cemmap

centre for microdata methods and practice

Quantile regression with censoring and endogeneity

Victor Chernozhukov
Iván Fernández-Val
Amanda Kowalski

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP20/11

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

An ESRC Research Centre

Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE **Tel:** 44 (0)20 7291 4800 **Email:** info@cemmap.ac.uk **www.cemmap.ac.uk**

QUANTILE REGRESSION WITH CENSORING AND ENDOGENEITY

VICTOR CHERNOZHUKOV† IVÁN FERNÁNDEZ-VAL§ AMANDA KOWALSKI‡

Abstract. In this paper, we develop a new censored quantile instrumental variable (CQIV) estimator and describe its properties and computation. The CQIV estimator combines Powell (1986) censored quantile regression (CQR) to deal semiparametrically with censoring, with a control variable approach to incorporate endogenous regressors. The CQIV estimator is obtained in two stages that are nonadditive in the unobservables. The first stage estimates a nonadditive model with infinite dimensional parameters for the control variable, such as a quantile or distribution regression model. The second stage estimates a nonadditive censored quantile regression model for the response variable of interest, including the estimated control variable to deal with endogeneity. For computation, we extend the algorithm for CQR developed by Chernozhukov and Hong (2002) to incorporate the estimation of the control variable. We give generic regularity conditions for asymptotic normality of the CQIV estimator and for the validity of resampling methods to approximate its asymptotic distribution. We verify these conditions for quantile and distribution regression estimation of the control variable. We illustrate the computation and applicability of the CQIV estimator with numerical examples and an empirical application on estimation of Engel curves for alcohol.

Date: April 23, 2011. We thank Denis Chetverikov and Sukjin Han for excellent comments and capable research assistance. We are grateful to Richard Blundell for providing us the data for the empirical application. Stata software to implement the methods developed in the paper is available in Amanda Kowalski's web site at http://www.econ.yale.edu/ ak669/research.html. We gratefully acknowledge research support from the NSF.

† Department of Economics, MIT, 50 Memorial Drive, Cambridge, MA 02142, vchern@mit.edu.
§ Boston University, Department of Economics, 270 Bay State Road,Boston, MA 02215, ivanf@bu.edu.
‡ Department of Economics, Yale University, 37 Hillhouse Avenue, New Haven, CT 06520, and NBER, kowalski@nber.org.

1. Introduction

Censoring and endogeneity are common problems in data analysis. For example, income survey data are often top-coded and many economic variables such as hours worked, wages and expenditure shares are naturally bounded from below by zero. Endogeneity is also an ubiquitous phenomenon both in experimental studies due to partial noncompliance (Angrist, Imbens, and Rubin, 1996), and in observational studies due to simultaneity (Koopmans and Hood, 1953), measurement error (Frish, 1934), sample selection (Heckman, 1979) or more generally to the presence of relevant omitted variables. Censoring and endogeneity often come together. Thus, for example, we motivate our analysis with the estimation of Engel curves for alcohol – the relationship between the share of expenditure on alcohol and the household's budget. For this commodity, more than 15% of the households in our sample report zero expenditure, and economic theory suggests that total expenditure and its composition are jointly determined in the consumption decision of the household. Either censoring or endogeneity lead to inconsistency of traditional mean and quantile regression estimators by inducing correlation between regressors and error terms. We introduce a quantile regression estimator that deals with both problems and name this estimator the censored quantile instrumental variable (CQIV) estimator.

Our procedure deals with censoring semiparametrically through the conditional quantile function following Powell (1986). This approach avoids the strong parametric assumptions of traditional Tobit estimators. The key ingredient here is the equivariance property of quantile functions to monotone transformations such as censoring. Powell's censored quantile regression estimator, however, has proven to be difficult to compute. We address this problem using the computationally attractive algorithm of Chernozhukov and Hong (2002). An additional advantage of focusing on the conditional quantile function is that we can capture nonadditive heterogeneity in the effects of the regressors across the distribution of the response variable by computing CQIV at different quantiles (Koenker, 2005). The traditional Tobit framework rules out this heterogeneity by imposing a location shift model.

We deal with endogeneity using a control variable approach. The basic idea is to add a variable to the regression such that, once we condition on this variable, regressors and error terms become independent. This so-called control variable is usually unobservable and need to be estimated in a first stage. Our main contribution here is to allow for semiparametric models with infinite dimensional parameters and nonadditive error terms, such as quantile regression and distribution regression, to model and estimate the first stage and back out the control variable. This part of the analysis constitutes the main theoretical difficulty because the first stage estimators do not live in spaces with nice entropic properties, unlike, for example, in Andrews (1994) or Newey (1994). To overcome this problem, we develop a new technique to derive asymptotic theory for two-stage procedures with plugged-in first stage estimators that, while not living in Donsker spaces themselves, can be suitably approximated by random functions that live in Donsker spaces. The CQIV estimator is therefore obtained in two stages that are nonadditive in the unobservables. The first stage estimates the control variable, whereas the second stage estimates a nonadditive censored quantile regression model for the response variable of interest, including the estimated control variable to deal with endogeneity.

We analyze the theoretical properties of the CQIV estimator in large samples. Under suitable regularity conditions, CQIV is $\sqrt{n}$-consistent and has a normal limiting distribution. We characterize the expression of the asymptotic variance. Although this expression can be estimated using standard methods, we find it more convenient to use resampling methods for inference. We focus on weighted bootstrap because it has practical advantages over nonparametric bootstrap to deal with discrete regressors with small cell sizes (Ma and Kosorok, 2005, and Chen and Pouzo, 2009). We give regularity conditions for the consistency of weighted bootstrap to approximate the distribution of the CQIV estimator. For our leading cases of quantile and distribution regression estimation of the control variable, we provide more primitive assumptions that verify the regularity conditions for asymptotic normality and weighted bootstrap consistency. The verification of these conditions for quantile and distribution regression estimators of the first stage is new to the best of our knowledge.

The CQIV estimator is simple to compute using standard statistical software. We demonstrate its implementation through Monte-Carlo simulations and an empirical application to the estimation of Engel curves for alcohol. The results of the Monte-Carlo exercise demonstrate that the performance of CQIV is comparable to that of Tobit IV in data generated to satisfy the Tobit IV assumptions, and it outperforms Tobit IV under heteroskedasticity. The results of the application to Engel curves demonstrate the importance of accounting for endogeneity and censoring in real data. Another application of our CQIV estimator to the estimation of the price elasticity of expenditure on medical care appears in Kowalski (2009).

1.1. **Literature review.** There is an extensive previous literature on the control variable approach to deal with endogeneity in models without censoring. Hausman (1978) and Wooldridge (2010) discussed parametric triangular linear and nonlinear models. Newey, Powell, and Vella (1999) described the use of this approach in nonparametric triangular systems of equations for the conditional mean, but limited the analysis to models with additive errors both in the first and the second stage. Lee (2007) set forth an estimation strategy using a control variable approach for a triangular system of equations for conditional quantiles with an additive nonparametric first stage. Imbens and Newey (2002, 2009) extended the analysis to triangular nonseparable models with nonadditive error terms in both the first and second stage. They focused on identification and nonparametric estimation rates for average, quantile and policy effects. Our paper complements Imbens and Newey (2002, 2009) by providing inference methods and allowing for censoring. Chesher (2003) and Jun (2009) considered local identification and semiparametric estimation of uncensored triangular quantile regression models with a nonseparable control variable. Relative to CQIV, these local methods impose less structure in the model at the cost of slower rates of convergence in estimation. While the previous papers focused on triangular models, Blundell and Matzkin (2010) have recently derived conditions for the existence of control variables in nonseparable simultaneous equations models. We refer also to Matzkin (2007) for an excellent comprehensive review of results on nonparametric identification of triangular and simultaneous equations models.

Our work is also closely related to Ma and Koenker (2006). They considered identification and estimation of quantile effects without censoring using a parametric control variable. Their parametric assumptions rule out the use of nonadditive models with infinite dimensional parameters in the first stage, such as quantile and distribution regression models in the first stage. In contrast, our approach is specifically designed to handle the latter, and in doing so, it puts the first stage and second stage models on the equally flexible footing. Allowing for a nonadditive infinite dimensional control variable makes the analysis of the asymptotic properties of our estimator very delicate and requires developing new proof techniques. In particular, we need to deal with control variable estimators depending on random functions that do not live in Donsker classes. We address this difficulty approximating these functions with sufficient degree of accuracy by smoother functions that live in Donsker classes. In the case of quantile and distribution regression, we carry out this approximation by smoothing the empirical quantile regression and distribution regression processes using third order kernels.

For models with censoring, the literature is more sparse. Smith and Blundell (1986) pioneered the use of the control variable approach to estimate a triangular parametric additive model for the conditional mean. More recently, Blundell and Powell (2007) proposed an alternative censored quantile instrumental variable estimator that assumes additive errors in the first stage. Our estimator allows for a more flexible nonadditive first stage specification.

1.2. **Plan of the paper.** The rest of the paper is organized as follows. In Section 2, we present the CQIV model, and develop estimation and inference methods for the parameters of interest of this model. In Section 3, we describe the associated computational algorithms and present results from a Monte-Carlo simulation exercise. In Section 4, we present an empirical application of CQIV to Engel curves. In Section 5, we provide conclusions and discuss potential empirical applications of CQIV. The proofs of the main results, and additional details on the computational algorithms and numerical examples are given in the Appendix.

2. Censored Quantile Instrumental Variable Regression

2.1. **The Model.** We consider the following triangular system of quantile equations:

$$Y = \max(Y^*, C), \tag{2.1}$$

$$Y^* = Q_{Y^*}(U \mid D, W, V), \tag{2.2}$$

$$D = Q_D(V \mid W, Z). \tag{2.3}$$

In this system, Y^* is a continuous latent response variable, the observed variable Y is obtained by censoring Y^* from below at the level determined by the variable C, D is the continuous regressor of interest, W is a vector of covariates, possibly containing C, V is a latent unobserved regressor that accounts for the possible endogeneity of D, and Z is a vector of "instrumental variables" excluded from (2.2).[1] Further, $u \mapsto Q_{Y^*}(u \mid D, W, V)$ is the conditional quantile function of Y^* given (D, W, V); and $v \mapsto Q_D(v \mid W, Z)$ is the conditional quantile function of the regressor D given (W, Z). Here, U is a Skorohod disturbance for Y that satisfies the independence assumption

$$U \sim U(0,1) \mid D, W, Z, V, C,$$

and V is a Skorohod disturbance for D that satisfies

$$V \sim U(0,1) \mid W, Z, C.$$

In the last two equations, we make the assumption that the censoring variable C is independent of the disturbances U and V. This variable can, in principle, be included in W. To recover the conditional quantile function of the latent response variable in equation (2.2), it is important to condition on an unobserved regressor V which plays the role of a "control variable." Equation (2.3) allows us to recover this unobserved regressor as a residual that explains movements in the variable D, conditional on the set of instruments and other covariates.

In the Engel curve application, Y is the expenditure share in alcohol, bounded from below at $C = 0$, D is total expenditure on nondurables and services, W are household demographic characteristics, and Z is labor income measured by the earnings of the head of the household. Total expenditure is likely to be jointly determined with the budget composition in the household's allocation of income across consumption goods and leisure. Thus, households

[1] We focus on left censored response variables without loss of generality. If Y is right censored at C, $Y = \min(Y^*, C)$, the analysis of the paper applies without change to $\widetilde{Y} = -Y$, $\widetilde{Y}^* = -Y^*$, $\widetilde{C} = -C$, and $Q_{\widetilde{Y}^*} = -Q_{Y^*}$, because $\widetilde{Y} = \max(\widetilde{Y}^*, \widetilde{C})$.

with a high preference to consume "non-essential" goods such as alcohol tend to expend a higher proportion of their incomes and therefore to have a higher expenditure. The control variable V in this case is the marginal propensity to consume, measured by the household ranking in the conditional distribution of expenditure given labor income and household characteristics. This propensity captures unobserved preference variables that affect both the level and composition of the budget. Under the conditions for a two stage budgeting decision process (Gorman, 1959), where the household first divides income between consumption and leisure/labor and then decide the consumption allocation, some sources of income can provide plausible exogenous variation with respect to the budget shares. For example, if preferences are weakly separable in consumption and leisure/labor, the consumption budget shares do not depend on labor income given the consumption expenditure (see, e.g., Deaton and Muellbauer, 1980). This justifies the use of labor income as an exclusion restriction.

An example of a structural model that has the triangular representation (2.2)-(2.3) is the following system of equations:

$$
\begin{aligned}
Y^* &= g_Y(D, W, \epsilon), && (2.4)\\
D &= g_D(W, Z, V), && (2.5)
\end{aligned}
$$

where g_Y and g_D are increasing in their third arguments, and $\epsilon \sim U(0,1)$ and $V \sim U(0,1)$ independent of (W, Z, C). By the Skorohod representation for ϵ, $\epsilon = Q_\epsilon(U \mid V) = g_\epsilon(V, U)$, where $U \sim U(0,1)$ independent of (D, W, Z, V, C). The corresponding conditional quantile functions have the form of (2.2) and (2.3) with

$$
\begin{aligned}
Q_{Y^*}(u \mid D, W, V) &= g_Y(D, W, g_\epsilon(V, u)),\\
Q_D(v \mid W, Z) &= g_D(W, Z, v).
\end{aligned}
$$

In the Engel curve application, we can interpret V as the marginal propensity to consume out of labor income and U as the unobserved household preference to spend on alcohol relative to households with the same characteristics W and marginal propensity to consume V.

In the system of equations (2.1)–(2.3), the observed response variable has the quantile representation

$$
Y = Q_Y(U \mid D, W, V, C) = \max(Q_{Y^*}(U \mid D, W, V), C), \tag{2.6}
$$

by the equivariance property of the quantiles to monotone transformations. For example, the quantile function for the observed response in the system of equations (2.4)–(2.5) has the form:

$$
Q_Y(u \mid D, W, V, C) = \max\{g_Y(D, W, g_\epsilon(V, u)), C\}.
$$

Whether the response of interest is the latent or observed variable depends on the source of censoring (e.g., Wooldridge, 2010). When censoring is due to data limitations such as top-coding, we are often interested in the conditional quantile function of the latent response variable Q_{Y^*} and marginal effects derived from this function. For example, in the system (2.4)–(2.5) the marginal effect of the endogenous regressor D evaluated at $(D, W, V, U) = (d, w, v, u)$ is

$$\partial_d Q_{Y^*}(u \mid d, w, v) = \partial_d g_Y(d, w, g_\epsilon(v, u)),$$

which corresponds to the ceteris paribus effect of a marginal change of D on the latent response Y^* for individuals with $(D, W, \epsilon) = (d, w, g_\epsilon(v, u))$. When the censoring is due to economic or behavioral reasons such are corner solutions, we are often interested in the conditional quantile function of the observed response variable Q_Y and marginal effects derived from this function. For example, in the system (2.4)–(2.5) the marginal effect of the endogenous regressor D evaluated at $(D, W, V, U, C) = (d, w, v, u, c)$ is

$$\partial_d Q_Y(u \mid d, w, v, c) = 1\{g_Y(d, w, g_\epsilon(v, u)) > c\} \partial_d g_Y(d, w, g_\epsilon(v, u)),$$

which corresponds to the ceteris paribus effect of a marginal change of D on the observed response Y for individuals with $(D, W, \epsilon, C) = (d, w, g_\epsilon(v, u), c)$. Since either of the marginal effects might depend on individual characteristics, average marginal effects or marginal effects evaluated at interesting values are often reported.

2.2. **Generic Estimation.** To make estimation both practical and realistic, we impose a flexible semiparametric restriction on the functional form of the conditional quantile function in (2.2). In particular, we assume that

$$Q_{Y^*}(u \mid D, W, V) = X'\beta_0(u), \quad X = x(D, W, V), \tag{2.7}$$

where $x(D, W, V)$ is a vector of transformations of the initial regressors (D, W, V). The transformations could be, for example, polynomial, trigonometric, B-spline or other basis functions that have good approximating properties for economic problems. An important property of this functional form is linearity in parameters, which is very convenient for computation. The resulting conditional quantile function of the censored random variable

$$Y = \max(Y^*, C),$$

is given by

$$Q_Y(u \mid D, W, V, C) = \max(X'\beta_0(u), C). \tag{2.8}$$

This is the standard functional form for the censored quantile regression (CQR) first derived by Powell (1984) in the exogenous case.

Given a random sample $\{Y_i, D_i, W_i, Z_i, C_i\}_{i=1}^n$, we form the estimator for the parameter $\beta_0(u)$ as

$$\widehat{\beta}(u) = \arg \min_{\beta \in \mathbb{R}^{\dim(X)}} \frac{1}{n} \sum_{i=1}^{n} 1(\widehat{S}_i' \widehat{\gamma} > \varsigma) \rho_u(Y_i - \widehat{X}_i' \beta), \tag{2.9}$$

where $\rho_u(z) = (u - 1(z < 0))z$ is the asymmetric absolute loss function of Koenker and Bassett (1978), $\widehat{X}_i = x(D_i, W_i, \widehat{V}_i)$, $\widehat{S}_i = s(\widehat{X}_i, C_i)$, $s(X, C)$ is a vector of transformations of (X, C), and $\widehat{V}_i$ is an estimator of V_i. This estimator adapts the algorithm for the CQR estimator developed in Chernozhukov and Hong (2002) to deal with endogeneity. We call the multiplier $1(\widehat{S}_i' \widehat{\gamma} > \varsigma)$ the selector, as its purpose is to predict the subset of individuals for which the probability of censoring is sufficiently low to permit using a linear – in place of a censored linear – functional form for the conditional quantile. We formally state the conditions on the selector in the next subsection. The estimator in (2.9) may be seen as a computationally attractive approximation to Powell estimator applied to our case:

$$\widehat{\beta}_p(u) = \arg \min_{\beta \in \mathbb{R}^{\dim(X)}} \frac{1}{n} \sum_{i=1}^{n} \rho_u[Y_i - \max(\widehat{X}_i' \beta, C_i)].$$

The CQIV estimator will be computed using an iterative procedure where each step will take the form specified in equation (2.9). We start selecting the set of "quantile-uncensored" observations for which the conditional quantile function is above the censoring point. We implement this step by estimating the conditional probabilities of censoring using a flexible binary choice model. Quantile-uncensored observations have probability of censoring lower than the quantile index u. We estimate the linear part of the conditional quantile function, $X_i'\beta_0(u)$, on the sample of quantile-uncensored observations by standard quantile regression. Then, we update the set of quantile-uncensored observations by selecting those observations with conditional quantile estimates that are above their censoring points and iterate. We provide more practical implementation details in the next section.

The control variable V can be estimated in several ways. Note that if $Q_D(v \mid W, Z)$ is invertible in v, the control variable has several equivalent representations:

$$V = \vartheta_0(D, W, Z) \equiv F_D(D \mid W, Z) \equiv Q_D^{-1}(D \mid W, Z) \equiv \int_0^1 1\{Q_D(v \mid W, Z) \leq D\} dv. \tag{2.10}$$

For any estimator of $F_D(D \mid W, Z)$ or $Q_D(V \mid W, Z)$, denoted by $\widehat{F}_D(D \mid W, Z)$ or $\widehat{Q}_D(V \mid W, Z)$, based on any parametric or semi-parametric functional form, the resulting estimator for the control variable is

$$\widehat{V} = \widehat{\vartheta}(D, W, Z) \equiv \widehat{F}_D(D \mid W, Z) \text{ or } \widehat{V} = \widehat{\vartheta}(D, W, Z) \equiv \int_0^1 1\{\widehat{Q}_D(v \mid W, Z) \leq D\} dv.$$

Here we consider several examples: in the classical additive location model, we have that $Q_D(v \mid W, Z) = R'\pi_0 + Q_V(v)$, where Q_V is a quantile function, and $R = r(W, Z)$ is a vector collecting transformations of W and Z. The control variable is

$$V = Q_V^{-1}(D - R'\pi_0),$$

which can be estimated by the empirical CDF of the least squares residuals. Chernozhukov, Fernandez-Val and Melly (2009) developed asymptotic theory for this estimator. If $D \mid W, Z \sim N(R'\pi_0, \sigma^2)$, the control variable has the common parametric form $V = \Phi^{-1}([D - R'\pi_0]/\sigma)$, where Φ^{-1} denotes the quantile function of the standard normal distribution. This control variable can be estimated by plugging in estimates of the regression coefficients and residual variance.

In a non-additive quantile regression model, we have that $Q_D(v \mid W, Z) = R'\pi_0(v)$, and

$$V = Q_D^{-1}(D \mid W, Z) = \int_0^1 1\{R'\pi_0(v) \leq D\} dv.$$

The estimator takes the form

$$\widehat{V} = \int_0^1 1\{R'\widehat{\pi}(v) \leq D\} dv, \tag{2.11}$$

where $\widehat{\pi}(v)$ is the Koenker and Bassett (1978) quantile regression estimator and the integral can be approximated numerically using a finite grid of quantiles. The use of the integral to obtain a generalized inverse is convenient to avoid monotonicity problems in $v \mapsto R'\widehat{\pi}(v)$ due to misspecification or sampling error. Chernozhukov, Fernandez-Val, and Galichon (2010) developed asymptotic theory for this estimator.

We can also estimate ϑ_0 using distribution regression. In this case we consider a semi-parametric model for the conditional distribution of D to construct a control variable

$$V = F_D(D \mid W, Z) = \Lambda(R'\pi_0(D)),$$

where Λ is a probit or logit link function. The estimator takes the form

$$\widehat{V} = \Lambda(R'\widehat{\pi}(D)),$$

where $\widehat{\pi}(d)$ is the maximum likelihood estimator of $\pi_0(d)$ at each d (see, e.g., Foresi and Peracchi, 1995, and Chernozhukov, Fernandez-Val and Melly, 2009). Chernozhukov, Fernandez-Val and Melly (2009) developed asymptotic theory for this estimator.

2.3. **Regularity Conditions for Estimation.** In what follows, we shall use the following notation. We let the random vector $A = (Y, D, W, Z, C, X, V)$ live on some probability space $(\Omega_0, \mathcal{F}_0, P)$. Thus, the probability measure P determines the law of A or any of its elements. We also let $A_1, ..., A_n$, i.i.d. copies of A, live on the complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$, which contains the infinite product of $(\Omega_0, \mathcal{F}_0, P)$. Moreover, this probability space can be suitably enriched to carry also the random weights that will appear in the weighted bootstrap. The distinction between the two laws P and $\mathbb{P}$ is helpful to simplify the notation in the proofs and in the analysis. Calligraphic letter such as $\mathcal{Y}$ and $\mathcal{X}$ denote the support of Y and X; and $\mathcal{YX}$ denotes the joint support of (Y, X). Unless explicitly mentioned, all functions appearing in the statements are assumed to be measurable.

We now state formally the assumptions. The first assumption is our model.

Assumption 1 (Model)**.** *We have $\{Y_i, D_i, W_i, Z_i, C_i\}_{i=1}^n$, a sample of size n of independent and identically distributed observations from the random vector (Y, D, W, Z, C) which obeys the model assumptions stated in equations (2.7) - (2.10), i.e.*

$$Q_Y(u \mid D, W, Z, V, C) = Q_Y(u \mid X, C) = \max(X'\beta_0(u), C), \quad X = x(D, W, V),$$
$$V = \vartheta_0(D, W, Z) \equiv F_D(D \mid W, Z) \sim U(0, 1) \mid W, Z.$$

The second assumption imposes compactness and smoothness conditions. Compactness can be relaxed at the cost of more complicated and cumbersome proofs, while the smoothness conditions are fairly tight.

Assumption 2 (Compactness and smoothness)**.** *(a) The set $\mathcal{YDWZCX}$ is compact. (b) The endogenous regressor D has a continuous conditional density $f_D(\cdot \mid w, z)$ that is bounded above by a constant uniformly in $(w, z) \in \mathcal{WZ}$. (c) The random variable Y has a conditional density $f_Y(y \mid x, c)$ on (c, ∞) that is uniformly continuous in $y \in (c, \infty)$ uniformly in $(x, c) \in \mathcal{XC}$, and bounded above by a constant uniformly in $(x, c) \in \mathcal{XC}$. (d) The derivative vector $\partial_v x(d, w, v)$ exists and its components are uniformly continuous in $v \in [0, 1]$ uniformly in $(d, w) \in \mathcal{DW}$, and are bounded in absolute value by a constant, uniformly in $(d, w, v) \in \mathcal{DWV}$.*

The following assumption is a high-level condition on the function-valued estimator of the control variable. We assume that it has an asymptotic functional linear representation. Moreover, this functional estimator, while not necessarily living in a Donsker class, can be approximated by a random function that does live in a Donsker class. We will fully verify this condition for the case of quantile regression and distribution regression under more primitive conditions.

Assumption 3 (Estimator of the control variable)**.** *We have an estimator of the control variable of the form $\widehat{V} = \widehat{\vartheta}(D, W, Z)$, such that uniformly over $(d, w, z) \in \mathcal{DWZ}$, (a)*

$$\sqrt{n}(\widehat{\vartheta}(d, w, z) - \vartheta_0(d, w, z)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \ell(A_i, d, w, z) + o_{\mathbb{P}}(1), \quad \mathrm{E}_P[\ell(A, d, w, z)] = 0,$$

where $\mathrm{E}_P[\ell(A, D, W, Z)^2] < \infty$ and $\|\frac{1}{\sqrt{n}} \sum_{i=1}^n \ell(A_i, \cdot)\|_\infty = O_{\mathbb{P}}(1)$, and (b)

$$\|\widehat{\vartheta} - \widetilde{\vartheta}\|_\infty = o_{\mathbb{P}}(1/\sqrt{n}), \quad \text{for} \quad \widetilde{\vartheta} \in \Upsilon,$$

where the entropy of the function class Υ is not too high, namely

$$\log N(\epsilon, \Upsilon, \|\cdot\|_\infty) \lesssim 1/(\epsilon \log^2(1/\epsilon)), \text{ for all } 0 < \epsilon < 1.$$

The following assumptions are on the selector. The first part is a high-level condition on the estimator of the selector. The second part is a smoothness condition on the index that defines the selector. We shall verify that the CQIV estimator can act as a legitimate selector itself. Although the statement is involved, this condition can be easily satisfied as explained below.

Assumption 4 (Selector)**.** *(a) The selection rule has the form*

$$1[s(x(D, W, \widehat{V}), C)'\widehat{\gamma} > \varsigma],$$

for some $\varsigma > 0$, where $\widehat{\gamma} \to_{\mathbb{P}} \gamma_0$ and, for some $\epsilon' > 0$,

$$1[S'\gamma_0 > \varsigma/2] \leq 1[X'\beta_0(u) > C + \epsilon'] \leq 1[X'\beta_0(u) > C] \ P\text{-a.e.},$$

where $S = s(X, V)$ and $1[X'\beta_0(u) > C] \equiv 1[P(Y = C \mid Z, W, V) < u]$. (b) The set $\mathcal{S}$ is compact. (c) The density of the random variable $s(x(D, W, \vartheta(D, W, Z)), C)'\gamma$ exists and is bounded above by a constant, uniformly in $\gamma \in \Gamma$ and in $\vartheta \in \Upsilon$, where Γ is an open neighborhood of γ_0 and Υ is defined in Assumption 3. (d) The components of the derivative vector $\partial_v s(x(d, w, v), c)$ are uniformly continuous at each $v \in [0, 1]$, uniformly in $(d, w, c) \in \mathcal{DWC}$, and are bounded in absolute value by a constant, uniformly in $(d, w, v, c) \in \mathcal{DWVC}$.

The next assumption is a sufficient condition to guarantee local identification of the parameter of interest as well as $\sqrt{n}$-consistency and asymptotic normality of the estimator.

Assumption 5 (Identification and non-degeneracy)**.** *(a) The matrix*

$$J(u) := \mathrm{E}_P[f_Y(X'\beta_0(u) \mid X, C)XX' \ 1(S'\gamma_0 > \varsigma)]$$

is of full rank. (b) The matrix

$$\Lambda(u) := \mathrm{Var}_P[f(A) + g(A) \ ],$$

is finite and is of full rank, where

$$f(A) := \{1(Y < X'\beta_0(u)) - u\}X1(S'\gamma_0 > \varsigma),$$

and, for $\dot{X} = \partial_v x(D, W, v)|_{v=V}$,

$$g(A) := \mathrm{E}_P[f_Y(X'\beta_0(u) \mid X, C)X\dot{X}'\beta_0(u)1(S'\gamma_0 > \varsigma)\ell(a, D, W, Z)]\big|_{a=A}.$$

Assumption 4(a) requires the selector to find a subset of the quantile-censored observations, whereas Assumption 5 requires the selector to find a nonempty subset. Given $\widehat{\beta}^0(u)$, an initial consistent estimator of $\beta_0(u)$, we can form the selector as $1[s(x(D, W, \widehat{V}), C)'\widehat{\gamma} > \varsigma]$, where

$s(x(D,W,\widehat{V}),C) = [x(D,W,\widehat{V})', C]'$, $\widehat{\gamma} = [\widehat{\beta}^0(u)', -1]'$, and ς is a small fixed cut-off that ensures that the selector is asymptotically conservative but nontrivial. To find $\widehat{\beta}^0(u)$, we use a selector based on a flexible model for the probability of censoring. This model does not need to be correctly specified under a mild separating hyperplane condition for the quantile-uncensored observations (Chernozhukov and Hong, 2002). Alternatively, we can estimate a fully nonparametric model for the censoring probabilities. We do not pursue this approach to preserve the computational appeal of the CQIV estimator.

2.4. **Main Estimation Results.** The following result states that the CQIV estimator is consistent, converges to the true parameter at a $\sqrt{n}$ rate, and is normally distributed in large samples.

Theorem 1 (Asymptotic distribution of CQIV)**.** *Under the stated assumptions*

$$\sqrt{n}(\widehat{\beta}(u) - \beta_0(u)) \to_d N(0, J^{-1}(u)\Lambda(u)J^{-1}(u)).$$

We can estimate the variance-covariance matrix using standard methods and carry out analytical inference based on the normal distribution. Estimators for the components of the variance can be formed, e.g., following Powell (1991) and Koenker (2005). However, this is not very convenient for practice due to the complicated form of these components and the need to estimate conditional densities. Instead, we suggest using weighted bootstrap (Chamberlain and Imbens, 2003, Ma and Kosorok, 2005, Chen and Pouzo, 2009) and prove its validity in what follows.

We focus on weighted bootstrap because it has practical advantages over nonparametric bootstrap to deal with discrete regressors with small cell sizes and the proof of its consistency is not overly complex, following the strategy set forth by Ma and Kosorok (2005). Moreover, a particular version of the weighted bootstrap, with exponentials acting as weights, has a nice Bayesian interpretation (Chamberlain and Imbens, 2003).

To describe the weighted bootstrap procedure in our setting, we first introduce the "weights".

Assumption 6 (Bootstrap weights)**.** *The weights* $(e_1, ..., e_n)$ *are i.i.d. draws from a random variable* $e \geq 0$*, with* $\mathrm{E}_P[e] = 1$ *and* $\mathrm{Var}_P[e] = 1$*, living on the probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *and are independent of the data* $\{Y_i, D_i, W_i, Z_i, C_i\}_{i=1}^n$ *for all* n*.*

Remark 1 (Bootstrap weights). The chief and recommended example of bootstrap weights is given by e set to be the standard exponential random variable. Note that for other positive random variables with $\mathrm{E}_P[e] = 1$ but $\mathrm{Var}_P[e] > 1$, we can take the transformation $\tilde{e} = 1 + (e-1)/\mathrm{Var}_P[e]^{1/2}$, which satisfies $\tilde{e} \geq 0$, $\mathrm{E}_P[\tilde{e}] = 1$, and $\mathrm{Var}_P[\tilde{e}] = 1$.

The weights act as sampling weights in the bootstrap procedure. In each repetition, we draw a new set of weights $(e_1, \ldots, e_n)$ and recompute the CQIV estimator in the weighted sample. We refer to the next section for practical details, and here we define the quantities needed to verify the validity of this bootstrap scheme. Specifically, let $\widehat{V}_i^e$ denote the estimator of the control variable for observation i in the weighted sample, such as the quantile regression or distribution regression based estimators described in the next section. The CQIV estimator in the weighted sample solves

$$\widehat{\beta}^e(u) = \arg \min_{\beta \in \mathbb{R}^{\dim(X)}} \frac{1}{n} \sum_{i=1}^n e_i 1(\widehat{\gamma}' \widehat{S}_i^e > \varsigma) \rho_u(Y_i - \beta' \widehat{X}_i^e), \tag{2.12}$$

where $\widehat{X}_i^e = x(D_i, W_i, \widehat{V}_i^e)$, $\widehat{S}_i^e = s(\widehat{X}_i^e, C_i)$, and $\widehat{\gamma}$ is a consistent estimator of the selector. Note that we do not need to recompute $\widehat{\gamma}$ in the weighted samples, which is convenient for computation.

We make the following assumptions about the estimator of the control variable in the weighted sample.

Assumption 7 (Weighted estimator of control variable)**.** *Let* $(e_1, \ldots, e_n)$ *be a sequence of weights that satisfies Assumption 6. We have an estimator of the control variable of the form* $\widehat{V}^e = \widehat{\vartheta}^e(D, W, Z)$*, such that uniformly over* $\mathcal{DWZ}$*,*

$$\sqrt{n}(\widehat{\vartheta}^e(d, w, z) - \vartheta_0(d, w, z)) = \frac{1}{\sqrt{n}} \sum_{i=1}^n e_i \ell(A_i, d, w, z) + o_{\mathbb{P}}(1), \quad \mathrm{E}_P[\ell(A, d, w, z)] = 0,$$

where $\mathrm{E}_P[\ell(A, D, W, Z)^2] < \infty$ *and* $\| \frac{1}{\sqrt{n}} \sum_{i=1}^n e_i \ell(A_i, \cdot) \|_\infty = O_{\mathbb{P}}(1)$*, and*

$$\|\widehat{\vartheta}^e - \widetilde{\vartheta}^e\|_\infty = o_{\mathbb{P}}(1/\sqrt{n}), \quad \textit{for} \quad \widetilde{\vartheta}^e \in \Upsilon,$$

where the entropy of the function class Υ *is not too high, namely*

$$\log N(\epsilon, \Upsilon, \| \cdot \|_\infty) \lesssim 1/(\epsilon \log^2(1/\epsilon)), \ \textit{for all } 0 < \epsilon < 1.$$

Basically this is the same condition as Assumption 3 in the unweighted sample, and therefore both can be verified using analogous arguments. Note also that the condition is stated under the probability measure $\mathbb{P}$, i.e. unconditionally on the data, which actually simplifies verification. We give primitive conditions that verify this assumption for quantile and distribution regression estimation of the control variable in the next section.

The following result shows the consistency of weighted bootstrap to approximate the asymptotic distribution of the CQIV estimator.

Theorem 2 (Weighted-bootstrap validity for CQIV)**.** *Under the stated assumptions, conditionally on the data*

$$\sqrt{n}(\widehat{\beta}^e(u) - \widehat{\beta}(u)) \to_d N(0, J^{-1}(u)\Lambda(u)J^{-1}(u)),$$

in probability under $\mathbb{P}$*.*

Note that the statement above formally means that the distance between the law of $\sqrt{n}(\widehat{\beta}^e(u)-\widehat{\beta}(u))$ conditional on the data and the law of the normal vector $N(0, J^{-1}(u)\Lambda(u)J^{-1}(u))$, as measured by any metric that metrizes weak convergence, conveges in probability to zero. More specifically,

$$d_{BL}\{\mathcal{L}[\sqrt{n}(\widehat{\beta}^e(u) - \widehat{\beta}(u))|\text{data}], \mathcal{L}[N(0, J^{-1}(u)\Lambda(u)J^{-1}(u))]\} \to_{\mathbb{P}} 0,$$

where d_{BL} denotes the bounded Lipshitz metric.

In practice, we approximate numerically the distribution of $\sqrt{n}(\widehat{\beta}^e(u) - \widehat{\beta}(u))$ conditional on the data by simulation. For $b = 1, \dots, B$, we compute $\widehat{\beta}^e_b(u)$ solving the problem (2.12) with the data fixed and a set of weights $(e_{1b}, ..., e_{nb})$ randomly drawn for a distribution that satisfies Assumption 6. By Theorem 2, we can use the empirical distribution of $\sqrt{n}(\widehat{\beta}^e_b(u) - \widehat{\beta}(u))$ to make asymptotically valid inference on $\beta_0(u)$ for large B.

2.5. **Quantile and distribution regression estimation of the control variable.** One of the main contributions of this paper is to allow for quantile and distribution regression estimation of the control variable. The difficulties here are multifold, since the control variable depends on the infinite dimensional function $\pi_0(\cdot)$, and more importantly the estimated version of this function, $\widehat{\pi}(\cdot)$, does not seem to lie in any class with good entropic properties. We overcome these difficulties by demonstrating that the estimated function can be approximated with sufficient degree of accuracy by a random function that lies in a class with good entropic properties. To carry out this approximation, we smooth the empirical quantile regression and distribution regression processes by third order kernels, after suitably extending the processes to deal with boundary issues. Such kernels can be obtained by reproducing kernel Hilbert space methods or via twicing kernel methods (Berlinet, 1993, and Newey, Hsieh, and Robins, 2004). In the case of quantile regression, we also use results of the asymptotic theory for rearrangement-related operators developed by Chernozhukov, Fernández-Val and Galichon (2010). Moreover, all the previous arguments carry over weighted samples, which is relevant for the bootstrap.

2.5.1. *Quantile regression.* We impose the following condition:

Assumption 8 (QR control variable)**.** *(a) The conditional quantile function of D given (W, Z) follows the quantile regression model, i.e.,*

$$Q_D(\cdot \mid W, Z) = Q_D(\cdot \mid R) = R'\pi_0(\cdot), \quad R = r(W, Z),$$

where the coefficients $v \mapsto \pi_0(v)$ are three times continuously differentiable with uniformly bounded derivatives, and $\mathcal{R}$ is compact; (b) The conditional density $f_D(\cdot \mid R)$ is uniformly bounded by a constant P-a.e., and is continuous at $R'\pi_0(v)$ uniformly in $v \in (0, 1)$ P-a.e. (c) The Gram matrix $E[RR']$ is finite and has full rank.

For $\rho_v(z) := (v - 1(z<0))z$, let

$$\widehat{\pi}^e(v) \in \arg\min_{\pi \in \mathbb{R}^{\dim(R)}} \frac{1}{n}\sum_{i=1}^n e_i \rho_v(D_i - R_i'\pi),$$

where either $e_i = 1$ for the unweighted sample, to obtain the estimates; or e_i is drawn from a positive random variable with unit mean and variance for the weighted sample, to obtain bootstrap estimates. Then set

$$\vartheta_0(d,r) = \int_{(0,1)} 1\{r'\pi_0(v) \le d\} dv; \; \widehat{\vartheta}^e(d,r) = \int_{(0,1)} 1\{r'\widehat{\pi}^e(v) \le d\} dv.$$

The following result verifies that our main high-level conditions for the control variable estimator in Assumptions 3 and 7 hold under Assumption 8. The verification is done simultaneously for weighted and unweighted samples by including weights that can be equal to the trivial unit weights, as mentioned above.

Theorem 3 (Validity of Assumptions 3 & 7 for QR)**.** *Suppose that Assumption 8 holds. (1) We have that*

$$\begin{aligned}
\sqrt{n}(\widehat{\vartheta}^e(d,r) - \vartheta_0(d,r)) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n e_i \ell(A_i,d,r) + o_{\mathbb{P}}(1) \rightsquigarrow \Delta^e(d,r) \text{ in } \ell^\infty(\mathcal{DR}),\\
\ell(A,d,r) &:= f_D(d \mid r) r' \mathrm{E}_P\left[f_D(R'\pi_0(\vartheta_0(d,r)) \mid R)RR'\right]^{-1} \times \\
&\quad \times [1\{D \le R'\pi_0(\vartheta_0(d,r))\} - \vartheta_0(d,r)]R,\\
\mathrm{E}_P[\ell(A,d,r)] &= 0, \; \mathrm{E}_P[\ell(A,D,R)^2] < \infty,
\end{aligned}$$

where $\Delta^e(d,r)$ is a Gaussian process with continuous paths and covariance function given by $\mathrm{E}_P[\ell(A,d,r)\ell(A,\tilde{d},\tilde{r})']$. (2) Moreover, there exists $\widetilde{\vartheta}^e : \mathcal{DR} \mapsto [0,1]$ that obeys the same first order representation, is close to $\widehat{\vartheta}^e$ in the sense that $\|\widetilde{\vartheta}^e - \widehat{\vartheta}^e\|_\infty = o_{\mathbb{P}}(1/\sqrt{n})$, and, with probability approaching one, belongs to a bounded function class Υ such that

$$\log N(\epsilon, \Upsilon, \|\cdot\|_\infty) \lesssim \epsilon^{-1/2}, \quad 0 < \epsilon < 1.$$

Thus, Assumption 3 holds for the case $e_i = 1$, and Assumption 7 holds for the case of e_i being drawn from a positive random variable with unit mean and variance as in Assumption 6. Thus, the results of Theorem 1 and 2 apply for the QR estimator of the control variable.

2.5.2. *Distribution regression.* We impose the following condition:

Assumption 9 (DR control variable)**.** *(a) The conditional distribution function of D given (W,Z) follows the distribution regression model, i.e.,*

$$F_D(\cdot \mid W,Z) = F_D(\cdot \mid R) = \Lambda(R'\pi_0(\cdot)), \quad R = r(W,Z),$$

where Λ is either the probit or logit link function, the coefficients $d \mapsto \pi_0(d)$ are three times continuously differentiable with uniformly bounded derivatives; (b) $\mathcal{D}$ and $\mathcal{R}$ are compact; (c) The Gram matrix ERR' has full rank.

Let

$$\widehat{\pi}^e(d) \in \arg \min_{\pi \in \mathbb{R}^{\dim(R)}} \frac{1}{n} \sum_{i=1}^{n} e_i \{1(D_i \leq d) \log \Lambda(R_i'\pi) + 1(D_i > d) \log[1 - \Lambda(R_i'\pi)]\},$$

where either $e_i = 1$ for the unweighted sample, to obtain the estimates; or e_i is drawn from a positive random variable with unit mean and variance for the weighted sample, to obtain bootstrap estimates. Then set

$$\vartheta_0(d, r) = \Lambda(r'\pi_0(d)); \ \widehat{\vartheta}^e(d, r) = \Lambda(r'\widehat{\pi}^e(d)).$$

The following result verifies that our main high-level conditions for the control variable estimator in Assumptions 3 and 7 hold under Assumption 9. The verification is done simultaneously for weighted and unweighted samples by including weights that can be equal to the trivial unit weights.

Theorem 4 (Validity of Assumptions 3 & 7 for DR)**.** *Suppose that Assumption 9 holds. (1) We have that*

$$\begin{aligned}
\sqrt{n}(\widehat{\vartheta}^e(d,r) - \vartheta_0(d,r)) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} e_i \ell(A_i, d, r) + o_{\mathbb{P}}(1) \rightsquigarrow \Delta^e(d,r) \ \textit{in} \ \ell^\infty(\mathcal{DR}), \\
\ell(A, d, r) &:= \partial\Lambda(r'\pi_0(d)) r' \mathrm{E}_P \left[\frac{\partial\Lambda(R'\pi_0(d))^2}{\Lambda(R'\pi_0(d))[1 - \Lambda(R'\pi_0(d))]} RR' \right]^{-1} \times \\
&\quad \times \frac{1\{D \leq d\} - \Lambda(R'\pi_0(d))}{\Lambda(R'\pi_0(d))[1 - \Lambda(R'\pi_0(d))]} \partial\Lambda(R'\pi_0(d)) R, \\
\mathrm{E}_P[\ell(A, d, r)] &= 0, \mathrm{E}_P[\ell(A, D, R)^2] < \infty,
\end{aligned}$$

where $\Delta^e(d,r)$ is a Gaussian process with continuous paths and covariance function given by $\mathrm{E}_P[\ell(A,d,r)\ell(A,\tilde{d},\tilde{r})']$, and $\partial\Lambda$ is the derivative of Λ. (2) Moreover, there exists $\widetilde{\vartheta}^e : \mathcal{DR} \mapsto [0,1]$ that obeys the same first order representation, is close to $\widehat{\vartheta}^e$ in the sense that $\|\widetilde{\vartheta}^e - \widehat{\vartheta}^e\|_\infty = o_{\mathbb{P}}(1/\sqrt{n})$ and, with probability approaching one, belongs to a bounded function class Υ such that

$$\log N(\epsilon, \Upsilon, \|\cdot\|_\infty) \lesssim \epsilon^{-1/2}, \quad 0 < \epsilon < 1.$$

Thus, Assumption 3 holds for the case $e_i = 1$, and Assumption 7 holds for the case of e_i being drawn from a positive random variable with unit mean and variance as in Assumption 6. Thus, the results of Theorem 1 and 2 apply for the DR estimator of the control variable.

3. Computation and Numerical Examples

This section describes the numerical algorithms to compute the CQIV estimator and weighted bootstrap confidence intervals, and shows the results of a Monte Carlo numerical example.

3.1. **CQIV Algorithm.**

The algorithm to obtain CQIV estimates is similar to Chernozhukov and Hong (2002). We add an initial step to estimate the control variable V. We name this step as Step 0 to facilitate comparison with the Chernozhukov and Hong (2002) 3-Step CQR algorithm.

Algorithm 1 (CQIV)**.** *For each desired quantile u, perform the following steps:*

0. *Obtain an estimate of the control variable for each individual, $\widehat{V}_i$, and construct $\widehat{X}_i = x(D_i, W_i, \widehat{V}_i)$.*
1. *Select a subset of quantile-uncensored observations, J_0, whose conditional quantile function is likely to be above the censoring point, namely select a subset of $\{i : X_i'\beta_0(u) > C\}$. To find these observations, we note that $X'\beta_0(u) > C$ is equivalent to $P(Y > C \mid X, C) > 1 - u$. Hence we predict the quantile-uncensored observations using a flexible binary choice model:*
$$P(Y > C \mid X, C) = \Lambda(S_i'\delta_0), \quad S_i = s(X_i, C_i),$$
where Λ is a known link function, typically a probit or a logit. In estimation, we replace S_i by $\widehat{S}_i = s(\widehat{X}_i, C_i)$. Then, we select the sample J_0 according to the following criterion:
$$J_0 = \{i : \Lambda(\widehat{S}_i'\widehat{\delta}) > 1 - u + k_0\}.$$
2. *Estimate a standard quantile regression on the subsample defined by J_0:*
$$\widehat{\beta}^0(u) = \arg \min_{\beta \in \mathbb{R}^{\dim(X)}} \sum_{i \in J_0} \rho_u(Y_i - \widehat{X}_i'\beta). \tag{3.1}$$
Next, using the predicted values, select another subset of quantile-uncensored observations, J_1, from the full sample according to the following criterion:
$$J_1 = \{i : \widehat{X}_i'\widehat{\beta}^0(u) > C_i + \varsigma_1\}. \tag{3.2}$$
3. *Estimate a standard quantile regression on the subsample defined by J_1. Formally, replace J_0 by J_1 in (3.1). The new estimates, $\widehat{\beta}^1(u)$, are the 3-Step CQIV coefficient estimates.*
4. *(Optional) With the results from the previous step, select a new sample J_2 replacing $\widehat{\beta}^0(u)$ by $\widehat{\beta}^1(u)$ in (3.2). Iterate this and the previous step a bounded number of times.*

Remark 2 (Step 0). A simple additive strategy is to estimate the control variable using the empirical CDF of the residuals from the first stage OLS regression of D on W and Z. More flexible non-additive strategies based on quantile regression or distribution regression are described in the previous section.

Remark 3 (Step 1). To predict the quantile-uncensored observations, a probit, logit, or any other model that fits the data well can be used. Note that the model does not need to be correctly specified; it suffices that it selects a nontrivial subset of observations with $X_i'\beta_0(u) > C_i$. To choose the value of k_0, it is advisable that a constant fraction of observations satisfying $\Lambda(\widehat{S}_i'\widehat{\delta}) > 1-u$ are excluded from J_0 for each quantile. To do so, set k_0 as the q_0th quantile of $\Lambda(\widehat{S}_i'\widehat{\delta})$ conditional on $\Lambda(\widehat{S}_i'\widehat{\delta}) > 1-u$, where q_0 is a percentage (10% worked well in our simulation). The empirical value of k_0 and the percentage of observations retained in J_0 can be computed as simple robustness diagnostic tests at each quantile.

Remark 4 (Step 2). To choose the cut-off ς_1, it is advisable that a constant fraction of observations satisfying $\widehat{X}_i'\widehat{\beta}^0(u) > C_i$ are excluded from J_1 for each quantile. To do so, set ς_1 to be the q_1th quantile of $\widehat{X}_i'\widehat{\beta}^0(u) - C_i$ conditional on $\widehat{X}_i'\widehat{\beta}^0(u) > C_i$, where q_1 is a percentage less than q_0 (3% worked well in our simulation). In practice, it is desirable that $J_0 \subset J_1$. If this is not the case, we recommend altering q_0, q_1, or the specification of the regression models. At each quantile, the empirical value of ς_1, the percentage of observations from the full sample retained in J_1, the percentage of observations from J_0 retained in J_1, and the number of observations in J_1 but not in J_0 can be computed as simple robustness diagnostic tests. The estimator $\widehat{\beta}^0(u)$ is consistent but will be inefficient relative to the estimator obtained in the subsequent step.

Remark 5 (Steps 1 and 2). In the notation of Assumption 4, the selector of Step 1 can be expressed as $1(\widehat{S}_i'\widehat{\gamma} > \varsigma_0)$, where $\widehat{S}_i'\widehat{\gamma} = \widehat{S}_i'\widehat{\delta} - \Lambda^{-1}(1-u)$ and $\varsigma_0 = \Lambda^{-1}(1-u+k_0) - \Lambda^{-1}(1-u)$. The selector of Step 2 can also be expressed as $1(\widehat{S}_i'\widehat{\gamma} > \varsigma_1)$, where $\widehat{S}_i = (\widehat{X}_i', C_i)'$ and $\widehat{\gamma} = (\widehat{\beta}^0(u)', -1)'$.

Remark 6 (Steps 2, 3 and 4). Beginning with Step 2, each successive iteration of the algorithm should yield estimates that come closer to minimizing the Powell objective function. As a simple robustness diagnostic test, we recommend computing the Powell objective function using the full sample and the estimated coefficients after each iteration, starting with Step 2. This diagnostic test is computationally straightforward because computing the objective function for a given set of values is much simpler than maximizing it. In practice, this test can be used to determine when to stop the CQIV algorithm for each quantile. If the Powell objective function increases from Step s to Step $s+1$ for $s \geq 2$, estimates from Step s can be retained as the coefficient estimates.

3.2. **Weighted Bootstrap Algorithm.** We recommend obtaining confidence intervals through a weighted bootstrap procedure, though analytical formulas can also be used. If the estimation runs quickly on the desired sample, it is straightforward to rerun the entire CQIV algorithm B times weighting all the steps by the bootstrap weights. To speed up the computation, we propose a procedure that uses a one-step CQIV estimator in each bootstrap repetition.

Algorithm 2 (Weighted bootstrap CQIV)**.** *For $b = 1, \ldots, B$, repeat the following steps:*

1. *Draw a set of weights $(e_{1b}, \ldots, e_{nb})$ i.i.d from a random variable e that satisfies Assumption 6. For example, we can draw the weights from a standard exponential distribution.*
2. *Reestimate the control variable in the weighted sample, $\widehat{V}_{ib}^e = \widehat{\vartheta}_b^e(D_i, W_i, Z_i)$, and construct $\widehat{X}_{ib}^e = x(D_i, W_i, \widehat{V}_{ib}^e)$.*
3. *Estimate the weighted quantile regression:*
$$\widehat{\beta}_b^e(u) = \arg \min_{\beta \in \mathbb{R}^{\dim(X)}} \sum_{i \in J_{1b}} e_{ib} \rho_u(Y_i - \beta' \widehat{X}_{ib}^e),$$
where $J_{1b} = \{i : \widehat{\beta}(u)' \widehat{X}_{ib}^e > C_i + \varsigma_1\}$, and $\widehat{\beta}(u)$ is a consistent estimator of $\beta_0(u)$, e.g., the 3-stage CQIV estimator $\widehat{\beta}^1(u)$.

Remark 7 (Step 2). The estimate of the control function $\widehat{\vartheta}_b^e$ can be obtained by weighted least squares, weighted quantile regression, or weighted distribution regression.

Remark 8 (Step 3). A computationally less expensive alternative is to set $J_{1b} = J_1$ in all the repetitions, where J_1 is the subset of selected observations in Step 2 of the CQIV algorithm.

We can construct an asymptotic $(1-\alpha)$-confidence interval for a function of the parameter vector $g(\beta_0(u))$ as $[\widehat{g}_{\alpha/2}, \widehat{g}_{1-\alpha/2}]$, where $\widehat{g}_\alpha$ is the sample α-quantile of $[g(\widehat{\beta}_1^e(u)), \ldots, g(\widehat{\beta}_B^e(u))]$. For example, the 0.025 and 0.975 quantiles of $(\widehat{\beta}_{1,k}^e(u), \ldots, \widehat{\beta}_{B,k}^e(u))$ form a 95% asymptotic confidence interval for the kth coefficient $\beta_{0,k}(u)$.

3.3. **Monte-Carlo illustration.** The goal of the following numerical example is to compare the performance of CQIV relative to tobit IV and other quantile regression estimators in finite samples. We generate data according to a normal design that satisfies the tobit parametric assumptions and a design with heteroskedasticity in the first stage equation for the endogenous regressor D that does not satisfy the tobit parametric assumptions. To facilitate the comparison, in both designs we consider a location model for the response variable Y^*, where the coefficients of the conditional expectation function and the conditional quantile function are equal (other than the intercept), so that tobit and CQIV estimate the same parameters. A comparison of the dispersion of the tobit estimates to the dispersion of the CQIV estimates at each quantile in the first design serves to quantify the relative efficiency of CQIV in a case where tobit IV can be expected to perform as well as possible. The appendix provides a more detailed description of the designs.

We consider two tobit estimators for comparison. Tobit-iv is the full information maximum likelihood estimator developed by Newey (1987), which is implemented in Stata with the command `ivtobit`. Tobit-cmle is the conditional maximum likelihood tobit estimator developed by Smith and Blundell (1986), which uses least squares residuals as a control variable. For additional comparisons, we present results from the censored quantile regression (cqr) estimator of Chernozhukov and Hong (2002), which does not address endogeneity; the quantile instrumental variables estimator (qiv-ols) of Lee (2007) with parametric first and second stage, which does not account for censoring; and the quantile regression (qr) estimator of Koenker and Bassett (1978), which does not account for endogeneity nor censoring. For CQIV we consider three different methods to estimate the control variable: cqiv-ols, which uses least squares; cqiv-qr, which uses quantile regression; and cqiv-dr, which uses probit distribution regression. The appendix also provides technical details for all CQIV estimators, as well as diagnostic test results for the cqiv-ols estimator.

We focus on the coefficient on the endogenous regressor D. We report mean bias and root mean square error (rmse) for all the estimators at the $\{.05, .10, ..., .95\}$ quantiles. For the homoskedastic design, the bias results are reported in the upper panel of Figure 1 and the rmse results are reported in the lower panel. In this figure, we see that tobit-cmle represents a substantial improvement over tobit-iv in terms of mean bias and rmse. Even though tobit-iv is theoretically efficient in this design, the CQIV estimators out-perform tobit-iv, and compare well to tobit-cmle. The figure also demonstrates that the CQIV estimators out-perform the other quantile estimators at all estimated quantiles. All of our qualitative findings hold when we consider unreported alternative measures of bias and dispersion such as median bias, interquartile range, and standard deviation.

The similar performance of tobit-cmle and cqiv can be explained by the homoskedasticity in the first stage of the design. Figure 2 reports mean bias and rmse results for the heteroskedastic design. Here cqiv-qr outperforms cqiv-ols and cqiv-dr at every quantile, which is expected because cqiv-ols and cqiv-dr are both misspecified for the control variable. Cqiv-dr has lower bias than cqiv-ols because it uses a more flexible specification for the control variable. Cqiv-qr also outperforms all other quantile estimators. Most importantly, at every quantile, cqiv-qr outperforms both tobit estimators, which are no longer consistent given the heteroskedasticity in the design of the first stage. In summary, CQIV performs well relative to tobit in a model that satisfies the parametric assumptions required for tobit-iv to be efficient, and it outperforms tobit in a model with heteroskedasticy.

4. Empirical Application: Engel Curve Estimation

In this section, we apply the CQIV estimator to the estimation of Engel curves. The Engel curve relationship describes how a household's demand for a commodity changes as the household's expenditure increases. Lewbel (2006) provides a recent survey of the extensive literature on Engel curve estimation. For comparability to the recent studies, we use data from the 1995 U.K. Family Expenditure Survey (FES) as in Blundell, Chen, and Kristensen (2007) and Imbens and Newey (2009). Following Blundell, Chen, and Kristensen (2007), we restrict the sample to 1,655 married or cohabitating couples with two or fewer children, in which the head of household is employed and between the ages of 20 and 55. The FES collects data on household expenditure for different categories of commodities. We focus on estimation of the Engel curve relationship for the alcohol category because 16% of families in our data report zero expenditure on alcohol. Although zero expenditure on alcohol arises as a corner solution outcome, and not from bottom coding, both types of censoring motivate the use of censored estimators such as CQIV.

Endogeneity in the estimation of Engel curves arises because the decision to consume a particular category of commodity may occur simultaneously with the allocation of income between consumption and savings. Following the literature, we rely on a two-stage budgeting argument to justify the use of labor income as an instrument for expenditure. Specifically, we estimate a quantile regression model in the first stage, where the logarithm of total expenditure, D, is a function of the logarithm of gross earnings of the head of the household, Z, and demographic household characteristics, W. The control variable, V, is obtained using the CQIV-QR estimator in (2.11), where the integral is approximated by a grid of 100 quantiles. For comparison, we also obtained control variable estimates using least squares and probit distribution regression. We do not report these comparison estimates because the correlation between the different control variable estimates was virtually 1, and all the methods resulted in very similar estimates in the second stage.

In the second stage we focus on the following quantile specification for Engel curve estimation:

$$Y_i = \max(X_i'\beta_0(U_i), 0), \;\; X_i = (1, D_i, D_i^2, W_i, \Phi^{-1}(V_i)), \;\; U_i \backsim U(0,1) \mid X_i,$$

where Y is the observed share of total expenditure on alcohol censored at zero, W is a binary household demographic variable that indicates whether the family has any children, and V is the control variable. We define our binary demographic variable following Blundell, Chen and Kristensen (2007).[2]

[2]Demographic variables are important shifters of Engel curves. In recent literature, "shape invariant" specifications for demographic variable have become popular. For comparison with this literature, we also estimate

To choose the specification, we rely on recent studies in Engel curve estimation. Thus, following Blundell, Browning, and Crawford (2003) we impose separability between the control variable and other regressors. Hausman, Newey, and Powell (1995) and Banks, Blundell, and Lewbel (1997) show that the quadratic specification in log-expenditure gives a better fit than the linear specification used in earlier studies. In particular, Blundell, Duncan, and Pendakur (1998) find that the quadratic specification gives a good approximation to the shape of the Engel curve for alcohol. To check the robustness of the specification to the linearity in the control variable, we also estimate specifications that include nonlinear terms in the control variable. The results are very similar to the ones reported.

Figure 3 reports the estimated coefficients $u \mapsto \widehat{\beta}(u)$ for a variety of estimators. In addition to reporting results for CQIV with a quantile estimate of the control variable (cqiv), as in the previous numerical examples, we report estimates from the censored quantile regression (cqr) of Chernozhukov and Hong (2002), the quantile instrumental variables estimator with a quantile regression estimate of the control variable (qiv) of Lee (2007), and the quantile regression (qr) estimator of Koenker and Bassett (1978). We also estimate a model for the conditional mean with the tobit-cmle of Smith and Blundell (1986) that incorporates a least squares estimate of the control variable. The tobit-iv algorithm implemented in Stata does not converge in this application. Given the level of censoring, we focus on conditional quantiles above the .15 quantile.

In the panels that depict the coefficients of expenditure and its square, the importance of controlling for censoring is especially apparent. Comparison between the censored quantile estimators (cqiv and cqr), plotted with thick light lines, and the uncensored quantile estimators (qiv and qr), plotted with thin dark lines, demonstrates that the censoring attenuates the uncorrected estimates toward zero at most quantiles in this application. In particular, censoring appears very important at the lowest quantiles. Relative to the tobit-cmle estimate of the conditional mean, cqiv provides a richer picture of the heterogenous effects of the variables. Comparison of the quantile estimators that account for endogeneity (cqiv and qiv), plotted with solid lines, and those that do not (cqr and qr), plotted with dashed lines, shows that endogeneity also influences the estimates, but the pattern is more difficult to interpret. The estimates of the coefficient of the control variable indicate that the endogeneity problem is more severe in the upper half of the distribution. This is consistent with a situation where a strong preference to consume alcohol raises total household expenditure.

Our quadratic quantile model is flexible in that it permits the expenditure elasticities to vary across quantiles of the alcohol share and across the level of total expenditure. These

an unrestricted version of shape invariant specification in which we include a term for the interaction between the logarithm of expenditure and our demographic variable. The results from the shape invariant specification are qualitatively similar but less precise than the ones reported in this application.

quantile elasticities are related to the coefficients of the model by

$$\partial_d Q_Y(u \mid x) = 1\{x'\beta_0(u) > 0\}\{\beta_{01}(u) + 2\beta_{02}(u)\ d\},$$

where $\beta_{01}(u)$ and $\beta_{02}(u)$ are the coefficients of D and D^2, respectively. Figure 4 reports point and interval estimates of average quantile elasticities as a function of the quantile index u, i.e., $u \mapsto \mathrm{E}_P[\partial_d Q_Y(u \mid X)]$. Here we see that accounting for endogeneity and censoring also has important consequences for these economically relevant quantities. The difference between the estimates is more pronounced along the endogeneity dimension than it is along the censoring dimension. The right panel plots 95% pointwise confidence intervals for the cqiv quantile elasticity estimates obtained by the weighted bootstrap method described in Section 3 with standard exponential weights and $B = 200$ repetitions. Here we can see that there is significant heterogeneity in the expenditure elasticity across quantiles. Thus, alcohol passes from being a normal good for low quantiles to being an inferior good for high quantiles. This heterogeneity is missed by conventional mean estimates of the elasticity.

In Figure 5 we report families of Engel curves based on the cqiv coefficient estimates. We predict the value of the alcohol share, Y, for a grid of values of log expenditure using the cqiv coefficients at each quartile. The subfigures depict the Engel curves for each quartile of the empirical values of the control variable, for individuals with and without kids, that is

$$d \mapsto \max\{(1, d, d^2, w, \Phi^{-1}(v))'\widehat{\beta}(u), 0\}$$

for $(w, \Phi^{-1}(v), u)$ evaluated at $w \in \{0, 1\}$, the quartiles of $\widehat{V}$ for v, and $u \in \{0.25, 0.50, 0.75\}$. Here we can see that controlling for censoring has an important effect on the shape of the Engel curves even at the median. The families of Engel curves are fairly robust to the values of the control variable, but the effect of children on alcohol shares is more pronounced. The presence of children in the household produces a downward shift in the Engel curves at all the levels of log-expenditure considered.

5. Conclusion

In this paper, we develop a new censored quantile instrumental variable estimator that incorporates endogenous regressors using a control variable approach. Censoring and endogeneity abound in empirical work, making the new estimator a valuable addition to the applied econometrician's toolkit. For example, Kowalski (2009) uses this estimator to analyze the price elasticity of expenditure on medical care across the quantiles of the expenditure distribution, where censoring arises because of the decision to consume zero care and endogeneity arises because marginal prices explicitly depend on expenditure. Since the new estimator can be implemented using standard statistical software, it should prove useful to applied researchers in many applications.

Appendix A. Notation

In what follows ϑ and γ denote generic values for the control function and the parameter of the selector $1(S_i'\gamma > \varsigma)$. It is convenient also to introduce some additional notation, which will be extensively used in the proofs. Let $X_i(\vartheta) := x(D_i, W_i, \vartheta(D_i, W_i, Z_i))$, $S_i(\vartheta) := s(X_i(\vartheta), C_i)$, $\dot{X}_i(\vartheta) := \partial_v x(D_i, W_i, v)|_{v=\vartheta(D_i,W_i,Z_i)}$, and $\dot{S}_i(\vartheta) := \partial_v s(X_i(v), C_i)|_{v=\vartheta(D_i,W_i,Z_i)}$. When the previous functions are evaluated at the true values we use $X_i = X_i(\vartheta_0)$, $S_i = S_i(\vartheta_0)$, $\dot{X}_i = \dot{X}_i(\vartheta_0)$, and $\dot{S}_i = \dot{S}_i(\vartheta_0)$. Also, let $\varphi_u(z) := [1(z < 0) - u]$. Recall that $A := (Y, D, W, Z, C, X, V)$. For a function $f : \mathcal{A} \mapsto \mathbb{R}$, we use $\|f\|_\infty = \sup_{a\in\mathcal{A}} |f(a)|$; for a K-vector of functions $f : \mathcal{A} \mapsto \mathbb{R}^K$, we use $\|f\|_\infty = \sup_{a\in\mathcal{A}} \|f(a)\|_2$. We make functions in Υ as well as estimates $\widehat{\vartheta}$ to take values in $[0, 1]$, the support of the control variable V. This allows us to simplify notation in what follows. We also adopt the standard notation in the empirical process literature (see, e.g., van der Vaart, 1998),

$$\mathbb{E}_n[f] = \mathbb{E}_n[f(A)] = n^{-1} \sum_{i=1}^n f(A_i),$$

and

$$\mathbb{G}_n[f] = \mathbb{G}_n[f(A)] = n^{-1/2} \sum_{i=1}^n (f(A_i) - \mathrm{E}_P[f(A)]).$$

When the function $\widehat{f}$ is estimated, the notation should interpreted as:

$$\mathbb{G}_n[\widehat{f}\,] = \mathbb{G}_n[f]\,|_{f=\widehat{f}}$$

Appendix B. Proof of Theorems 1 and 2

B.1. **Proof of Theorem 1.** Step 1. This step shows that $\sqrt{n}(\widehat{\beta}(u) - \beta_0(u)) = O_{\mathbb{P}}(1)$.

By Assumption 4 on the selector, for large enough n:

$$1\{S(\widehat{\vartheta})'\widehat{\gamma} \geq \varsigma\} \leq 1\{S'\gamma_0 > \varsigma - \epsilon_n\} \leq 1\{S'\gamma_0 > \varsigma/2\} \leq 1\{X'\beta_0(u) > C + \epsilon'\} =: \chi,$$

P-a.e., since

$$|S(\widehat{\vartheta})'\widehat{\gamma} - S'\gamma_0| \leq \epsilon_n := L_S(\|\widehat{\vartheta} - \vartheta_0\|_\infty + \|\widehat{\gamma} - \gamma_0\|_2) \to_{\mathbb{P}} 0,$$

where $L_S := (\|\partial_v s\|_\infty \vee \|s\|_\infty)$ is a finite constant by assumption.

Hence, with probability approaching one,

$$\widehat{\beta}(u) = \arg \min_{\beta\in\mathbb{R}^{\dim(X)}} \mathbb{E}_n[\rho_u(Y - X(\widehat{\vartheta})'\beta)1(S(\widehat{\vartheta})'\widehat{\gamma} > \varsigma)\chi],$$

Due to convexity of the objective function, it suffices to show that for any $\epsilon > 0$ there exists a finite positive constant B_ϵ such that

$$\liminf_{n\to\infty} \mathbb{P}\left(\inf_{\|\eta\|_2=1} \sqrt{n}\eta'\mathbb{E}_n\left[\widehat{f}_{\eta,B_\epsilon}\right] > 0\right) \geq 1-\epsilon, \tag{B.1}$$

where

$$\widehat{f}_{\eta,B_\epsilon}(A) := \varphi_u\left\{Y - X(\widehat{\vartheta})'(\beta_0(u) + B_\epsilon\eta/\sqrt{n})\right\} X(\widehat{\vartheta})1\{S(\widehat{\vartheta})'\widehat{\gamma} > \varsigma\}\chi.$$

Let

$$f(A) := \varphi_u\left\{Y - X'\beta_0(u)\right\} X1\{S'\gamma_0 > \varsigma\}.$$

Then uniformly in $\|\eta\|_2 = 1$,

$$\begin{aligned}
\sqrt{n}\eta'\mathbb{E}_n[\widehat{f}_{\eta,B_\epsilon}] &= \eta'\mathbb{G}_n[\widehat{f}_{\eta,B_\epsilon}] + \sqrt{n}\eta'\mathrm{E}_P[\widehat{f}_{\eta,B_\epsilon}] \\
&=_{(1)} \eta'\mathbb{G}_n[f] + o_{\mathbb{P}}(1) + \eta'\sqrt{n}\mathrm{E}_P[\widehat{f}_{\eta,B_\epsilon}] \\
&=_{(2)} \eta'\mathbb{G}_n[f] + o_{\mathbb{P}}(1) + \eta'J(u)\eta B_\epsilon + \eta'\mathbb{G}_n[g] + o_{\mathbb{P}}(1) \\
&=_{(3)} O_{\mathbb{P}}(1) + o_{\mathbb{P}}(1) + \eta'J(u)\eta B_\epsilon + O_{\mathbb{P}}(1) + o_{\mathbb{P}}(1),
\end{aligned}$$

where relations (1) and (2) follow by Lemma 1 and Lemma 2 with $\widetilde{\beta} = \beta_0(u) + B_\epsilon\eta/\sqrt{n}$, respectively, using that $\|\widehat{\vartheta} - \widetilde{\vartheta}\|_\infty = o_{\mathbb{P}}(1/\sqrt{n})$, $\widetilde{\vartheta} \in \Upsilon$, $\|\widetilde{\vartheta} - \vartheta_0\|_\infty = O_{\mathbb{P}}(1/\sqrt{n})$ and $\|\beta_0(u) + B_\epsilon\eta/\sqrt{n} - \beta_0(u)\|_2 = O(1/\sqrt{n})$; relation (3) holds by Chebyshev inequality. Since $J(u)$ is positive definite, with minimal eigenvalue bounded away from zero, the inequality (B.1) follows by choosing B_ϵ as a sufficiently large constant.

Step 2. In this step we show the main result. From the subgradient characterization of the solution to the quantile regression problem we have

$$\sqrt{n}\mathbb{E}_n\left[\widehat{f}\right] = \delta_n; \quad \|\delta_n\|_2 \leq \dim(X) \max_{1\leq i\leq n} \|X_i\|_2/\sqrt{n} = o_{\mathbb{P}}(1), \tag{B.2}$$

where

$$\widehat{f}(A) := \varphi_u\left\{Y - X(\widehat{\vartheta})'\widehat{\beta}(u)\right\} X(\widehat{\vartheta})1\{S(\widehat{\vartheta})'\widehat{\gamma} > \varsigma\}\chi.$$

Therefore

$$\begin{aligned}
o_{\mathbb{P}}(1) = \sqrt{n}\mathbb{E}_n\left[\widehat{f}\right] &= \mathbb{G}_n\left[\widehat{f}\right] + \sqrt{n}\mathrm{E}_P\left[\widehat{f}\right] \\
&=_{(1)} \mathbb{G}_n[f] + o_{\mathbb{P}}(1) + \sqrt{n}\mathrm{E}_P\left[\widehat{f}\right] \\
&=_{(2)} \mathbb{G}_n[f] + o_{\mathbb{P}}(1) + J(u)\sqrt{n}(\widehat{\beta}(u) - \beta_0(u)) + \mathbb{G}_n[g] + o_{\mathbb{P}}(1),
\end{aligned}$$

where relations (1) and (2) follow by Lemma 1 and Lemma 2 with $\widetilde{\beta} = \widehat{\beta}(u)$, respectively, using that $\|\widehat{\vartheta} - \widetilde{\vartheta}\|_\infty = o_{\mathbb{P}}(1/\sqrt{n})$, $\widetilde{\vartheta} \in \Upsilon$, $\|\widetilde{\vartheta} - \vartheta\|_\infty = O_{\mathbb{P}}(1/\sqrt{n})$ and $\|\widehat{\beta}(u) - \beta_0(u)\|_2 = O_{\mathbb{P}}(1/\sqrt{n})$.

Therefore by invertibility of $J(u)$,

$$\sqrt{n}(\widehat{\beta}(u) - \beta_0(u)) = -J(u)^{-1}\mathbb{G}_n(f+g) + o_{\mathbb{P}}(1).$$

By the Central Limit Theorem, $\mathbb{G}_n(f+g) \to_d N(0, \mathrm{Var}_P(f+g))$, so that

$$\sqrt{n}(\widehat{\beta}(u) - \beta_0(u)) \to_d N(0, J(u)^{-1}\mathrm{Var}_P(f+g)J(u)^{-1}).$$

□

B.2. **Proof of Theorem 2.** Step 1. This step shows that $\sqrt{n}(\widehat{\beta}^e(u) - \beta_0(u)) = O_{\mathbb{P}}(1)$ under the unconditional probability $\mathbb{P}$.

By Assumption 4 on the selector, with probability approaching one,

$$\widehat{\beta}^e(u) = \arg \min_{\beta \in \mathbb{R}^{\dim(X)}} \mathbb{E}_n[e\rho_u(Y - X(\widehat{\vartheta}^e)'\beta)1(S(\widehat{\vartheta}^e)'\widehat{\gamma} > \varsigma)\chi],$$

where e is the random variable used in the weighted bootstrap, and $\chi = 1(X'\beta_0(u) > C + \epsilon')$. Due to convexity of the objective function, it suffices to show that for any $\epsilon > 0$ there exists a finite positive constant B_ϵ such that

$$\liminf_{n\to\infty} \mathbb{P}\left(\inf_{\|\eta\|_2 = 1} \sqrt{n}\eta'\mathbb{E}_n\left[\widehat{f}^e_{\eta,B_\epsilon}\right] > 0 \right) \geq 1 - \epsilon, \tag{B.3}$$

where

$$\widehat{f}^e_{\eta,B_\epsilon}(A) := e \cdot \varphi_u\left\{Y - X(\widehat{\vartheta}^e)'(\beta_0(u) + B_\epsilon\eta/\sqrt{n})\right\} X(\widehat{\vartheta}^e)1\{S(\widehat{\vartheta}^e)'\widehat{\gamma} > \varsigma\}\chi.$$

Let

$$f^e(A) := e \cdot \varphi_u\left\{Y - X'\beta_0(u)\right\} X1\{S'\gamma_0 > \varsigma\}.$$

Then uniformly in $\|\eta\|_2 = 1$,

$$\begin{aligned}
\sqrt{n}\eta'\mathbb{E}_n[\widehat{f}^e_{\eta,B_\epsilon}] &= \eta'\mathbb{G}_n[\widehat{f}^e_{\eta,B_\epsilon}] + \sqrt{n}\eta'\mathrm{E}_P[\widehat{f}^e_{\eta,B_\epsilon}] \\
&=_{(1)} \eta'\mathbb{G}_n[f^e] + o_{\mathbb{P}}(1) + \eta'\sqrt{n}\mathrm{E}_P[\widehat{f}^e_{\eta,B_\epsilon}] \\
&=_{(2)} \eta'\mathbb{G}_n[f^e] + o_{\mathbb{P}}(1) + \eta' J(u)\eta B_\epsilon + \eta'\mathbb{G}_n[g^e] + o_{\mathbb{P}}(1) \\
&=_{(3)} O_{\mathbb{P}}(1) + o_{\mathbb{P}}(1) + \eta' J(u)\eta B_\epsilon + O_{\mathbb{P}}(1) + o_{\mathbb{P}}(1),
\end{aligned}$$

where relations (1) and (2) follow by Lemma 1 and Lemma 2 with $\widetilde{\beta} = \beta_0(u) + B_\epsilon\eta/\sqrt{n}$, respectively, using that $\|\widehat{\vartheta}^e - \widetilde{\vartheta}^e\|_\infty = o_{\mathbb{P}}(1/\sqrt{n})$, $\widetilde{\vartheta}^e \in \Upsilon$, $\|\widetilde{\vartheta}^e - \vartheta_0\|_\infty = O_{\mathbb{P}}(1/\sqrt{n})$ and $\|\beta_0(u) + B_\epsilon\eta/\sqrt{n} - \beta_0(u)\|_2 = O(1/\sqrt{n})$; relation (3) holds by Chebyshev inequality. Since $J(u)$ is positive definite, with minimal eigenvalue bounded away from zero, the inequality (B.3) follows by choosing B_ϵ as a sufficiently large constant.

Step 2. In this step we show that $\sqrt{n}(\widehat{\beta}^e(u) - \beta_0(u)) = -J(u)^{-1}\mathbb{G}_n(f^e + g^e) + o_{\mathbb{P}}(1)$ under the unconditional probability $\mathbb{P}$.

From the subgradient characterization of the solution to the quantile regression problem we have

$$\sqrt{n}\mathbb{E}_n\left[\widehat{f}^e\right] = \delta_n^e; \quad \|\delta_n^e\|_2 \leq \dim(X) \max_{1\leq i\leq n} \|e_i X_i\|_2/\sqrt{n} = o_{\mathbb{P}}(1), \tag{B.4}$$

where

$$\widehat{f}^e(A) := e \cdot \varphi_u\left\{Y - X(\widehat{\vartheta}^e)'\widehat{\beta}^e(u)\right\} X(\widehat{\vartheta}^e)1\{S(\widehat{\vartheta}^e)'\widehat{\gamma} > \varsigma\}\chi.$$

Therefore

$$\begin{aligned} o_{\mathbb{P}}(1) = \sqrt{n}\mathbb{E}_n\left[\widehat{f}^e\right] &= \mathbb{G}_n\left[\widehat{f}^e\right] + \sqrt{n}\mathrm{E}_P\left[\widehat{f}^e\right] \\ &=_{(1)} \mathbb{G}_n[f^e] + o_{\mathbb{P}}(1) + \sqrt{n}\mathrm{E}_P\left[\widehat{f}^e\right] \\ &=_{(2)} \mathbb{G}_n[f^e] + o_{\mathbb{P}}(1) + J(u)\sqrt{n}(\widehat{\beta}^e(u) - \beta_0(u)) + \mathbb{G}_n[g^e] + o_{\mathbb{P}}(1), \end{aligned}$$

where relations (1) and (2) follow by Lemma 1 and Lemma 2 with $\widetilde{\beta} = \widehat{\beta}^e(u)$, respectively, using that $\|\widehat{\vartheta}^e - \widetilde{\vartheta}^e\|_\infty = o_{\mathbb{P}}(1/\sqrt{n})$, $\widetilde{\vartheta}^e \in \Upsilon$, $\|\widetilde{\vartheta}^e - \vartheta_0\|_\infty = O_{\mathbb{P}}(1/\sqrt{n})$ and $\|\widehat{\beta}^e(u) - \beta_0(u)\|_2 = O_{\mathbb{P}}(1/\sqrt{n})$.

Therefore by invertibility of $J(u)$,

$$\sqrt{n}(\widehat{\beta}^e(u) - \beta_0(u)) = -J(u)^{-1}\mathbb{G}_n(f^e + g^e) + o_{\mathbb{P}}(1).$$

Step 3. In this final step we establish the behavior of $\sqrt{n}(\widehat{\beta}^e(u) - \widehat{\beta}(u))$ under $\mathbb{P}^e$. Note that $\mathbb{P}^e$ denotes the conditional probability measure, namely the probability measure induced by draws of $e_1, ..., e_n$ conditional on the data $A_1, ..., A_n$. By Step 2 of the proof of Theorem 1 and Step 2 of this proof, we have that under $\mathbb{P}$:

$$\sqrt{n}(\widehat{\beta}^e(u)-\beta_0(u)) = -J(u)^{-1}\mathbb{G}_n(f^e+g^e)+o_{\mathbb{P}}(1), \ \sqrt{n}(\widehat{\beta}(u)-\beta_0(u)) = -J(u)^{-1}\mathbb{G}_n(f+g)+o_{\mathbb{P}}(1).$$

Hence, under $\mathbb{P}$

$$\sqrt{n}(\widehat{\beta}^e(u)-\widehat{\beta}(u)) = -J(u)^{-1}\mathbb{G}_n(f^e-f+g^e-g)+r_n = -J(u)^{-1}\mathbb{G}_n((e-1)(f+g))+r_n, \ r_n = o_{\mathbb{P}}(1).$$

Note that it is also true that

$$r_n = o_{\mathbb{P}^e}(1) \text{ in } \mathbb{P}\text{-probability},$$

where the latter statement means that for every $\epsilon > 0$, $\mathbb{P}^e(\|r_n\|_2 > \epsilon) \to_{\mathbb{P}} 0$. Indeed, this follows from Markov inequality and by

$$\mathrm{E}_{\mathbb{P}}[\mathbb{P}^e(\|r_n\|_2 > \epsilon)] = \mathbb{P}(\|r_n\|_2 > \epsilon) = o(1),$$

where the latter holds by the Law of Iterated Expectations and $r_n = o_{\mathbb{P}}(1)$.

By the Conditional Multiplier Central Limit Theorem, e.g., Lemma 2.9.5 in van der Vaart and Wellner (1996), we have that conditional on the data $A_1, ..., A_n$

$$\mathbb{G}_n((e-1)(f+g)) \to_d Z := N(0, \mathrm{Var}_{\mathbb{P}}(f+g)), \text{ in } \mathbb{P}\text{-probability},$$

where the statement means that for each $z \in \mathbb{R}^{\dim(X)}$

$$\mathbb{P}^e(\mathbb{G}_n((e-1)(f+g)) \leq z) \to_{\mathbb{P}} \Pr(Z \leq z).$$

Conclude that conditional on the data $A_1, ..., A_n$

$$\sqrt{n}(\widehat{\beta}^e(u) - \widehat{\beta}(u)) \to_d N(0, J^{-1}(u)\mathrm{Var}_{\mathbb{P}}(f+g)J^{-1}(u)), \text{ in } \mathbb{P}\text{-probability},$$

where the statement means that for each $z \in \mathbb{R}^{\dim(X)}$

$$\mathbb{P}^e(\sqrt{n}(\widehat{\beta}^e(u) - \widehat{\beta}(u)) \leq z) \to_{\mathbb{P}} \Pr(-J^{-1}(u)Z \leq z).$$

□

B.3. **Lemma on Stochastic Equicontinuity.**

Lemma 1 (Stochastic equicontinuity)**.** *Let $e \geq 0$ be a positive random variable with unit mean and finite variance that is independent of (Y, D, W, Z, X, V), including as a special case $e = 1$, and set, for $A = (e, Y, D, W, Z, X, V)$ and $\chi = 1(X'\beta_0(u) > C + \epsilon')$,*

$$f(A, \vartheta, \beta, \gamma) := e \cdot [1(Y \leq X(\vartheta)'\beta) - u] \cdot X(\vartheta) \cdot 1(S(\vartheta)'\gamma > \varsigma) \cdot \chi.$$

Under the assumptions of the paper, the following relations are true.

(a) *Consider the set of functions*

$$\mathcal{F} = \{f(A, \vartheta, \beta, \gamma)'\alpha : (\vartheta, \beta) \in \Upsilon_0 \times \mathcal{B}, \gamma \in \Gamma, \alpha \in \mathbb{R}^{\dim(X)}, \|\alpha\|_2 \leq 1\},$$

where Γ is an open neighborhood of γ_0 under the $\|\cdot\|_2$ metric, $\mathcal{B}$ is an open neighborhood of $\beta_0(u)$ under the $\|\cdot\|_2$ metric, Υ_0 is the intersection of Υ, defined in Assumption 3, with a small neighborhood of ϑ_0 under the $\|\cdot\|_\infty$ metric, which are chosen to be small enough so that:

$$|X(\vartheta)'\beta - X'\beta_0(u)| \leq \epsilon'/2, \text{ P-a.e. } \forall (\vartheta, \beta) \in \Upsilon_0 \times \mathcal{B},$$

where ϵ' is defined in Assumption 4. This class is P-Donsker with a square integrable envelope of the form e times a constant.

(b) *Moreover, if $(\vartheta, \beta, \gamma) \to (\vartheta_0, \beta_0(u), \gamma_0)$ in the $\|\cdot\|_\infty \vee \|\cdot\|_2 \vee \|\cdot\|_2$ metric, then*

$$\|f(A, \vartheta, \beta, \gamma) - f(A, \vartheta_0, \beta_0(u), \gamma_0)\|_{P,2} \to 0.$$

(c) *Hence for any* $(\widetilde{\vartheta}, \widetilde{\beta}, \widehat{\gamma}) \to_{\mathbb{P}} (\vartheta_0, \beta_0(u), \gamma_0)$ *in the* $\|\cdot\|_\infty \vee \|\cdot\|_2 \vee \|\cdot\|_2$ *metric such that* $\widetilde{\vartheta} \in \Upsilon_0$,

$$\|\mathbb{G}_n f(A, \widetilde{\vartheta}, \widetilde{\beta}, \widehat{\gamma}) - \mathbb{G}_n f(A, \vartheta_0, \beta_0(u), \gamma_0)\|_2 \to_{\mathbb{P}} 0.$$

(d) *For for any* $(\widehat{\vartheta}, \widetilde{\beta}, \widehat{\gamma}) \to_{\mathbb{P}} (\vartheta_0, \beta_0(u), \gamma_0)$ *in the* $\|\cdot\|_\infty \vee \|\cdot\|_2 \vee \|\cdot\|_2$ *metric, so that*

$$\|\widehat{\vartheta} - \widetilde{\vartheta}\|_\infty = o_{\mathbb{P}}(1/\sqrt{n}), \text{ where } \widetilde{\vartheta} \in \Upsilon_0,$$

we have that

$$\|\mathbb{G}_n f(A, \widehat{\vartheta}, \widetilde{\beta}, \widehat{\gamma}) - \mathbb{G}_n f(A, \vartheta_0, \beta_0(u), \gamma_0)\|_2 \to_{\mathbb{P}} 0.$$

Proof of Lemma 1. The proof is divided in proofs of each of the claims.

Proof of Claim (a). The proof proceeds in several steps.

Step 1. Here we bound the bracketing entropy for

$$\mathcal{I}_1 = \{[1(Y \le X(\vartheta)'\beta) - u]\chi : \beta \in \mathcal{B}, \vartheta \in \Upsilon_0\}.$$

For this purpose consider a mesh $\{\vartheta_k\}$ over Υ_0 of $\|\cdot\|_\infty$ width δ, and a mesh $\{\beta_l\}$ over $\mathcal{B}$ of $\|\cdot\|_2$ width δ. A generic bracket over $\mathcal{I}_1$ takes the form

$$[i_1^0, i_1^1] = [\{1(Y \le X(\vartheta_k)'\beta_l - \kappa\delta) - u\}\chi, \{1(Y \le X(\vartheta_k)'\beta_l + \kappa\delta) - u\}\chi],$$

where $\kappa = L_X \max_{\beta \in \mathcal{B}} \|\beta\|_2 + L_X$, and $L_X := \|\partial_v x\|_\infty \vee \|x\|_\infty$.

Note that this is a valid bracket for all elements of $\mathcal{I}_1$ induced by any ϑ located within δ from ϑ_k and any β located within δ from β_l, since

$$\begin{aligned} |X(\vartheta)'\beta - X(\vartheta_k)'\beta_l| &\le |(X(\vartheta) - X(\vartheta_k))'\beta| + |X(\vartheta_k)'(\beta - \beta_k)| \\ &\le L_X \delta \max_{\beta \in \mathcal{B}} \|\beta\|_2 + L_X \delta \le \kappa\delta, \end{aligned} \tag{B.5}$$

and the $L_2(P)$ size of this bracket is given by

$$\begin{aligned} \|i_1^0 - i_1^1\|_{P,2} &\le \sqrt{\mathrm{E}_P[P\{Y \in [X(\vartheta_k)'\beta_l \pm \kappa\delta] \mid D, W, Z, C, \chi = 1\}]} \\ &\le \sqrt{\mathrm{E}_P[\sup_{y \in (C + \kappa\delta, \infty)} P\{Y \in [y \pm \kappa\delta] \mid X, C, \chi = 1\}]} \\ &\le \sqrt{\|f_Y(\cdot \mid \cdot)\|_\infty 2\kappa\delta}, \end{aligned}$$

provided that $2\kappa\delta < \epsilon'/2$. In order to derive this bound we use the condition $|X(\vartheta)'\beta - X'\beta_0(u)| \le \epsilon'/2$, P-a.e. $\forall (\vartheta, \beta) \in \Upsilon_0 \times \mathcal{B}$, so that conditional on $\chi = 1$ we have that $X(\vartheta)'\beta > C + \epsilon'/2$; and

$$P\{Y \in \cdot \mid D, W, Z, C, \chi = 1\} = P\{Y \in \cdot \mid D, W, Z, V, C, \chi = 1\} = P\{Y \in \cdot \mid X, C, \chi = 1\},$$

because $V = \vartheta_0(D, W, Z)$ and the exclusion restriction for Z. Hence, conditional on X, C and $\chi = 1$, Y does not have point mass in the region $[X(\vartheta_k)'\beta_l \pm \kappa\delta] \subset (C, \infty)$, and by assumption the density of Y conditional on X, C is uniformly bounded over the region (C, ∞).

It follows that

$$\log N_{[]}(\epsilon, \mathcal{I}_1, L_2(P)) \lesssim \log N(\epsilon^2, \Upsilon_0, \|\cdot\|_\infty) + \log N(\epsilon, \mathcal{B}, \|\cdot\|_2) \lesssim 1/(\epsilon^2 \log^2 \epsilon) + \log(1/\epsilon),$$

and so $\mathcal{I}_1$ is P-Donsker with a constant envelope.

Step 2. Similarly to Step 1, it follows that

$$\mathcal{I}_2 = \{X(\vartheta)'\alpha : \vartheta \in \Upsilon_0, \alpha \in \mathbb{R}^{\dim(X)}, \|\alpha\|_2 \leq 1\}$$

also obeys a similar bracketing entropy bound

$$\log N_{[]}(\epsilon, \mathcal{I}_2, L_2(P)) \lesssim 1/(\epsilon^2 \log^2 \epsilon) + \log(1/\epsilon)$$

with a generic bracket taking the form $[i_2^0, i_2^1] = [X(\vartheta_k)'\beta_l - \kappa\delta, X(\vartheta_k)'\beta_l + \kappa\delta]$. Hence, this class is also P-Donsker with a constant envelope.

Step 3. Here we bound the bracketing entropy for

$$\mathcal{I}_3 = \{1(S(\vartheta)'\gamma \geq \varsigma) : \vartheta \in \Upsilon_0, \gamma \in \Gamma\}.$$

For this purpose consider the mesh $\{\vartheta_k\}$ over Υ_0 of $\|\cdot\|_\infty$ width δ, and a mesh $\{\gamma_l\}$ over Γ of $\|\cdot\|_2$ width δ. A generic bracket over $\mathcal{I}_3$ takes the form

$$[i_3^0, i_3^1] = [1(S(\vartheta_k)'\gamma_l - \kappa\delta \geq \varsigma), 1(S(\vartheta_k)'\gamma_l + \kappa\delta \geq \varsigma)],$$

where $\kappa = L_S \max_{\gamma \in \Gamma} \|\gamma\|_2 + L_S$, and $L_S := \|\partial_v s\|_\infty \vee \|s\|_\infty$.

Note that this is a valid bracket for all elements of $\mathcal{I}_3$ induced by any ϑ located within δ from ϑ_k and any γ located within δ from γ_l, since

$$\begin{aligned} |S(\vartheta)'\gamma - S(\vartheta_k)'\gamma_l| &\leq |(S(\vartheta) - S(\vartheta_k))'\gamma| + |S(\vartheta_k)'(\gamma - \gamma_k)| \\ &\leq L_S\delta \max_{\gamma \in \Gamma} \|\gamma\|_2 + L_S\delta \leq \kappa\delta, \end{aligned} \tag{B.6}$$

and the size of this bracket is given by

$$\|i_3^0 - i_3^1\|_{P,2} \leq \sqrt{P\{|S(\vartheta_k)'\gamma_l - \varsigma| \leq 2\kappa\delta\}} \leq \sqrt{\bar{f}_S 2\kappa\delta},$$

where $\bar{f}_S$ is a constant representing the uniform upper bound on the density of random variable $S(\vartheta)'\gamma$, where the uniformity is over $\vartheta \in \Upsilon_0$ and $\gamma \in \Gamma$.

It follows that

$$\log N_{[]}(\epsilon, \mathcal{I}_3, L_2(P)) \lesssim \log N(\epsilon^2, \Upsilon_0, \|\cdot\|_\infty) + \log N(\epsilon, \Gamma, \|\cdot\|_2) \lesssim 1/(\epsilon^2 \log^2 \epsilon) + \log(1/\epsilon)$$

and so $\mathcal{I}_3$ is P-Donsker with a constant envelope.

Step 4. In this step we verify the claim (a). Note that $\mathcal{F} = e \cdot \mathcal{I}_1 \cdot \mathcal{I}_2 \cdot \mathcal{I}_3$. This class has a square-integrable envelope under P.

If we make a further assumption that the random variable e obeys $\mathrm{E}_P|e|^{2+\delta} < \infty$, for $\delta > 0$, then the class $\mathcal{F}$ is P-Donsker by the following argument. Note that the product $\mathcal{I}_1 \cdot \mathcal{I}_2 \cdot \mathcal{I}_3$ of uniformly bounded classes is P-Donsker, e.g., by Theorem 2.10.6 of van der Vaart and Wellner (1996). Under the stated assumption the final product of the random variable e with the P-Donsker class remains to be P-Donsker by the Multiplier Donsker Theorem, namely Theorem 2.9.2 in van der Vaart and Wellner (1996).

It is also interesting to give a direct argument based on bracketing, which also requires the somewhat weaker assumption of $e \geq 0$ having finite variance. Indeed, our generic brackets over $\mathcal{F}$ take the form

$$[f^0, f^1] = \left[\min_{(g,h,j) \in \{0,1\}^3} e \; i_1^g \; i_2^h \; i_3^j, \max_{(g,h,j) \in \{0,1\}^3} e \; i_1^g \; i_2^h \; i_3^j \right].$$

For example, if $i_1^0, i_2^0, i_3^0 \geq 0$, then $f^0 = e \; i_1^0 \; i_2^0 \; i_3^0$ and $f^1 = e \; i_1^1 \; i_2^1 \; i_3^1$. Note that these are valid brackets for all f that are induced by an element of $\mathcal{I}_1$ located in the bracket $[i_1^0, i_1^1]$, an element of $\mathcal{I}_2$ located in the bracket $[i_2^0, i_2^1]$, and an element of $\mathcal{I}_3$ located in the bracket $[i_3^0, i_3^1]$. Using that the constructed brackets are uniformly bounded, by say k, the size of this bracket is

$$\|f^0 - f^1\|_{P,2} \leq \sqrt{\mathrm{E}_P e^2} k^2 (\|i_1^0 - i_1^1\|_{P,2} + \|i_2^0 - i_2^1\|_{P,2} + \|i_3^0 - i_3^1\|_{P,2}) \leq \sqrt{\mathrm{E}_P e^2} k^2 3\delta,$$

where we used the independence of random variable e from random vector A. Therefore,

$$\log N_{[]}(\epsilon, \mathcal{F}, L_2(P)) \lesssim \log N_{[]}(\epsilon, \mathcal{I}_1, L_2(P)) + \log N_{[]}(\epsilon, \mathcal{I}_2, L_2(P)) + \log N_{[]}(\epsilon, \mathcal{I}_3, L_2(P)),$$

and the class is thus P-Donsker by the previous steps.

Proof of Claim (b). The claim follows by the Dominated Convergence Theorem, since any $f \in \mathcal{F}$ is dominated by a square-integrable envelope under P, and by the following three facts:

(1) in view of the relation such as (B.5), $1(Y \leq X(\vartheta)'\beta)\chi \to 1(Y \leq X'\beta_0(u))\chi$ everywhere, except for the set $\{A \in \mathcal{A} : Y = X'\beta_0(u)\}$ whose measure under P is zero by Y having a uniformly bounded density conditional on X, C;
(2) in view of the relation such as (B.5), $|X(\vartheta)'\beta - X'\beta_0(u)| \to 0$ everywhere;
(3) in view of the relation such as (B.6), $1(S(\vartheta)'\gamma \geq \varsigma) \to 1(S'\gamma_0 \geq \varsigma)$ everywhere, except for the set $\{A \in \mathcal{A} : S'\gamma_0 = \varsigma\}$ whose measure under P is zero by $S'\gamma_0$ having a bounded density.

Proof of Claim (c). This claim follows from the asymptotic equicontinuity of the empirical process $(\mathbb{G}_n[f], f \in \mathcal{F})$ under the $L_2(P)$ metric, and hence also with respect to the $\|\cdot\|_\infty \vee \|\cdot\|_2 \vee \|\cdot\|_2$ metric in view of Claim (b).

Proof of Claim (d). It is convenient to set $\widehat{f} := f(A, \widehat{\vartheta}, \widetilde{\beta}, \widehat{\gamma})$ and $\widetilde{f} := f(A, \widetilde{\vartheta}, \widetilde{\beta}, \widehat{\gamma})$. Note that

$$
\begin{aligned}
|\mathbb{G}_n[\widehat{f} - \widetilde{f}]| &\leq |\sqrt{n}\mathbb{E}_n[\widehat{f} - \widetilde{f}]| + |\sqrt{n}\mathrm{E}_P(\widehat{f} - \widetilde{f})| && \text{(B.7)}\\
&\lesssim \sqrt{n}\mathbb{E}_n[\widehat{\zeta}\,] + \sqrt{n}\mathrm{E}_P[\widehat{\zeta}\,] && \text{(B.8)}\\
&\lesssim \mathbb{G}_n[\widehat{\zeta}\,] + 2\sqrt{n}\mathrm{E}_P[\widehat{\zeta}\,], && \text{(B.9)}
\end{aligned}
$$

where $|f|$ denote an application of absolute value to each element of the vector f, and $\widehat{\zeta}$ is defined by the following relationship, which holds with with probability approaching one,

$$
|\widehat{f} - \widetilde{f}| \lesssim |e| \cdot \|X(\widehat{\vartheta}) - X(\widetilde{\vartheta})\|_2 + \widehat{g} + \widehat{h} \lesssim \widehat{\zeta} := e \cdot L_X \Delta_n + \widehat{g} + \widehat{h}, \quad \Delta_n \geq \|\widehat{\vartheta} - \widetilde{\vartheta}\|_\infty, \tag{B.10}
$$

where $L_X = \|\partial_v x\|_\infty \vee \|x\|_\infty$, and, for some constant k,

$$
\widehat{g} := e \cdot 1\{|Y - X(\widetilde{\vartheta})'\widetilde{\beta}| \leq k\Delta_n\}\chi, \quad \text{and} \quad \widehat{h} := e \cdot 1\{|S(\widetilde{\vartheta})'\widehat{\gamma} - \varsigma| \leq k\Delta_n\},
$$

and $\Delta_n = o(1/\sqrt{n})$ is a deterministic sequence.

Hence it suffices to show that the result follows from

$$
\mathbb{G}_n[\widehat{\zeta}\,] = o_{\mathbb{P}}(1), \tag{B.11}
$$

and

$$
\sqrt{n}\mathrm{E}_P[\widehat{\zeta}\,] = o_{\mathbb{P}}(1). \tag{B.12}
$$

Note that since $\Delta_n \to 0$, with probability approaching one, $\widehat{g}$ and $\widehat{h}$ are elements of the function classes

$$
\begin{aligned}
\mathcal{G} &= \{e \cdot 1(|Y - X(\vartheta)'\beta| \leq k)\chi : \vartheta \in \Upsilon_0, \beta \in \mathcal{B}, k \in [0, \epsilon'/4]\},\\
\mathcal{H} &= \{e \cdot 1(|S(\vartheta)'\gamma - \varsigma| \leq k) : \vartheta \in \Upsilon_0, \gamma \in \Gamma, k \in [0, 1]\}.
\end{aligned}
$$

By the argument similar to that in the proof of claim (a), we have that

$$
\log N_{[]}(\epsilon, \mathcal{G}, L_2(P)) \lesssim 1/(\epsilon^2 \log^2 \epsilon) \text{ and } \log N_{[]}(\epsilon, \mathcal{H}, L_2(P)) \lesssim 1/(\epsilon^2 \log^2 \epsilon).
$$

Hence these classes are P-Donsker with unit envelopes. Also note that

$$
\|\widehat{g}\|_{P,2} \leq \sqrt{\mathrm{E}_P[e^2] \cdot P\{|Y - X(\widetilde{\vartheta})'\widetilde{\beta}| \leq k\Delta_n\}} \leq \sqrt{4\|f_Y(\cdot \mid \cdot)\|_\infty k\Delta_n} = o(1), \tag{B.13}
$$

$$
\|\widehat{h}\|_{P,2} \leq \sqrt{\mathrm{E}_P[e^2] \cdot P\{|S(\widetilde{\vartheta})'\widehat{\gamma} - \varsigma| \leq k\Delta_n\}} \leq \sqrt{4\bar{f}_S k\Delta_n} = o(1), \tag{B.14}
$$

by the assumption on bounded densities and $\mathrm{E}_P[e^2] = 2$.

Conclude that the relation (B.11) holds by (B.10), (B.13), (B.14), the P-Donskerity of the empirical processes $(\mathbb{G}_n[h], h \in \mathcal{H})$ and $(\mathbb{G}_n[g], g \in \mathcal{G})$ and hence their asymptotic equicontinuity under the $\|\cdot\|_{P,2}$ metric. Indeed, these properties imply

$$\|\widehat{\zeta}\|_{P,2} = o(1) \Rightarrow \mathbb{G}_n[\widehat{\zeta}\,] = o(1).$$

To show (B.12) note that

$$\sqrt{n}\mathrm{E}_P|\widehat{\zeta}| \lesssim \mathrm{E}_P|e| \cdot L_X\sqrt{n}\Delta_n + \|\widehat{g}\|_{P,1} + \|\widehat{h}\|_{P,1} = o(1),$$

since $\Delta_n = o(1/\sqrt{n})$, and

$$\begin{aligned}
\|\widehat{g}\|_{P,1} &\leq \mathrm{E}_P|e| \cdot P\{|Y - X(\widetilde{\vartheta})'\widetilde{\beta}| \leq k\Delta_n\} \leq 2k\|f_Y(\cdot \mid \cdot)\|_\infty \Delta_n = o(1/\sqrt{n}) \\
\|\widehat{h}\|_{P,1} &\leq \mathrm{E}_P|e| \cdot P\{|S(\widetilde{\vartheta})'\widehat{\gamma} - \varsigma| \leq k\Delta_n\} \leq 2k\bar{f}_S\Delta_n = o(1/\sqrt{n}),
\end{aligned}$$

by the assumption on bounded densities.

B.4. **Lemma on Local Expansion.**

Lemma 2 (Local expansion)**.** *Under the assumptions stated in the paper, for*

$$\begin{aligned}
&\widehat{\delta} = \sqrt{n}(\widetilde{\beta} - \beta_0(u)) = O_{\mathbb{P}}(1); \widehat{\gamma} = \gamma_0 + o_{\mathbb{P}}(1); \\
&\widehat{\Delta}(d, w, z) = \sqrt{n}(\widehat{\vartheta}(d, w, z) - \vartheta_0(d, w, z)) = \sqrt{n}\ \mathbb{E}_n[\ell(A, d, w, z)] + o_{\mathbb{P}}(1) \ \textit{in}\ \ell^\infty(\mathcal{DR}), \\
&\|\sqrt{n}\ \mathbb{E}_n[\ell(A, \cdot)]\|_\infty = O_{\mathbb{P}}(1),
\end{aligned}$$

we have that

$$\sqrt{n}\ \mathrm{E}_P 1\{Y \leq X(\widehat{\vartheta})'\widetilde{\beta}\}X(\widehat{\vartheta})1\{S(\widehat{\vartheta})'\gamma \geq \varsigma\}\chi = J(u)\widehat{\delta} + \sqrt{n}\ \mathbb{E}_n\left[g(A)\right] + o_{\mathbb{P}}(1),$$

where

$$g(A) = \int B(a)\ell(A, d, r)dP(a, d, r), \quad B(A) := f_Y(X'\beta_0(u)|X, C)X\dot{X}'\beta_0(u)1(S'\gamma_0 \geq \varsigma).$$

Proof of Lemma 2. We have that with probability approaching one,

$$1\{S(\widehat{\vartheta})'\widehat{\gamma} \geq \varsigma\} \leq 1\{S'\gamma_0 > \varsigma - \epsilon_n\} \leq 1\{S'\gamma_0 > \varsigma/2\} \leq 1\{X'\beta_0(u) > C + \epsilon'\} = \chi,$$

P-a.e., by assumption on the selector since

$$|S(\widehat{\vartheta})'\widehat{\gamma} - S'\gamma_0| \leq \epsilon_n := L_S(\|\widehat{\vartheta} - \vartheta_0\|_\infty + \|\widehat{\gamma} - \gamma_0\|_2) \to_{\mathbb{P}} 0,$$

where $L_S = (\|\partial_v s\|_\infty \vee \|s\|_\infty)$ is a finite constant by assumption.

Hence uniformly in X such that $X'\beta_0(u) > C + \epsilon'$,

$$
\begin{aligned}
&\sqrt{n}\mathrm{E}_P[1\{Y \leq X(\widehat{\vartheta})'\widetilde{\beta}\} \mid D, W, Z, C] \\
&= f_Y(X(\bar{\vartheta}_X)'\bar{\beta}_X \mid D, W, Z, C)\{X(\bar{\vartheta}_X)'\widehat{\delta} + \dot{X}(\bar{\vartheta}_X)'\bar{\beta}_X\widehat{\Delta}(D, W, Z)\} \\
&= f_Y(X'\beta_0(u) \mid D, W, Z, C)\{X'\widehat{\delta} + \dot{X}'\beta_0(u)\widehat{\Delta}(D, W, Z)\} + R_X, \\
&= f_Y(X'\beta_0(u) \mid X, C)\{X'\widehat{\delta} + \dot{X}'\beta_0(u)\widehat{\Delta}(D, W, Z)\} + R_X, \\
&\bar{R} = \sup_{\{X: X'\beta_0(u) > C + \epsilon'\}} |R_X| = o_{\mathbb{P}}(1),
\end{aligned}
$$

where $\bar{\vartheta}_X$ is on the line connecting ϑ_0 and $\widehat{\vartheta}$ and $\bar{\beta}_X$ is on the line connecting $\beta_0(u)$ and $\widetilde{\beta}$. The first equality follows by the mean value expansion. The second equality follows by the uniform continuity assumption of $f_Y(\cdot \mid X, C)$ uniformly in X, C, uniform continuity of $X(\cdot)$ and $\dot{X}(\cdot)$, and by $\|\widehat{\vartheta} - \vartheta_0\|_\infty \to_{\mathbb{P}} 0$ and $\|\widetilde{\beta} - \beta_0(u)\|_2 \to_{\mathbb{P}} 0$. The third equality follows by

$$f_Y(\cdot \mid D, W, Z, C) = f_Y(\cdot \mid D, W, Z, V, C) = f_Y(\cdot \mid X, C)$$

because $V = \vartheta_0(D, W, Z)$ and the exclusion restriction for Z.

Since $f_Y(\cdot \mid \cdot)$ and the entries of X and $\dot{X}$ are bounded, and $\widehat{\delta} = O_{\mathbb{P}}(1)$ and $\|\widehat{\Delta}\|_\infty = O_{\mathbb{P}}(1)$, with probability approaching one

$$
\begin{aligned}
&\mathrm{E}_P 1(Y \leq X(\widehat{\vartheta})'\widetilde{\beta})X(\widehat{\vartheta})1(S(\widehat{\vartheta})'\widehat{\gamma} \geq \varsigma)\chi \\
&= \mathrm{E}_P[f_Y(X'\beta_0(u)|X, C)XX'1\{S(\widehat{\vartheta})'\widehat{\gamma} \geq \varsigma\}]\widehat{\delta} \\
&+\mathrm{E}_P[f_Y(X'\beta_0(u)|X, C)X\dot{X}'\beta_0(u)1\{S(\widehat{\vartheta})'\widehat{\gamma} \geq \varsigma\}\widehat{\Delta}(D, W, Z)] + O_{\mathbb{P}}(\bar{R}). \qquad \text{(B.15)}
\end{aligned}
$$

Furthermore since

$$\mathrm{E}_P|1(S'\gamma_0 \geq \varsigma) - 1(S(\widehat{\vartheta})'\widehat{\gamma} \geq \varsigma)| \leq \mathrm{E}_P[1(S'\gamma_0 \in [\varsigma \pm \epsilon_n])] \lesssim \bar{f}_S\epsilon_n \to_{\mathbb{P}} 0,$$

where $\bar{f}_S$ is a constant representing the uniform upper bound on the density of random variable $S'\gamma_0$, the expression (B.15) is equal to

$$J(u)\widehat{\delta} + \mathrm{E}_P[f_Y(X'\beta_0(u)|X, C)X\dot{X}'\beta_0(u)1(S'\gamma_0 \geq \varsigma)\widehat{\Delta}(D, W, Z)] + O_{\mathbb{P}}(\bar{f}_S\epsilon_n + \bar{R}).$$

Substituting in $\widehat{\Delta}(d, w, z) = \sqrt{n}\ \mathbb{E}_n[\ell(A, d, w, z)] + o_{\mathbb{P}}(1)$ and interchanging E_P and $\mathbb{E}_n$, we obtain

$$\mathrm{E}_P[f_Y(X'\beta_0(u)|X, C)X\dot{X}'\beta_0(u)1(S'\gamma_0 \geq \varsigma)\widehat{\Delta}(D, W, Z)] = \sqrt{n}\ \mathbb{E}_n[g(A)] + o_{\mathbb{P}}(1).$$

The claim of the lemma follows. $\square$

Appendix C. Proof of Theorem 3

To show claim (1), we first note that by Chernozhukov, Fernández-Val and Melly (2009),

$$\sqrt{n}(\widehat{\pi}(v)-\pi_0(v))=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}e_i\mathrm{E}_P\left[f_D(R'\pi_0(v)\mid R)RR'\right]^{-1}[v-1\{D\leq R'\pi_0(v)\}]R.$$

By the Hadarmard differentiability of rearrangement-related operators in Chernozhukov, Fernández-Val and Galichon (2010), the mapping $\pi\mapsto\phi_\pi$ from $\ell^\infty(0,1)^{\dim(R)}$ to $\ell^\infty(\mathcal{DR})$ defined by

$$\phi_\pi(d,r)=\int_{(0,1)}1\{r'\pi(v)\leq d\}dv$$

is Hadamard differentiable at $\pi=\pi_0$, tangentially to the set of continuous directions, with the derivative given by

$$\dot{\phi}_{\pi_0}[h]=-f_D(d\mid r)r'h(\vartheta_0(d,r)),$$

where $\vartheta_0(d,r)=\phi_{\pi_0}(d,r)$. Therefore by the Functional Delta Method (Theorem 20.8 in van der Vaart, 1998), we have that in $\ell^\infty(\mathcal{DR})$, for $\widehat{\vartheta}(d,r)=\phi_{\widehat{\pi}}(d,r)$,

$$\begin{aligned}\sqrt{n}(\widehat{\vartheta}(d,r)-\vartheta_0(d,r)) &= -f_D(d\mid r)r'\tfrac{1}{\sqrt{n}}\textstyle\sum_{i=1}^{n}e_i\mathrm{E}_P\left[f_D(R'\pi_0(\vartheta_0(d,r))\mid R)RR'\right]^{-1}\times\\ &\quad\times[\vartheta_0(d,r)-1\{D\leq R'\pi_0(\vartheta_0(d,r))\}]R+o_{\mathbb{P}}(1).\end{aligned}$$

The claim (1) then follows immediately. Also for future reference, note that the result also implies that

$$\sqrt{n}(\widehat{\pi}(\cdot)-\pi_0(\cdot))\Rightarrow Z_\pi\ \text{ in }\ell^\infty(0,1),\quad\text{and}\quad r'\sqrt{n}(\widehat{\pi}(\cdot)-\pi_0(\cdot))\Rightarrow r'Z_\pi\quad\text{in }\ell^\infty(\mathcal{R}\times(0,1)),\tag{C.1}$$

where Z_π is a Gaussian process with continuous sample paths.

The proof of claim (2) is divided in several steps:

Step 1. In this step we construct Υ and bound its entropy. Let $C^2_M(0,1)$ denote the class of functions $f:(0,1)\to\mathbb{R}$ with all derivatives up to order 2 bounded by a constant M, including the zero order derivative. The covering entropy of this class is known to obey $\log N(\epsilon,C^2_M(0,1),\|\cdot\|_\infty)\lesssim\epsilon^{-1/2}$. Hence also $\log N(\epsilon,\times_{j=1}^{\dim(R)}C^2_M(0,1),\|\cdot\|_\infty)\lesssim\epsilon^{-1/2}$. Next construct for some small $m>0$:

$$\Upsilon=\left\{\int_{(0,1)}1\{R'\pi(v)\leq D\}dv:\pi=(\pi_1,...,\pi_{\dim(R)})\in\times_{j=1}^{\dim(R)}C^2_M(0,1),R'\partial\pi(v)>m\ P\text{-a.e.}\right\}.$$

Then for any π and $\bar{\pi}$ obeying the conditions in the display and such that $\|\pi-\bar{\pi}\|_\infty\leq\delta$,

$$\begin{aligned}&\left|\int_{(0,1)}1\{R'\pi(v)\leq D\}dv-\int_{(0,1)}1\{R'\bar{\pi}(v)\leq D\}dv\right|\\ &\leq\int_{(0,1)}1\Big\{R'\pi(v)-D\in[-\|R\|_2\delta,\|R\|_2\delta]\Big\}dv\lesssim\frac{1}{m}\|R\|_2\delta\lesssim\delta,\end{aligned}$$

P-a.e., since the density of $r'\pi(V)$, $V \sim U(0,1)$, is bounded above by $1/m$. We conclude that

$$\log N(\epsilon, \Upsilon, \|\cdot\|_\infty) \lesssim \log N(\epsilon, \times_{j=1}^{\dim(R)} C_M^2(0,1), \|\cdot\|_\infty) \lesssim \epsilon^{-1/2}.$$

Step 2. In this step we show that there exists $\widetilde{\vartheta} \in \Upsilon$ such that $\|\widetilde{\vartheta} - \widehat{\vartheta}\|_\infty = o_{\mathbb{P}}(1/\sqrt{n})$.

We first construct $\widetilde{\pi}$ such that

$$\sqrt{n}\|\widetilde{\pi} - \widehat{\pi}\|_\infty = o_{\mathbb{P}}(1), \quad \text{and} \quad \max_{r \in \mathcal{R}} \sqrt{n}\|r'(\widetilde{\pi} - \widehat{\pi})\|_\infty = o_{\mathbb{P}}(1), \tag{C.2}$$

where with probability approaching one, $\widetilde{\pi} \in \times_{j=1}^{\dim(R)} C_M^2(0,1)$ and $R'\partial\widetilde{\pi}(v) > m$ P-a.e., for some M and some $m > 0$.

We construct $\widetilde{\pi}$ by smoothing $\widehat{\pi}$ component by component. Let the components of $\widehat{\pi}$ be indexed by $1 \le j \le \dim(R)$. Before smoothing, we need to extend $\widehat{\pi}_j$ outside $(0,1)$. Start by extending the estimand π_{0j} outside $(0,1)$ onto the ϵ-expansion $(0,1)^\epsilon$ smoothly so that the extended function is in the class C^3. This is possible by first extending $\partial^3\pi_{0j}$ smoothly and then integrating up to obtain lower order derivatives and the function. Then we extend the estimator $\widehat{\pi}_j$ to the outer region by simply setting $\widehat{\pi}_j(v) = \pi_{0j}(v)$ over $v \notin (0,1)$. Note that the extended empirical process $\sqrt{n}(\widehat{\pi}_j(v) - \pi_{0j}(v))$ remains to be stochastically equicontinuous, since it vanishes asymptotically at the boundary of $(0,1)$. Then we construct $\widetilde{\pi}_j$ as the smoothed version of $\widehat{\pi}_j$, namely

$$\widetilde{\pi}_j(v) = \int_{(0,1)^\epsilon} \widehat{\pi}_j(z)[K((z-v)/h)/h]dz, \quad v \in (0,1),$$

where $0 \le h \le \epsilon$ is bandwidth such that $\sqrt{n}h^3 \to 0$ and $\sqrt{n}h^2 \to \infty$; $K : \mathbb{R} \to \mathbb{R}$ is a third order kernel with the properties: $\partial^\mu K$ are continuous on $[-1,1]$ and vanish outside of $[-1,1]$ for $\mu = 0,1,2$, $\int K(z)dz = 1$, and $\int z^\mu K(z)dz = 0$ for $\mu = 1,2$. Such kernel exists and can be obtained by reproducing kernel Hilbert space methods or via twicing kernel transformations (Berlinet, 1993, and Newey, Hsieh, and Robins, 2004). We then have

$$\begin{aligned} \sqrt{n}(\widetilde{\pi}_j(v) - \widehat{\pi}_j(v)) &= \int_{(0,1)^\epsilon} \sqrt{n}[\widehat{\pi}_j(z) - \pi_{0j}(z) - (\widehat{\pi}_j(v) - \pi_{0j}(v))][K([z-v]/h)/h]dz \\ &+ \int_{(0,1)^\epsilon} \sqrt{n}(\pi_{0j}(z) - \pi_{0j}(v))[K([z-v]/h)/h]dz. \end{aligned}$$

The first term is bounded uniformly in $v \in (0,1)$ by $\omega(2h)\|K\|_\infty \to_{\mathbb{P}} 0$ where

$$\omega(2h) = \sup_{|z-u| \le 2h} |\sqrt{n}[\widehat{\pi}_j(z) - \pi_{0j}(z) - (\widehat{\pi}_j(v) - \pi_{0j}(v))]| \to_{\mathbb{P}} 0.$$

The second term is bounded uniformly in $v \in (0,1)$, up to a constant, by

$$\sqrt{n}\|\partial^3\pi_{0j}\|_\infty h^3 \int \lambda^3 K(\lambda)d\lambda \lesssim \sqrt{n}h^3 \to 0.$$

This establishes the equivalence (C.2), in view of compactness of $\mathcal{R}$.

Next we show that $\|\partial^2\widetilde{\pi}_j\|_\infty \leq 2\|\partial^2\pi_{0j}\|_\infty =: M$ with probability approaching 1. Note that

$$\partial^2\widetilde{\pi}_j(v) - \partial^2\pi_{0j}(v) = \int_{(0,1)^\epsilon} \widehat{\pi}_j(z)[\partial^2 K([z-v]/h)/h^3]dz - \partial^2\pi_{0j}(v),$$

which can be decomposed into two pieces:

$$\begin{aligned}&n^{-1/2}h^{-2}\textstyle\int_{(0,1)^\epsilon} n^{1/2}(\widehat{\pi}_j(z)-\pi_{0j}(z))[\partial^2 K([z-v]/h)/h]dz\\ &+\textstyle\int_{(0,1)^\epsilon}[\partial^2 K([z-v]/h)/h^3]\pi_{0j}(z)dz - \partial^2\pi_{0j}(v).\end{aligned}$$

The first piece is bounded uniformly in $v \in (0,1)$ by $n^{-1/2}h^{-2}\omega(2h)\|\partial^2 K\|_\infty \to 0$, while, using the integration by parts, the second piece is equal to

$$\int_{(0,1)^\epsilon}[\partial^2\pi_{0j}(z) - \partial^2\pi_{0j}(v)][K([z-v]/h)/h]dz.$$

This expression is bounded in absolute value by

$$\|K\|_\infty \sup_{|z-v|\leq 2h} |\partial^2\pi_{0j}(z) - \partial^2\pi_{0j}(v)| \to 0,$$

by continuity of $\partial^2\pi_{0j}$ and compactness of $(0,1)^\epsilon$. Thus, we conclude that $\|\partial^2\widetilde{\pi}_j - \partial^2\pi_{0j}\|_\infty \to_{\mathbb{P}} 0$, and we can also deduce similarly that $\|\partial\widetilde{\pi}_j - \partial\pi_{0j}\|_\infty \to_{\mathbb{P}} 0$, all uniformly in $1 \leq j \leq \dim(R)$, since $\dim(R)$ is finite and fixed.

Finally, since by Assumption 8(b) the conditional density is uniformly bounded above by a constant, this implies that $R'\partial\pi_0(v) > k$ P-a.e., for some constant $k > 0$, and therefore we also have that with probability approaching one, $R'\partial\widetilde{\pi}(v) > k/2 =: m$ P-a.e.

Next we construct

$$\widetilde{\vartheta}(d,r) = \phi_{\widetilde{\pi}}(d,r) = \int_{(0,1)} 1\{r'\widetilde{\pi}(v) \leq d\}dv.$$

Note that by construction $\widetilde{\vartheta} \in \Upsilon$ for some M with probability approaching one. It remains to show the first order equivalence with $\widehat{\vartheta}$.

By the Hadarmard differentiability for the mapping ϕ_π stated earlier and by the functional delta method (Theorem 20.8 in van der Vaart, 1998), $\widetilde{\vartheta}$ and $\widehat{\vartheta}$ have the same first order representation in $\ell^\infty(\mathcal{DR})$,

$$\sqrt{n}(\widetilde{\vartheta}(\cdot) - \vartheta_0(\cdot)) = \sqrt{n}(\widehat{\vartheta}(\cdot) - \vartheta_0(\cdot)) + o_{\mathbb{P}}(1),$$

i.e., $\sqrt{n}\|\widetilde{\vartheta} - \widehat{\vartheta}\|_\infty \to_{\mathbb{P}} 0$. $\square$

Appendix D. Proof of Theorem 4

Claim (1) follows from the results of Chernozhukov, Fernández-Val and Melly (2009). Also for future reference, note that these results also imply that

$$\sqrt{n}(\widehat{\pi}(\cdot) - \pi_0(\cdot)) \Rightarrow Z_\pi \text{ in } \ell^\infty(\mathcal{D}), \quad \text{and} \quad r'\sqrt{n}(\widehat{\pi}(\cdot) - \pi_0(\cdot)) \Rightarrow r'Z_\pi \ \text{ in } \ell^\infty(\mathcal{R}\mathcal{D}), \qquad \text{(D.1)}$$

where Z_π is a Gaussian process with continuous sample paths.

The proof of claim (2) is divided in several steps:

Step 1. In this step we construct Υ and bound its entropy. Let $C^2_M(\mathcal{D})$ denote the class of functions $f : \mathcal{D} \to \mathbb{R}$ with and all the derivatives up to order 2 bounded by a constant M, including the zero order derivative. The covering entropy of this class is known to obey $\log(\epsilon, C^2_M(\mathcal{D}), \|\cdot\|_\infty) \lesssim \epsilon^{-1/2}$. Hence

$$\log(\epsilon, \times_{j=1}^{\dim(R)} C^2_M(\mathcal{D}), \|\cdot\|_\infty) \lesssim \epsilon^{-1/2}.$$

Next construct

$$\Upsilon = \left\{ \Lambda(R'\pi(D)) : \pi = (\pi_1, ..., \pi_{\dim(R)}) \in \times_{j=1}^{\dim(R)} C^2_M(\mathcal{D}) \right\}.$$

Then, for any π and $\bar{\pi}$ obeying the condition in the definition of the preceding class such that $\|\pi - \bar{\pi}\|_\infty \le \delta$,

$$|\Lambda(R'\pi(D)) - \Lambda(R'\bar{\pi}(D))| \le \|\partial\Lambda\|_\infty \sup_{r\in\mathcal{R}} \|r\|_\infty \delta.$$

We conclude that

$$\log N(\epsilon, \Upsilon, \|\cdot\|_\infty) \lesssim \log N(\epsilon, \times_{j=1}^{\dim(R)} C^2_M(\mathcal{D}), \|\cdot\|_\infty) \lesssim \epsilon^{-1/2}.$$

Step 2. In this step we show that there exists $\widetilde{\vartheta} \in \Upsilon$ such that $\|\widetilde{\vartheta} - \widehat{\vartheta}\|_\infty = o_{\mathbb{P}}(1/\sqrt{n})$.

We first construct $\widehat{\pi}$ and $\widetilde{\pi}$ such that,

$$\sqrt{n}\|\widetilde{\pi} - \widehat{\pi}\|_\infty = o_{\mathbb{P}}(1), \quad \text{and} \quad \max_{r\in\mathcal{R}} \sqrt{n}\|r'(\widetilde{\pi} - \widehat{\pi})\|_\infty = o_{\mathbb{P}}(1), \qquad \text{(D.2)}$$

where with probability approaching one, $\widetilde{\pi} \in \times_{j=1}^{\dim(R)} C^2_M(\mathcal{D})$, for some M.

We construct $\widetilde{\pi}$ by smoothing $\widehat{\pi}$ component by component. Before smoothing, we extend the estimand π_{0j} outside $\mathcal{D}$, onto the ϵ-expansion $\mathcal{D}^\epsilon$ smoothly so that the extended function is of class C^3. This is possible by first extending the third derivative of π_{0j} smoothly and then integrating up to obtain lower order derivatives and the function. Then we extend $\widehat{\pi}_j$ to the outer region by simply setting $\widehat{\pi}_j(d) = \pi_{0j}(d)$ over $d \notin \mathcal{D}$. Note that the extended process $\sqrt{n}(\widehat{\pi}_j(d) - \pi_{0j}(d))$ remains to be stochastically equicontinuous, since it vanishes asymptotically at the boundary of $\mathcal{D}$. Then we define the smoothed version of $\widehat{\pi}_j$ as

$$\widetilde{\pi}_j(d) = \int_{\mathcal{D}^\epsilon} \widehat{\pi}_j(z)[K((z-d)/h)/h]dz, \quad d \in \mathcal{D},$$

where $0 \leq h \leq \epsilon$ is bandwidth such that $\sqrt{n}h^3 \to 0$ and $\sqrt{n}h^2 \to \infty$; $K : \mathbb{R} \to \mathbb{R}$ is a third order kernel with the properties: $\partial^\mu K$ are continuous on $[-1,1]$ and vanish outside of $[-1,1]$ for $\mu = 0,1,2$, $\int K(z)dz = 1$, and $\int z^\mu K(z)dz = 0$ for $\mu = 1,2$. Such kernel exists and can be obtained by reproducing kernel Hilbert space methods or via twicing kernel methods (Berlinet, 1993, and Newey, Hsieh, and Robins, 2004). We then have

$$\begin{aligned}\sqrt{n}(\widetilde{\pi}_j(d) - \widehat{\pi}_j(d)) &= \int_{\mathcal{D}^\epsilon} \sqrt{n}[\widehat{\pi}_j(z) - \pi_{0j}(z) - (\widehat{\pi}_j(d) - \pi_{0j}(d))][K([z-d]/h)/h]dz \\ &+ \int_{\mathcal{D}^\epsilon} \sqrt{n}(\pi_{0j}(z) - \pi_{0j}(d))[K([z-d]/h)/h]dz.\end{aligned}$$

The first term is bounded uniformly in $d \in \mathcal{D}$ by $\omega(2h)\|K\|_\infty \to_{\mathbb{P}} 0$ where

$$\omega(2h) = \sup_{|z-u|\leq 2h} |\sqrt{n}[\widehat{\pi}_j(z) - \pi_{0j}(z) - (\widehat{\pi}_j(d) - \pi_{0j}(d))]| \to_{\mathbb{P}} 0.$$

The second term is bounded uniformly in $d \in \mathcal{D}$, up to a constant, by

$$\sqrt{n}\|\partial^3\pi_{0j}\|_\infty h^3 \int \lambda^3 K(\lambda)d\lambda \lesssim \sqrt{n}h^3 \to 0.$$

This establishes the equivalence (D.2), in view of compactness of $\mathcal{R}$.

Next we show that $\|\partial^2\widetilde{\pi}_j\|_\infty \leq 2\|\partial^2\pi_{0j}\|_\infty := M$ with probability approaching 1. Note that

$$\partial^2\widetilde{\pi}_j(d) - \partial^2\pi_{0j}(d) = \int_{\mathcal{D}^\epsilon} \widehat{\pi}_j(z)[\partial^2 K([z-d]/h)/h^3]dz - \partial^2\pi_{0j}(d),$$

which can be decomposed into two pieces:

$$\begin{aligned}&n^{-1/2}h^{-2}\textstyle\int_{\mathcal{D}^\epsilon} n^{1/2}(\widehat{\pi}_j(z) - \pi_{0j}(z))[\partial^2 K([z-d]/h)/h]dz \\ &+ \textstyle\int_{\mathcal{D}^\epsilon}[\partial^2 K([z-d]/h)/h^3]\pi_{0j}(z)dz - \partial^2\pi_{0j}(d).\end{aligned}$$

The first piece is bounded uniformly in $d \in \mathcal{D}$ by $n^{-1/2}h^{-2}\omega(2h)\|\partial^2 K\|_\infty \to 0$, while, using the integration by parts, the second piece is equal to

$$\int_{\mathcal{D}^\epsilon}[\partial^2\pi_{0j}(z) - \partial^2\pi_{0j}(d)][K([z-d]/h)/h]dz,$$

which converges to zero uniformly in $d \in \mathcal{D}$ by the uniform continuity of $\partial^2\pi_{0j}$ on $\mathcal{D}^\epsilon$ and by boundedness of the kernel function. Thus $\|\partial^2\widetilde{\pi}_j - \partial^2\pi_{0j}\|_\infty \to_{\mathbb{P}} 0$, and similarly conclude that $\|\partial\widetilde{\pi}_j - \partial\pi_{0j}\|_\infty \to_{\mathbb{P}} 0$, where convergence is uniform in $1 \leq j \leq \dim(R)$, since $\dim(R)$ is finite and fixed.

We then construct $\widetilde{\vartheta}(r,d) = \Lambda(r'\widetilde{\pi}(d))$. Note that by the preceding arguments $\widetilde{\vartheta} \in \Upsilon$ for some M with probability approaching one. Finally, the first order equivalence $\sqrt{n}\|\widetilde{\vartheta} - \widehat{\vartheta}\|_\infty \to_{\mathbb{P}} 0$ follows immediately from (D.2), boundedness of $\|\partial\Lambda\|_\infty$ and compactness of $\mathcal{R}$. $\square$

Appendix E. Monte Carlo designs

For the homoskedastic design, we use the following parametric linear version of the system of equations (2.4)–(2.5) to generate the observations:

$$D = \pi_{00} + \pi_{01}Z + \pi_{02}W + \Phi^{-1}(V), \quad V \backsim U(0,1), \tag{E.1}$$

$$Y^* = \beta_{00} + \beta_{01}D + \beta_{02}W + \Phi^{-1}(\epsilon), \quad \epsilon \backsim U(0,1), \tag{E.2}$$

where Φ^{-1} denotes the quantile function of the standard normal distribution, and ($\Phi^{-1}(V)$, $\Phi^{-1}(\epsilon)$) is jointly normal with correlation ρ_0. Though we can observe Y^* in the simulated data, we artificially censor the data to observe

$$Y = \max(Y^*, C) = \max(\beta_{00} + \beta_{01}D + \beta_{02}W + \Phi^{-1}(\epsilon), C). \tag{E.3}$$

From the properties of the multivariate normal distribution, $\Phi^{-1}(\epsilon) = \rho_0\Phi^{-1}(V) + (1 - \rho_0^2)^{1/2}\Phi^{-1}(U)$, where $U \backsim U(0,1)$. Using this expression, we can combine (E.1) and (E.3) for an alternative formulation of the censored model in which the control term V_i is included in the equation for the observed response:

$$Y = \max(Y^*, C) = \max(\beta_{00} + \beta_{01}D + \beta_{02}W + \rho_0\Phi^{-1}(V) + (1-\rho_0^2)^{1/2}\Phi^{-1}(U), C).$$

This formulation is useful because it indicates that when we include the control term in the quantile function, its true coefficient is ρ_0.

In our simulated data, we create extreme endogeneity by setting $\rho_0 = .9$. We set $\pi_{00} = \beta_{00} = 0$, and $\pi_{01} = \pi_{02} = \beta_{01} = \beta_{02} = 1$. We draw the disturbances $[\Phi^{-1}(V), \Phi^{-1}(\epsilon)]$ from a bivariate normal distribution with zero means, unit variances and correlation ρ_0. We draw Z from a standard normal distribution, and we generate W to be a log-normal random variable that is censored from the right at its 95th percentile, q_W. Formally, we draw $\widetilde{W}$ from a standard normal distribution. We then calculate $q_W = Q_W(.95)$, which differs across replication samples. Next, we set $W = \min(e^{\widetilde{W}}, q_W)$. For comparative purposes, we set the amount of censoring in the dependent variable to be comparable to that in Kowalski (2009). Specifically, we set $C = C = Q_{Y^*}(.38)$ in each replication sample. We report results from 1,000 simulations with $n = 1,000$.

For the heteroskedastic design, we replace the first stage equation for D in (E.1) by the following equation:

$$D = \pi_{00} + \pi_{01} Z + \pi_{02} W + (\pi_{03} + \pi_{04} W)\Phi^{-1}(V), \quad V \backsim U(0,1) \tag{E.4}$$

where we set $\pi_{03} = \pi_{04} = 1$. The corresponding conditional quantile function is

$$Q_D(v \mid W, Z) = \pi_{00} + \pi_{01} Z + \pi_{02} W + (\pi_{03} + \pi_{04} W)\Phi^{-1}(v),$$

and can be consistently estimated by quantile regression or other estimator for location-scale shift models.

Appendix F. CQIV Technical Details and Robustness Diagnostic Test

For the OLS estimator of the control variable, we run an OLS first stage and retain the predicted residuals from the OLS first stage as the control variable. For the quantile estimator of the control variable, we run first stage quantile regressions at each quantile from .01 to .99 in increments of .01. Next, for each observation, we compute the fraction of the quantile estimates for which the predicted value of the endogenous variable is less than or equal to the true value of the endogenous variable. We then evaluate the standard normal quantile function at this value and retain the result as the estimate of the control variable. In this way, the quantile estimate of the control variable allows for heteroskedasticity in the first stage.

For the distribution regression estimator of the control variable, we first create a matrix $n * n$ of indicators, where n is the sample size. For each value of the endogenous variable in the data set y_j in columns, each row i gives if the log-expenditure of the individual i is less or equal than y_j $(1(y_i \leq y_j))$. Second, for each column j of the matrix of indicators, we run a probit regression of the column on the exogenous variables. Finally, the estimate of the control variable for the observation i is the quantile function of the standard normal evaluated at the predicted value for the probability of the observation $i = j$.

In Table B1, we present the CQIV robustness diagnostic tests suggested in section 3 for the CQIV estimator with an OLS estimate of the control variable. In our estimates, we used a probit model in the first step, and we set $q_0 = 10$ and $q_1 = 3$. In practice, we do not necessarily recommend reporting the diagnostics in Table B1, but we have included them here for expositional purposes. In the top section of the table, we present diagnostics computed after CQIV Step 1. At the 0.05 quantile, observations are retained in J_0 if their predicted probability of being uncensored exceeds $1-u+k_0 = 1-.05+.0445 = .9945$. Empirically, this leaves 47.0% of the total sample in J_0 in the median replication sample. In all statistics, the variation across replication samples appears small. However, as intended by the algorithm, there is meaningful variation across the estimated quantiles. As the estimated quantile increases, the percentage of observations retained in J_0 increases. From these diagnostics, the CQIV estimator appears well-behaved in the sense that the percentage of observations retained in J_0 is never very close to 0 or 100.

In the second section of Table B1, we present robustness test diagnostics computed after CQIV Step 2. Observations are retained in J_1 if the predicted Y_i exceeds $C_i + \varsigma_1$, where the median value of C_i, as shown in the table, is 1.60, and the median value of ς_1 at the .05 quantile is 1.70. As desired, at each quantile, the percentage of observations retained in J_1 is smaller than the percentage of observations with predicted values above C_i but larger than the percentage of observations retained in J_0. As shown in sections of the table labeled "Percent J_0 in J_1" and "Count J_1 not in J_0" J_0 is almost a proper subset of J_1.

In the last section of Table B1, we report the value of the Powell objective function obtained after CQIV Step 2 and CQIV Step 3. The last column shows that on average the final CQIV step represents an improvement in the objective function in 36-51% of replication samples across the estimated quantiles. In our CQIV simulation results, we report the results from the third step. Researchers might prefer to select select results from the second or third step based on the value of the objective function.

References

[1] Andrews, Donald W. K. "Asymptotics for semiparametric econometric models via stochastic equicontinuity. " Econometrica. 1994. 62 no. 1. pp 43-72.

[2] Angrist, Joshua D., Imbens, Guido W., and Rubin, Donald B. "Identification of Causal Effects Using Instrumental Variables." Journal of the American Statistical Association. 1996. 91. pp 444-455.

[3] Banks, James, Blundell, Richard, and Lewbel, Arthur. "Quadratic Engel Curves and Consumer Demand." Review of Economics and Statistics. 1997. 79(4). pp 527-539.

[4] Berlinet, Alain. "Hierarchies of higher order kernels." Probability Theory and Related Fields. 1993. 94(4). pp 489-504.

[5] Blundell, Richard, Browning, Martin, and Crawford, Ian. "Nonparametric Engel Curves and Revealed Preference." Econometrica. 2003. 71(1). pp 205-240.

[6] Blundell, Richard, Chen, Xiaohong, and Kristensen, Dennis. "Semi-nonparametric IV Estimation of Shape-Invariant Engel Curves." Econometrica. 2007. 75(6). pp. 1613-1669.

[7] Blundell, Richard, and Matzkin, Rosa. "Conditions for the Existence of Control Functions in Nonseparable Simultaneous Equations Models." CEMMAP Working Paper 28/10.

[8] Blundell, Richard, Duncan, Alan, and Pendakur, Krishna. "Semiparametric Estimation and Consumer Demand." Journal of Applied Econometrics. 1998. 13(5). pp. 435-461.

[9] Blundell, Richard, and Powell, James. "Censored Regression Quantiles with Endogenous Regressors." Journal of Econometrics. 2007. 141. pp. 65-83.

[10] Chamberlain, Gary, and Imbens, Guido. "Nonparametric applications of Bayesian inference." Journal of Business and Economic Statistics. 2003. 21(1). pp. 12-18.

[11] Chen, Xiaohong, and Pouzo, Demian. "Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals." Journal of Econometrics. 2009. 152(1). pp. 46-60.

[12] Chernozhukov, Victor, Fernandez-Val, Ivan, and Galichon, Alfred. "Quantile and Probability Curves without Crossing." Econometrica. 2010. 78(3, May) 1093-1125.

[13] Chernozhukov, Victor, Fernandez-Val, Ivan, and Melly, Blaise. "Inference on Counterfactual Distributions." MIT Department of Economics Working Paper. 08-16. 2009.

[14] Chernozhukov, Victor, and Hansen, Christian. "Instrumental variable quantile regression: A robust inference approach." Journal of Econometrics. January 2008. 142(1). pp.379-398.

[15] Chernozhukov, Victor, and Hong, Han. "Three-Step Quantile Regression and Extramarital Affairs." Journal of The American Statistical Association. September 2002. 97(459). pp. 872-882.

[16] Chesher, A. "Identification in Nonseparable Models." Econometrica, 2003, 71(5), pp. 1405-1441.

[17] Deaton, Angus and Muelbauer, John. Economics and consumer behavior. Cambridge University Press. 1980.

[18] Foresi, Silverio and Peracchi, Franco. "The Conditional Distribution of Excess Returns: An Empirical Analysis." Journal of the American Statistical Association, 1995, 90(430), pp.451-466.

[19] Frisch, R. "Circulation Planning: : Proposal For a National Organization of a Commodity and Service Exchange." Econometrica. 1934. 2(3), 258-336.

[20] Gorman, W.M. "Separable Utility and Aggregation." Econometrica. 1959 27(3), 469-481.

[21] Hausman, Jerry A."Specification Tests in Econometrics." Econometrica. 1978. 46(6). pp. 1251-71.

[22] Hausman, Jerry, Newey, Whitney, and Powell, James. "Nonlinear Errors in Variables Estimation of Some Engel Curves." Journal of Econometrics. 1995. 65. pp. 203-233.

[23] Heckman, James J. "Sample Selection Bias as a Specification Error" Econometrica. 1979. 47(1), 153-161.

[24] Imbens, Guido W., and Newey, Whitney K.. "Identification and Estimation of Triangular Simultaneous Equations Models without Additivity."NBER Technical Working Paper 285. 2002.

[25] Imbens, Guido W., and Newey, Whitney K.. "Identification and Estimation of Triangular Simultaneous Equations Models without Additivity."Econometrica. 2009. 77(5) 1481-1512.

[26] Jun, Sung Jae. "Local structural quantile effects in a model with a nonseparable control variable."Journal of Econometrics. 2009. 151(1) 82-97.

[27] Koenker, Roger. Quantile Regression. Cambridge University Press. 2005.

[28] Koenker, Roger, and Bassett, Gilbert Jr. "Regression Quantiles." Econometrica, 1978, 46(1), pp. 33-50.

[29] Koopmans, T.C. and Hood, W.C.. "The estimation of simultaneous linear economic relationships." W.C. Hood and T.C. Koopmans, Editors, Studies in econometric method, Wiley, New York (1953).

[30] Kowalski, Amanda E. "Censored Quantile Instrumental Variable Estimates of the Price Elasticity of Expenditure on Medical Care." NBER Working Paper 15085. 2009.

[31] Lee, Sokbae. "Endogeneity in quantile regression models: A control function approach." Journal of Econometrics. 2007. 141, pp. 1131-1158.

[32] Lewbel, Arthur. "Entry for the New Palgrave Dictionary of Economics, 2nd Edition. " Boston College. 2006.

[33] Ma, Lingjie and Koenker, Roger. "Quantile regression methods for recursive structural equation models." Journal of Econometrics. 2006. 134(2). pp. 471-506.

[34] Ma, Shuangge, and Kosorok, Michael. "Robust semiparametric M-estimation and the weighted bootstrap." Journal of Multivariate Analysis. 2005. 96(1). pp. 190-217.

[35] Matzkin, Rosa L. "Nonparametric Identification."In *Handbook of Econometrics*, Vol. 6B, ed. by J. Heckman and E. Leamer. 2007. Amsterdam : Elsevier.

[36] Newey, Whitney K. "Efficient Estimation of Limited Dependent Variable Models with Endogenous Explanatory Variables." Journal of Econometrics, 1987, 36, pp. 231-250.

[37] Newey, Whitney K. "The asymptotic variance of semiparametric estimators. " Econometrica, 1994. 62 no. 6. 1349-1382.

[38] Newey, Whitney K., Hsieh, Fushing, Robins, James M. "Twicing kernels and a small bias property of semiparametric estimators. " Econometrica. 2004. 72(3). pp 947-962.

[39] Newey, Whitney K., Powell, James L., Vella, Francis. "Nonparametric Estimation of Triangular Simultaneous Equations Models." Econometrica. 1999. 67(3), 565-603.

[40] Powell, James L. "Censored Regression Quantiles." Journal of Econometrics, 1986. 23. pp-143-155.

[41] Powell, James L. "Least absolute deviations estimation for the censored regression model." Journal of Econometrics, 1984, 25(3), pp. 303-325.

[42] Powell, James L. "Chapter 14: Estimation of Monotonic Regression Models under Quantile Restrictions." Nonparametric and Semiparametric Methods in Econometrics and Statistics: Proceedings of the Fifth International Symposium in Economic Theory and Econometrics. 1991

[43] Smith, Richard J. and Blundell, Richard W. "An Exogeneity Test for a Simultaneous Equation Tobit Model with an Application to Labor Supply." Econometrica, 1986, 54(3), pp. 679-685.

[44] van der Vaart. Asymptotic Statistics. Cambridge University Press. 1998.

[45] van der Vaart, A.W. and Wellner, Jon A. Weak convergence and empirical processes. Springer. 1996.

[46] Wooldridge, Jeffrey M. Econometric Analysis of Cross Section and Panel Data. MIT Press. Cambridge, MA. Second Edition. 2010.

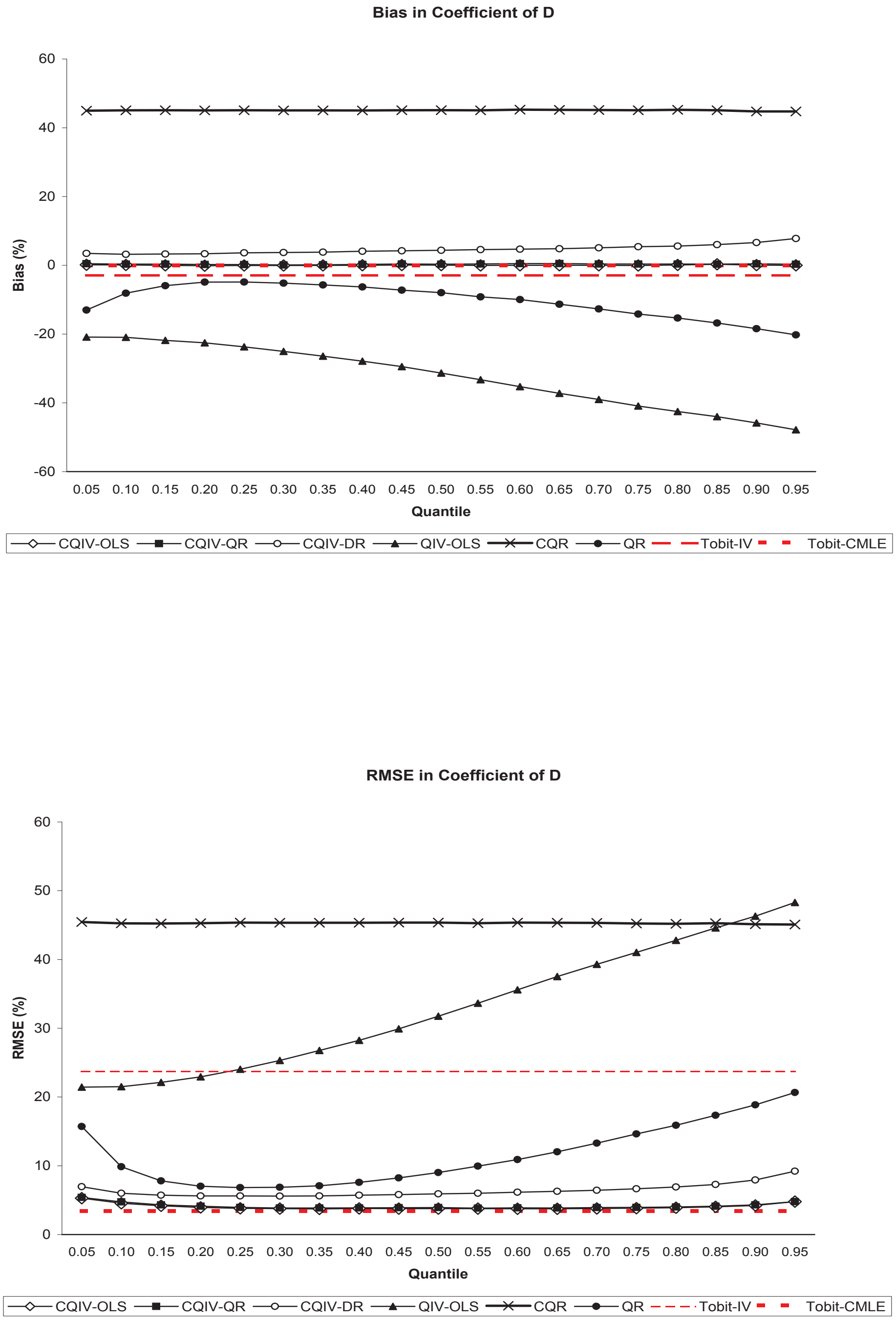

Figure 1: Homoskedatic design: Mean bias and RMSE of Tobit and QR estimators. Results obtained from 1,000 samples of size $n = 1,000$.

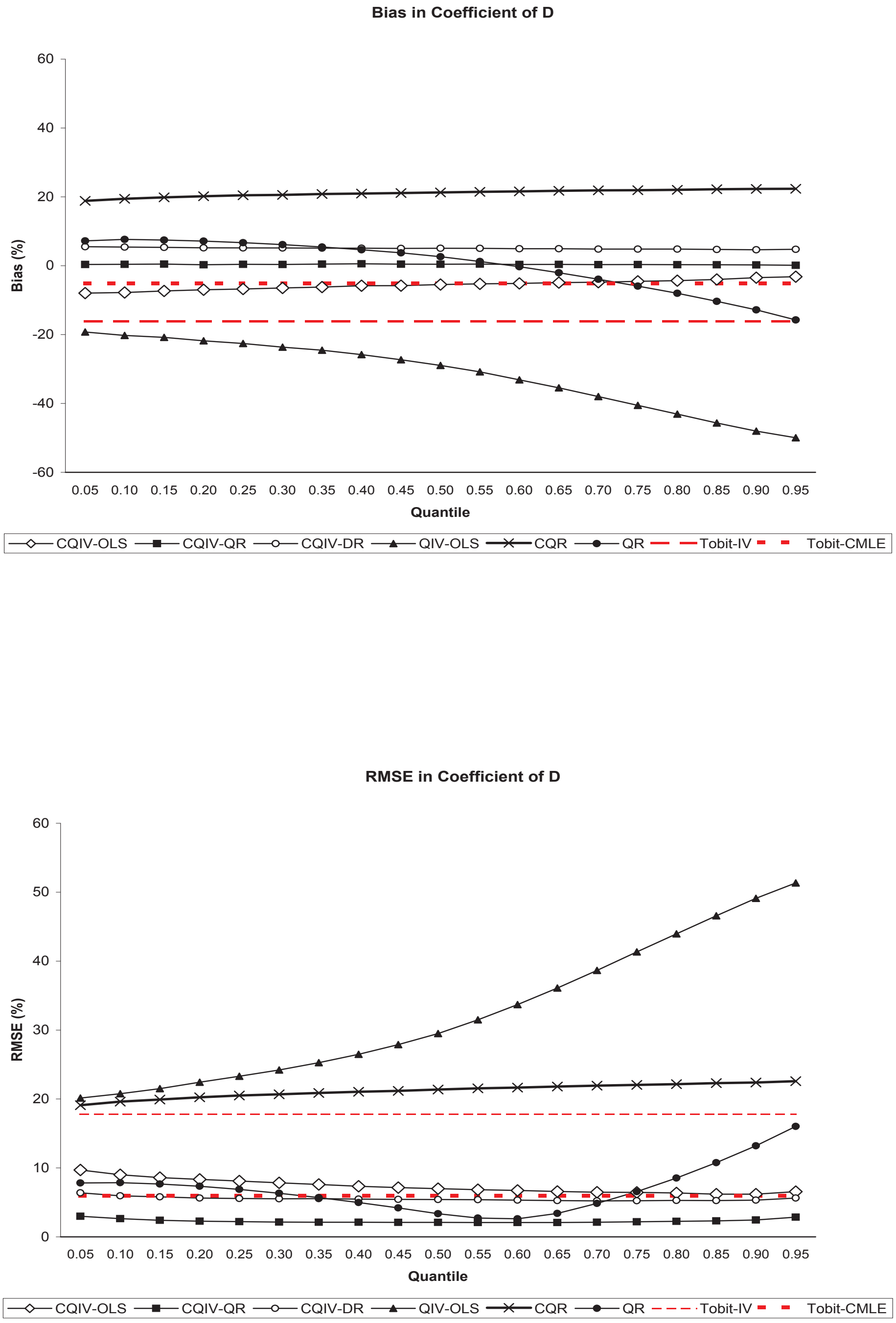

Figure 2: Heteroskedastic design: Mean bias and RMSE of Tobit and QR estimators. Results obtained from 1,000 samples of size $n = 1,000$.

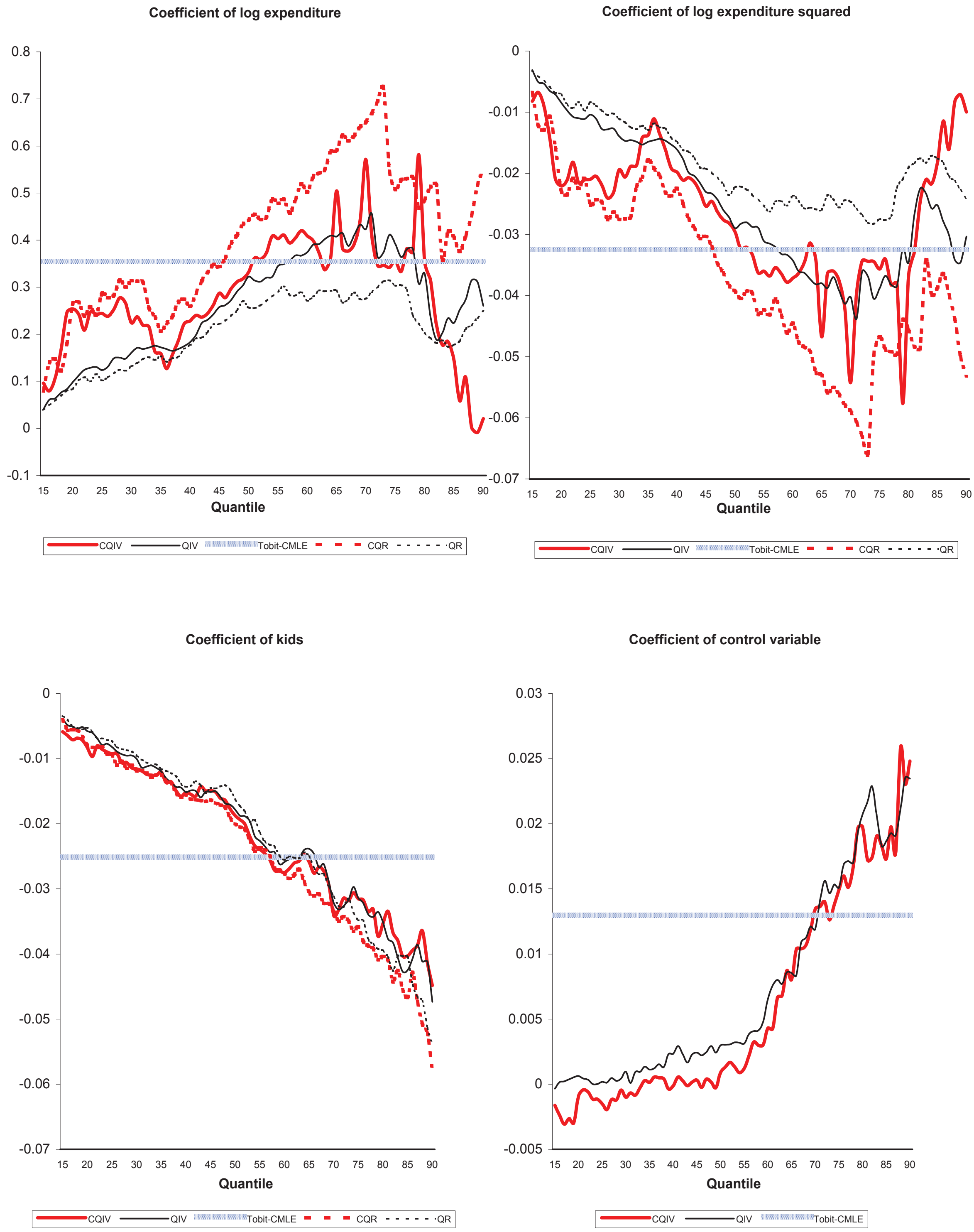

Figure 3: Coefficients of Engel Curves

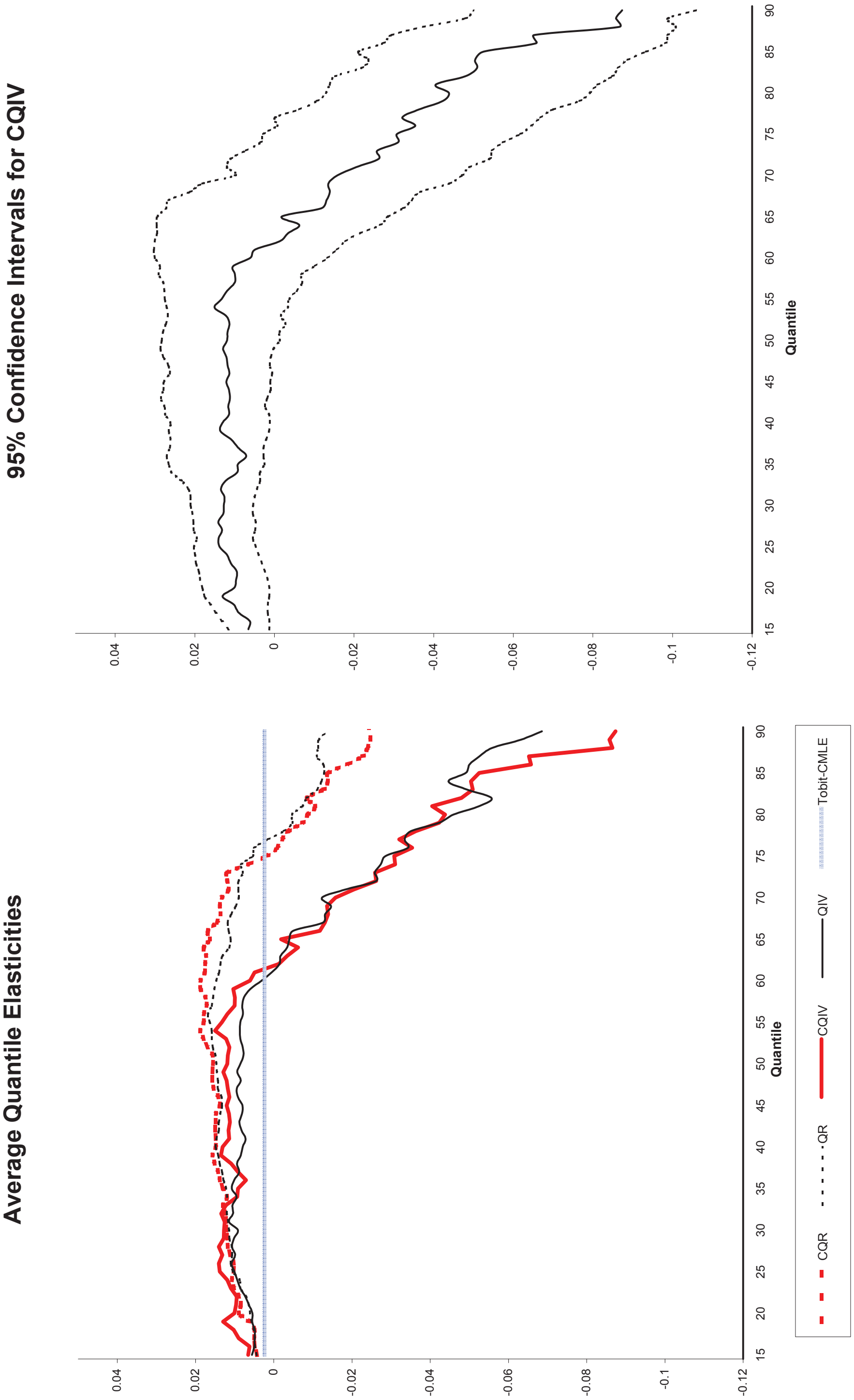

Figure 4: Estimates and 95% pointwise confidence intervals for average quantile expenditure elasticities. The intervals are obtained by weighted bootstrap with 200 replications and exponentially distributed weights.

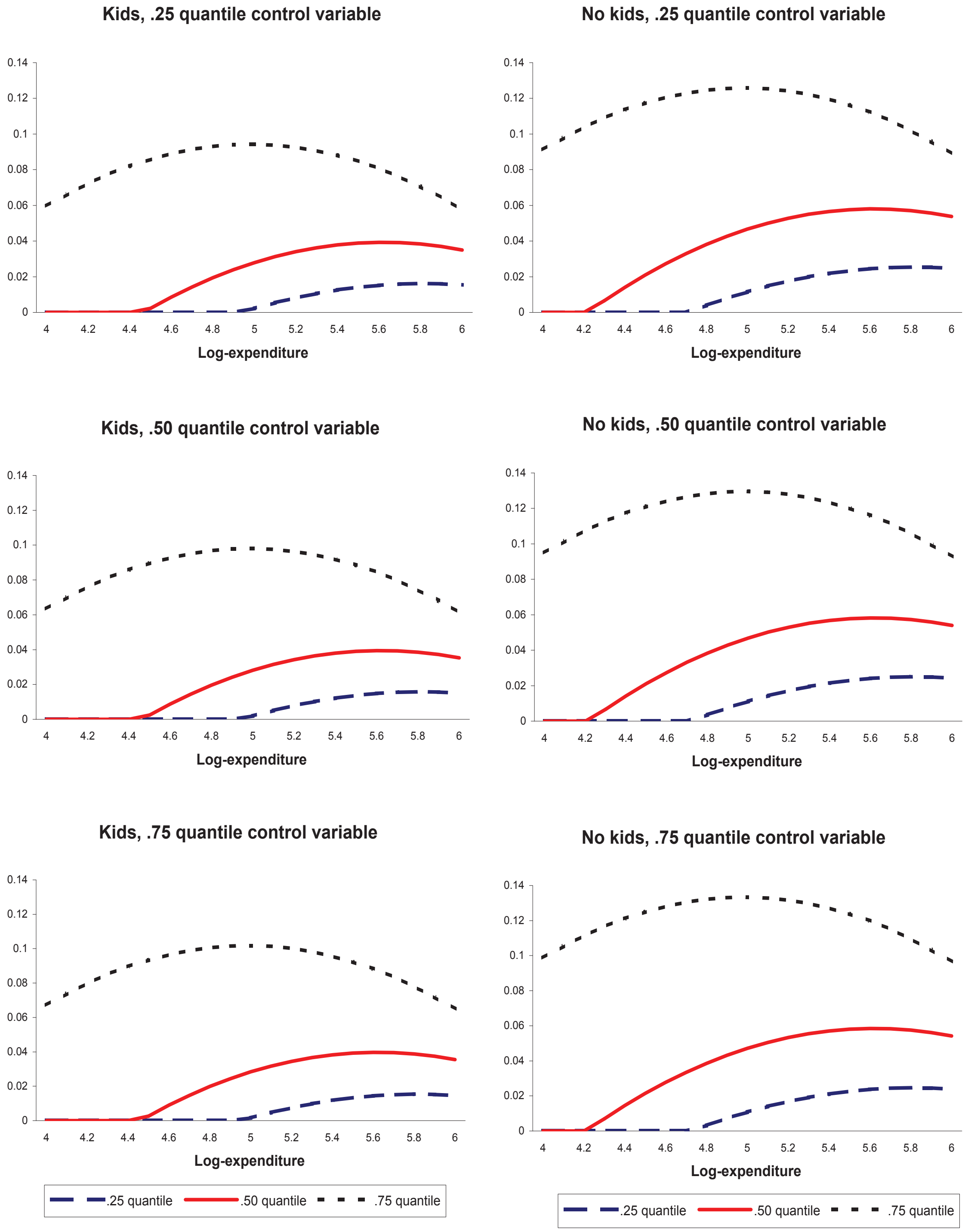

Figure 5: Family of Engel curves: each panel plots Engel curves for the three quantiles of alcohol share.

Table B1: CQIV Robustness Diagnostic Test Results
for CQIV with OLS Estimate of the Control Variable - Homoskedastic Design

CQIV-OLS Step 1

		k_0			Percent J0	
Quantile	Median	Min	Max	Median	Min	Max
0.05	0.04	0.04	0.05	47.20	43.30	50.30
0.1	0.09	0.06	0.10	49.10	46.00	51.30
0.25	0.20	0.15	0.24	52.20	50.50	53.70
0.5	0.36	0.26	0.46	55.80	54.80	56.80
0.75	0.43	0.29	0.58	59.40	57.70	61.10
0.9	0.37	0.22	0.58	62.40	60.30	65.10
0.95	0.30	0.18	0.54	64.20	61.40	67.50

CQIV-OLS Step 2

		ς_1			Percent J1			Percent Predicted Above C	
Quantile	Median	Min	Max	Median	Min	Max	Median	Min	Max
0.05	1.70	1.45	2.01	50.70	46.70	54.90	52.30	48.20	56.70
0.1	1.71	1.44	1.96	52.80	49.50	55.50	54.50	51.10	57.30
0.25	1.71	1.46	1.98	56.30	53.60	58.70	58.10	55.30	60.60
0.5	1.72	1.44	2.02	60.10	57.60	63.40	62.00	59.40	65.40
0.75	1.73	1.47	1.99	64.00	61.20	66.80	66.00	63.10	68.90
0.9	1.75	1.44	2.01	67.40	64.60	70.60	69.50	66.60	72.80
0.95	1.76	1.49	2.02	69.30	65.60	72.80	71.50	67.70	75.10
		C			Percent J0 in J1			Count in J1 not in J0	
Quantile	Median	Min	Max	Median	Min	Max	Median	Min	Max
0.05	1.60	1.33	1.85	100	97.7	100	36	0	81
0.1	1.60	1.33	1.85	100	99.0	100	37	7	74
0.25	1.60	1.33	1.85	100	99.6	100	40	15	68
0.5	1.60	1.33	1.85	100	99.6	100	43	23	78
0.75	1.60	1.33	1.85	100	99.7	100	47	17	74
0.9	1.60	1.33	1.85	100	99.7	100	50	15	88
0.95	1.60	1.33	1.85	100	99.1	100	51	16	97

Comparison of Objective Functions

		Objective Step 3			Objective Step 2		Objective Step 3<Objective Step 2	
Quantile	Median	Min	Max	Median	Min	Max	Median	Mean
0.05	5058	4458	5674	5054	4400	5753	0	0.44
0.1	8939	7925	9946	8927	7888	10049	0	0.47
0.25	17292	15100	19839	17271	14741	20052	0	0.44
0.5	22859	18692	27022	22837	18306	27091	0	0.45
0.75	16073	9603	22872	15895	8737	22866	0	0.42
0.9	-1016	-9624	7150	-1047	-10834	9265	0	0.45
0.95	-13815	-24602	-2884	-14034	-27816	-1919	0	0.44

N=1,000, Replications=1,000