histogram of years of schooling, WLS weighting function and inverse of the conditional variance for the linear regression of log-earnings on years of schooling.



**Figure 10:** Weighting Functions in 2000 Census (US-born white and black men aged 40-49). Panels A-E plot the histogram of years of schooling, QR weighting function and importance weighting function for QR's of log-earnings on years of schooling. Panel F plots the histogram of years of schooling, WLS weighting function and inverse of the conditional variance for the linear regression of log-earnings on years of schooling.



**Figure 11:** Importance and Density Weights in 1990 Census (US-born white and black men aged 40-49).



**Figure 12:** Importance and Density Weights in 2000 Census (US-born white and black men aged 40-49).