

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 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. Our analysis covers two-stage (uncensored) quantile regression with nonadditive first stage as an important special case. We illustrate the computation and applicability of the CQIV estimator with a Monte-Carlo numerical example and an empirical application on estimation of Engel curves for alcohol.

Date: March 13, 2014. 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. We thank the editor Cheng Hsiao, two referees, and seminar participants at EIEF, Georgetown, Rochester, and Penn State for useful comments. 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 censored due to top-coding 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 relevant omitted variables. Censoring and endogeneity often come together in economic applications. For example, both of them arise in the estimation of Engel curves for alcohol – the relationship between the share of expenditure on alcohol and the household’s budget. For this commodity, a significant fraction of households report zero expenditure, and economic theory suggests that the 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 unobservables. 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 heterogeneous effects across the distribution 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 unobservables become independent. This so-called control variable is usually unobservable and needs to be estimated in a first stage. Our main contribution here is to allow for semiparametric models with infinite dimensional parameters and nonadditive unobservables, 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. This technique applies to semiparametric two-stage estimators where the two stages can be nonadditive in the unobservables. CQIV is an example where the first stage estimates a nonadditive quantile or distribution regression model for the control variable, whereas the second stage estimates a nonadditive censored quantile regression model, including the estimated control variable to deal with endogeneity. Two-stage (uncensored) quantile regression with distribution or quantile regression in the first stage is an important special case of CQIV.

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 the proof of its consistency is not overly complex following the strategy set forth by Ma and Kosorok (2005). 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 two-stage censored and uncensored quantile regression estimators based on 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.¹ 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 in data that do not satisfy these assumptions. 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,

¹We have developed a Stata command to implement the methods developed in this paper (see Chernozhukov, Fernandez-Val, Han, and Kowalski 2011). It is available at <http://EconPapers.repec.org/RePEc:boc:bocode:s457478>.

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 unobservables both in the first and the second stage. Blundell and Powell (2004) and Rothe (2009) applied the control variable approach to semiparametric binary response models. 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 unobservables 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 methods have the advantage that they impose less structure in the model at the cost of slower rates of convergence in estimation. In particular, they leave the dependence on the control variable unspecified, whereas CQIV uses a flexible parametric specification. 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 Blundell and Powell (2003) and Matzkin (2007) for excellent comprehensive reviews of results on semi and nonparametric identification and estimation 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 contrast, our approach is specifically designed to handle the latter, and in doing so, it puts the first stage and second stage models on 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 because of the difficulties discussed above.

For models with censoring and exogenous regressors, Powell (1986), Fitzenberger (1997), Buchinsky and Hahn (1998), Khan and Powell (2001), Chernozhukov and Hong (2002), Honoré, Khan and Powell (2003), and Portnoy (2003) developed quantile regression methods. The literature on models combining both endogeneity and censoring is more sparse. Smith and Blundell (1986) pioneered the use of the control variable approach to estimate a triangular parametric additive location model. More recently, Blundell and Powell (2007) proposed an alternative censored quantile instrumental variable estimator building on Chen

and Khan (2001). Compared to our estimator, Blundell and Powell estimator assumes additive unobservables in the first and second stages, but permits a flexible local nonparametric endogeneity correction in the second stage. Hong and Tamer (2003) and Khan and Tamer (2006) also considered censored regression models with endogenous regressors. They dealt with endogeneity with an instrumental variable quantile approach that is not nested with the control variable approach used here; see Blundell and Powell (2003) for a comparison of these two approaches. They dealt with censoring using a more flexible moment inequality formulation that allows for endogenous censoring and partial identification, but that leads to a more complicated estimator. A referee has pointed to us the possibility of applying the control variable approach as pursued in this paper to Buchinsky and Hahn (1998) estimator to deal with endogenous regressors. We believe that this is indeed possible using the results of this paper, though we leave formal developments to future work.

Relative to the previous literature, the paper makes three main contributions. First, it develops a two-stage quantile regression estimator for a triangular nonseparable model where the first stage is nonadditive in the unobservables. Our analysis here builds on Chernozhukov, Fernandez-Val, and Galichon (2010) and Chernozhukov, Fernandez-Val, and Melly (2013), which established the properties of the quantile and distribution regression estimators that we use in the first stage. The theory for the second stage estimator, however, does not follow from these results using standard techniques due to the dimensionality and entropy properties of the first stage estimators. Second, it adapts the two-stage quantile regression estimator to models with censoring by extending Chernozhukov and Hong (2002) algorithm to the presence of a generated regressor (control variable). Third, it establishes the validity of weighted bootstrap for two-stage censored and uncensored quantile regression estimators where the first stage is estimated by quantile or distribution regression.

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 Sections 3 and 4, we describe the associated computational algorithms and present results from a Monte-Carlo simulation exercise. In Section 5, we present an empirical application of CQIV to Engel curves. In Section 6, we provide conclusions and discuss potential empirical applications of CQIV. The proofs of the main results are given in the appendix.

2. CENSORED QUANTILE INSTRUMENTAL VARIABLE REGRESSION

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

$$Y = \max(Y^*, C), \quad (2.1)$$

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

$$D = Q_D(V \mid W, Z). \quad (2.3)$$

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).² The uncensored case is covered by making C arbitrarily small.

The function $u \mapsto Q_{Y^*}(u \mid D, W, V)$ 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. The main identification conditions are the exclusion restriction of Z in equation (2.2), and the relevance condition of Z in equation (2.3). These conditions permit V to have independent variation of D and W .

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

$$Y^* = \beta_1 D + \beta_2' W + (\beta_3 D + \beta_4' W)\epsilon, \quad (2.4)$$

$$D = \pi_1' Z + \pi_2' W + (\pi_3' Z + \pi_4' W)\eta, \quad (2.5)$$

²We 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 $\tilde{Y} = -Y$, $\tilde{Y}^* = -Y^*$, $\tilde{C} = -C$, and $Q_{\tilde{Y}^*} = -Q_{Y^*}$, because $\tilde{Y} = \max(\tilde{Y}^*, \tilde{C})$.

where (ϵ, η) are jointly standard bivariate normal with correlation $\rho'W$ conditional on (W, Z, C) , $(\beta_3 D + \beta_4' W) > 0$ a.s., and $(\pi_3' Z + \pi_4' W) > 0$ a.s. By the properties of the normal distribution, $\eta = \Phi^{-1}(V)$ with $V \sim U(0, 1)$ independent of (W, Z, C) , and $\epsilon = (\rho'W)\Phi^{-1}(V) + [1 - (\rho'W)^2]^{1/2}\Phi^{-1}(U)$ with $U \sim U(0, 1)$ independent of (W, Z, C, V, D) , where Φ^{-1} is the inverse distribution function of the standard normal. The corresponding conditional quantile functions have the form of (2.2) and (2.3) with

$$\begin{aligned} Q_{Y^*}(U \mid D, W, V) &= \beta_1 D + \beta_2' W + (\beta_3 D + \beta_4' W)\{(\rho'W)\Phi^{-1}(V) + [1 - (\rho'W)^2]^{1/2}\Phi^{-1}(U)\}, \\ Q_D(V \mid W, Z) &= \pi_1' Z + \pi_2' W + (\pi_3' Z + \pi_4' W)\Phi^{-1}(V). \end{aligned}$$

Both of these quantile functions are nonadditive in U and V . We use a simplified version of the system (2.4)–(2.5) to generate the data for the numerical examples in Section 4.

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

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

by the equivariance property of the quantiles to monotone transformations. Whether the response of interest is the latent or observed variable depends on the source of censoring (e.g., Wooldridge, 2010, Chap. 17). 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

$$\partial_d Q_{Y^*}(u \mid d, w, v) = \beta_1 + \beta_3\{(\rho'w)\Phi^{-1}(v) + [1 - (\rho'w)^2]^{1/2}\Phi^{-1}(u)\},$$

which corresponds to the ceteris paribus effect of a marginal change of D on the latent response Y^* for individuals with $(D, W, V, U) = (d, w, v, u)$. When the censoring is due to economic or behavioral reasons such as 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, the marginal effect of the endogenous regressor D evaluated at $(D, W, V, U, C) = (d, w, v, u, c)$ is

$$\partial_d Q_Y(u \mid d, w, v, c) = 1\{Q_{Y^*}(u \mid d, w, v) > c\}\partial_d Q_{Y^*}(u \mid d, w, v),$$

which corresponds to the ceteris paribus effect of a marginal change of D on the observed response Y for individuals with $(D, W, V, C, U) = (d, w, v, c, u)$. 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

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

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. For the control variable, it is convenient to take a strictly monotonic transformation to adjust the location and scale (Newey, 2009), and to include interactions with the basis of D and W to account for nonseparabilities.³ 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

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

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

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

where $\rho_u(z) = (u - 1(z < 0))z$ is the asymmetric absolute loss function of Koenker and Bassett (1978), $\hat{X}_i = x(D_i, W_i, \hat{V}_i)$, $\hat{S}_i = s(\hat{X}_i, C_i)$, $s(X, C)$ is a vector of transformations of (X, C) , $\varsigma(u)$ is a positive cut-off, \hat{V}_i is an estimator of V_i , and T_i is an exogenous trimming indicator defined in Assumption 2 that we include for technical reasons. The estimator in (2.9) adapts the algorithm for the CQR estimator developed in Chernozhukov and Hong (2002) to deal with endogeneity. This algorithm is based on the property of the model

$$P(Y \leq X'\beta_0(u) \mid X, C, X'\beta_0(u) > C) = P(Y^* \leq X'\beta_0(u) \mid X, C, X'\beta_0(u) > C) = u,$$

provided that $P(X'\beta_0(u) > C) > 0$. In other words, $X'\beta_0(u)$ is the conditional u -quantile of the observed outcome for the observations for which $X'\beta_0(u) > C$, i.e., the conditional u -quantile of the latent outcome is above the censoring point. These observations change with the quantile index u and may include censored observations. We refer to them as the “ u -quantile uncensored” observations. The multiplier $1(\hat{S}_i' \hat{\gamma}(u) \geq \varsigma(u))$ is a selector that predicts if observation i is u -quantile uncensored. We formally state the conditions on this selector in Assumption 5. The estimator in (2.9) may also be seen as a computationally

³For example, the transformation $\Phi^{-1}(V)$, where Φ is the distribution function of the standard normal, yields the control variable in the system (2.4)–(2.5).

attractive approximation to Powell estimator applied to our case:

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

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 u -quantile uncensored observations by estimating the conditional probabilities of censoring using a flexible binary choice model. These observations have conditional probability of censoring lower than the quantile index u because of the equivalence of the events $\{X' \beta_0(u) > C\} \equiv \{P(Y^* \leq C \mid X, C) < u\}$. We estimate the linear part of the conditional quantile function, $X_i' \beta_0(u)$, on the sample of u -quantile uncensored observations by standard quantile regression. Then, we update the set of u -quantile uncensored observations by selecting those observations with conditional quantile estimates that are above their censoring points, $X_i' \widehat{\beta}(u) > C_i$, 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 two equivalent representations:

$$V = \vartheta_0(D, W, Z) \equiv F_D(D \mid W, Z) \equiv Q_D^{-1}(D \mid W, Z). \quad (2.10)$$

For any estimator of $F_D(D \mid W, Z)$ or $Q_D^{-1}(V \mid W, Z)$, denoted by $\widehat{F}_D(D \mid W, Z)$ or $\widehat{Q}_D^{-1}(V \mid W, Z)$, based on any parametric or semiparametric functional form, the resulting estimator for the control variable is

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

Here we consider several examples: in the classical additive location model, $Q_D(v \mid W, Z) = R' \pi_0 + Q_V(v)$, 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 (2013) developed asymptotic theory for this estimator. If $D \mid W, Z \sim N(R' \pi_0, \sigma^2)$, the control variable has the parametric form $V = \Phi^{-1}([D - R' \pi_0]/\sigma)$, where Φ denotes the distribution 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 nonadditive 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

$$\widehat{V} = \tau + \int_{(\tau, 1-\tau)} 1\{R'\widehat{\pi}(v) \leq D\}dv, \quad (2.11)$$

where $\widehat{\pi}(v)$ is the Koenker and Bassett (1978) quantile regression estimator, τ is small positive trimming cut-off that avoids estimation of tail quantiles (Koenker, 2005, p. 148), and the integral can be approximated numerically using a finite grid of quantiles. The use of the integral representation of Q_D^{-1} with respect to Q_D is convenient to avoid potential noninvertibility of \widehat{Q}_D caused by nonmonotonicity of $v \mapsto R'\widehat{\pi}(v)$. 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

$$\widehat{V} = \Lambda(R'\widehat{\pi}(D)), \quad (2.12)$$

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, 2013). Chernozhukov, Fernandez-Val and Melly (2013) developed asymptotic theory for this estimator.

The classical additive location model is an special case of the quantile regression model, where only the coefficient of the intercept varies across quantiles. The quantile and distribution regression models are flexible in the sense that by allowing for a sufficiently rich R , we can approximate any conditional distributions arbitrarily well. These models are not nested, so they cannot be ranked on the basis of generality. We refer to Chernozhukov, Fernandez-Val and Melly (2013) for a detailed comparison of these models.

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, \dots, 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 letters such as \mathcal{Y} and \mathcal{X} denote the closures of the supports of Y and X ; and \mathcal{YX} denotes the closure of 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 observe $\{Y_i, D_i, W_i, Z_i, C_i\}_{i=1}^n$, 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*

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

We define a trimming indicator that equals one whenever D lies in a region that exclude extreme values. The purpose of the trimming is to avoid the far tails in the modeling and estimation of the control variable in the first stage. We consider a fixed trimming rule, which greatly simplifies the derivation of the asymptotic properties. Alternative random, data driven rules are also possible at the cost of more complicated proofs. We discuss the choice of the trimming rule in Section 3.

Assumption 2 (Trimming indicator). *The tail trimming indicator has the form*

$$T = 1(D \in \overline{\mathcal{D}}),$$

where $\overline{\mathcal{D}} = [\underline{d}, \overline{d}]$ for some $-\infty < \underline{d} < \overline{d} < \infty$, such that $P(T = 1) > 0$.

Throughout the paper we use bars to denote trimmed supports with respect to D , e.g., $\overline{\mathcal{DWZ}} = \{(d, w, z) \in \mathcal{DWZ} : d \in \overline{\mathcal{D}}\}$, and $\overline{\mathcal{V}} = \{\vartheta_0(d, w, z) : (d, w, z) \in \overline{\mathcal{DWZ}}\}$. The next assumption imposes compactness and smoothness conditions. Compactness is imposed over the trimmed supports and can be relaxed at the cost of more complicated and cumbersome proofs. Moreover, we do not require compactness of the support of Y , which is important to cover the tobit model. The smoothness conditions are fairly tight.

Assumption 3 (Compactness and smoothness). *(a) The set $\overline{\mathcal{DWZCX}}$ 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 $(w, z) \in \overline{\mathcal{WZ}}$. (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 $(x, c) \in \overline{\mathcal{XC}}$, and bounded above by a constant uniformly in $(x, c) \in \overline{\mathcal{XC}}$. (d) The derivative vector $\partial_v x(d, w, v)$ exists and its components are uniformly continuous in $v \in \overline{\mathcal{V}}$ uniformly in $(d, w) \in \overline{\mathcal{DW}}$, and are bounded in absolute value by a constant, uniformly in $(d, w, v) \in \overline{\mathcal{DWV}}$.*

The following assumption is a high-level condition on the function-valued estimator of the control variable. We assume that it has a uniform asymptotic functional linear representation. The trimming device facilitates this assumption because it limits the convergence to a region that excludes the tails of the control variable. Moreover, the function-valued

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. Let $T(d) := 1(d \in \overline{\mathcal{D}})$ and $\|f\|_{T,\infty} := \sup_{a \in \mathcal{A}} |T(d)f(a)|$ for any function $f : \mathcal{A} \mapsto \mathbb{R}$.

Assumption 4 (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 $\overline{\mathcal{D}\mathcal{W}\mathcal{Z}}$, (a)*

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

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

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

with probability approaching one, where the covering entropy of the function class Υ is not too high, namely

$$\log N(\epsilon, \Upsilon, \|\cdot\|_{T,\infty}) \lesssim 1/(\epsilon \log^4(1/\epsilon)), \quad \text{for all } 0 < \epsilon < 1.$$

The following assumptions are on the u -quantile uncensored 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 5 (Quantile-uncensored selector). (a) *The selection rule has the form*

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

for some $\varsigma(u) > 0$, where $\widehat{\gamma}(u) \rightarrow_{\mathbb{P}} \gamma_0(u)$ and, for some $\epsilon' > 0$,

$$1[S'\gamma_0(u) \geq \varsigma(u)/2]T \leq 1[X'\beta_0(u) \geq C + \epsilon']T \leq 1[X'\beta_0(u) > C]T \text{ } P\text{-a.e.},$$

where $S = s(X, C)$. (b) *The set $\overline{\mathcal{S}}$ 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 $\gamma \in \Gamma(u)$ and in $\vartheta \in \Upsilon$, where $\Gamma(u)$ is an open neighborhood of $\gamma_0(u)$ and Υ is defined in Assumption 4.* (d) *The components of the derivative vector $\partial_v s(x(d, w, v), c)$ are uniformly continuous at each $v \in \overline{\mathcal{V}}$ uniformly in $(d, w, c) \in \overline{\mathcal{D}\mathcal{W}\mathcal{C}}$, and are bounded in absolute value by a constant, uniformly in $(d, w, v, c) \in \overline{\mathcal{D}\mathcal{W}\mathcal{V}\mathcal{C}}$.*

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 6 (Identification and nondegeneracy). (a) *The matrix*

$$J(u) := E_P[f_Y(X'\beta_0(u) \mid X, C)XX' 1(S'\gamma_0(u) \geq \varsigma(u))T]$$

is of full rank. (b) *The matrix*

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

is finite and is of full rank, where

$$f(A) := \{1(Y < X'\beta_0(u)) - u\}X1(S'\gamma_0(u) \geq \varsigma(u))T,$$

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

$$g(A) := E_P[f_Y(X'\beta_0(u) \mid X, C)X\dot{X}'\beta_0(u)1(S'\gamma_0(u) \geq \varsigma(u))T\ell(a, D, W, Z)]|_{a=A}.$$

Assumption 5(a) requires the selector to find a subset of the u -quantile-censored observations, whereas Assumption 6 requires the selector to find a nonempty subset. Let $\hat{\beta}^0(u)$ be an initial consistent estimator of $\beta_0(u)$ that uses a selector based on a flexible model for the conditional probability of censoring as described in Section 3. This model does not need to be correctly specified under a mild separating hyperplane condition for the u -quantile uncensored observations (Chernozhukov and Hong, 2002). Then, we update the selector to $1[s(x(D, W, \hat{V}), C)'\hat{\gamma}(u) \geq \varsigma(u)]$, where $s(x(D, W, \hat{V}), C) = [x(D, W, \hat{V})', C]'$, and $\hat{\gamma}(u) = [\hat{\beta}^0(u)', -1]'$. The parameter $\varsigma(u)$ is a small fixed cut-off that ensures that the selector is asymptotically conservative but nontrivial. We provide guidelines for the choice of $\varsigma(u)$ in Section 3 and show that the CQIV estimates are not very sensitive to this choice in the numerical examples of Section 4.

The full rank conditions of Assumption 6 hold if there are not perfectly collinear components in the vector $X = x(D, W, \vartheta_0(D, W, Z))$ and $P(S'\gamma_0(u) \geq \varsigma(u), T = 1) > 0$. To avoid reliance on functional form assumptions for x and ϑ_0 , the noncollinearity requires the exclusion restriction for Z in Assumption 1, $Q_Y(u \mid D, W, Z, V, C) = Q_Y(u \mid D, W, V, C)$ a.s., and a global relevance or rank condition for Z such as $\text{Var}_P[\vartheta_0(D, W, Z)|D, W] > 0$ a.s. Chesher (2003) and Jun (2009) impose local versions of the exclusion and relevance conditions for Z at a point of interest for V .

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*

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

We can estimate the variance-covariance matrix $J^{-1}(u)\Lambda(u)J^{-1}(u)$ using standard methods and carry out analytical inference based on the normal distribution. Estimators for the components of the variance can be formed 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 (Ma and Kosorok, 2005, Chen and Pouzo, 2009) and prove its validity in what follows.

We focus on weighted bootstrap because the proof of its consistency is not overly complex, following the strategy set forth by Ma and Kosorok (2005). This bootstrap also has practical advantages over nonparametric bootstrap to deal with discrete regressors with small cell sizes, because it avoids having singular designs under the bootstrap data generating process. Moreover, a particular version of the weighted bootstrap, with exponentials acting as weights, has a nice Bayesian interpretation (Hahn, 1997, Chamberlain and Imbens, 2003).

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

Assumption 7 (Bootstrap weights). *The weights (e_1, \dots, e_n) are i.i.d. draws from a random variable $e \geq 0$, with $E_P[e] = 1$, $\text{Var}_P[e] = 1$, and $E_P|e|^{2+\delta} < \infty$ for some $\delta > 0$; live 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 $E_P[e] = 1$ but $\text{Var}_P[e] > 1$, we can take the transformation $\tilde{e} = 1 + (e - 1)/\text{Var}_P[e]^{1/2}$, which satisfies $\tilde{e} \geq 0$, $E_P[\tilde{e}] = 1$, and $\text{Var}_P[\tilde{e}] = 1$.

The weights act as sampling weights in the bootstrap procedure. In each repetition, we draw a new set of weights (e_1, \dots, 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 \hat{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 below. The CQIV estimator in the weighted sample solves

$$\hat{\beta}^e(u) = \arg \min_{\beta \in \mathbb{R}^{\dim(X)}} \frac{1}{n} \sum_{i=1}^n e_i 1(\hat{\gamma}(u)' \hat{S}_i^e \geq \varsigma(u)) T_i \rho_u(Y_i - \beta' \hat{X}_i^e), \quad (2.13)$$

where $\hat{X}_i^e = x(D_i, W_i, \hat{V}_i^e)$, $\hat{S}_i^e = s(\hat{X}_i^e, C_i)$, and $\hat{\gamma}(u)$ is a consistent estimator of the selector. Note that we do not need to recompute $\hat{\gamma}(u)$ 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 8 (Weighted estimator of control variable). *Let (e_1, \dots, e_n) be a sequence of weights that satisfies Assumption 7. We have an estimator of the control variable of the form $\widehat{V}^e = \widehat{\vartheta}^e(D, W, Z)$ such that uniformly over \overline{DWZ} ,*

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

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

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

with probability approaching one, where the covering entropy of the function class Υ is not too high, namely

$$\log N(\epsilon, \Upsilon, \|\cdot\|_{T, \infty}) \lesssim 1/(\epsilon \log^4(1/\epsilon)), \quad \text{for all } 0 < \epsilon < 1.$$

Basically this is the same condition as Assumption 4 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 below.

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*

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

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, converges in probability to zero. More specifically,

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

where d_{BL} denotes the bounded Lipschitz 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 $\widehat{\beta}_b^e(u)$ solving the problem (2.13) with the data fixed and a set of weights (e_{1b}, \dots, e_{nb}) randomly drawn from a distribution that satisfies Assumption 7. By Theorem 2, we can use the empirical distribution of $\sqrt{n}(\widehat{\beta}_b^e(u) - \widehat{\beta}(u))$ 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, $\hat{\pi}(\cdot)$, 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 9 (QR control variable). *(a) The conditional quantile function of D given (W, Z) follows the quantile regression model:*

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

for all $v \in \mathcal{T} = [\tau, 1 - \tau]$, for some $\tau > 0$, where $\{F_D(d \mid w, z) : (d, w, z) \in \overline{\mathcal{D}\mathcal{W}\mathcal{Z}}\} \subseteq \mathcal{T}$, and the coefficients $v \mapsto \pi_0(v)$ are three times continuously differentiable with uniformly bounded derivatives on $v \in \mathcal{T}$; (b) $\overline{\mathcal{R}}$ is compact; (c) the conditional density $f_D(d \mid r)$ exists, is uniformly continuous in (d, r) over $\overline{\mathcal{D}\mathcal{R}}$, and is uniformly bounded; and (d) the minimal eigenvalue of $E_P[f_D(R' \pi_0(v) \mid R) R R']$ is bounded away from zero uniformly over $v \in \mathcal{T}$.

We impose that $\{F_D(d \mid w, z) : (d, w, z) \in \overline{\mathcal{D}\mathcal{W}\mathcal{Z}}\} \subseteq \mathcal{T}$ to ensure that the untrimmed observations are not at the tails of the conditional distribution, restricting the support of the control variable for these observations, i.e., $\overline{\mathcal{V}} \subseteq \mathcal{T}$. The differentiability of $v \mapsto \pi_0(v)$ is used in the proofs to construct a smooth approximation to the quantile regression process. The rest of the conditions are standard in quantile regression models (see, e.g., Koenker, 2005).

For $\rho_v(z) := (v - 1(z < 0))z$ and $v \in \mathcal{T}$, let

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

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

$$\vartheta_0(d, r) = \tau + \int_{\mathcal{T}} 1\{r'\pi_0(v) \leq d\}dv; \quad \widehat{\vartheta}^e(d, r) = \tau + \int_{\mathcal{T}} 1\{r'\widehat{\pi}^e(v) \leq d\}dv,$$

if $(d, r) \in \overline{\mathcal{DR}}$ and $\vartheta_0(d, r) = \tau$ otherwise.

The following result verifies that our main high-level conditions for the control variable estimator in Assumptions 4 and 8 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, as mentioned above.

Theorem 3 (Validity of Assumptions 4 & 8 for QR). *Suppose that Assumptions 2 and 9 hold. Then, (1)*

$$\begin{aligned} \sqrt{n}(\widehat{\vartheta}^e(d, r) - \vartheta_0(d, r)) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n e_i \ell(A_i, d, r) + o_{\mathbb{P}}(1) \rightsquigarrow \Delta^e(d, r) \text{ in } \ell^\infty(\overline{\mathcal{DR}}), \\ \ell(A, d, r) &:= f_D(d | r) r' E_P [f_D(R' \pi_0(\vartheta_0(d, r)) | R) R R']^{-1} \times \\ &\quad \times [1\{D \leq R' \pi_0(\vartheta_0(d, r))\} - \vartheta_0(d, r)] R, \\ E_P[\ell(A, d, r)] &= 0, \quad E_P[T \ell(A, D, R)^2] < \infty, \end{aligned}$$

where $\Delta^e(d, r)$ is a Gaussian process with continuous paths and covariance function given by $E_P[\ell(A, d, r) \ell(A, \tilde{d}, \tilde{r})']$. (2) Moreover, there exists $\widetilde{\vartheta}^e : \overline{\mathcal{DR}} \mapsto [0, 1]$ that obeys the same first order representation uniformly over $\overline{\mathcal{DR}}$, is close to $\widehat{\vartheta}^e$ in the sense that $\|\widetilde{\vartheta}^e - \widehat{\vartheta}^e\|_{T, \infty} = o_{\mathbb{P}}(1/\sqrt{n})$, and, with probability approaching one, belongs to a bounded function class Υ such that

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

Thus, Assumption 4 holds for the case $e_i = 1$, and Assumption 8 holds for the case of e_i being drawn from a positive random variable with unit mean and variance as in Assumption 7. 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 10 (DR control variable). (a) *The conditional distribution function of D given (W, Z) follows the distribution regression model, i.e.,*

$$F_D(d | W, Z) = F_D(d | R) = \Lambda(R' \pi_0(d)), \quad R = r(W, Z),$$

for all $d \in \overline{\mathcal{D}}$, where Λ is either the probit or logit link function, and the coefficients $d \mapsto \pi_0(d)$ are three times continuously differentiable with uniformly bounded derivatives; (b) $\overline{\mathcal{R}}$ is compact; (c) the minimum eigenvalue of

$$E_P \left[\frac{\partial \Lambda(R' \pi_0(d))^2}{\Lambda(R' \pi_0(d)) [1 - \Lambda(R' \pi_0(d))]} R R' \right]$$

is bounded away from zero uniformly over $d \in \overline{\mathcal{D}}$, where $\partial\Lambda$ is the derivative of Λ .

The differentiability of $d \mapsto \pi_0(d)$ is used in the proofs to construct a smooth approximation to the distribution regression process. The rest of the conditions are standard in distribution regression models (see, e.g., Chernozhukov, Fernandez-Val, and Melly, 2013).

For $d \in \overline{\mathcal{D}}$, 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

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

if $(d, r) \in \overline{\mathcal{DR}}$, and $\vartheta_0(d, r) = \widehat{\vartheta}^e(d, r) = 0$ otherwise.

The following result verifies that our main high-level conditions for the control variable estimator in Assumptions 4 and 8 hold under Assumption 10. 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 4 & 8 for DR). *Suppose that Assumptions 2 and 10 hold. Then, (1)*

$$\begin{aligned} \sqrt{n}(\widehat{\vartheta}^e(d, r) - \vartheta_0(d, r)) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n e_i \ell(A_i, d, r) + o_{\mathbb{P}}(1) \rightsquigarrow \Delta^e(d, r) \text{ in } \ell^\infty(\overline{\mathcal{DR}}), \\ \ell(A, d, r) &:= \partial \Lambda(r' \pi_0(d)) r' \mathbb{E}_P \left[\frac{\partial \Lambda(R' \pi_0(d))^2}{\Lambda(R' \pi_0(d)) [1 - \Lambda(R' \pi_0(d))]} R R' \right]^{-1} \times \\ &\quad \times \frac{1\{D \leq d\} - \Lambda(R' \pi_0(d))}{\Lambda(R' \pi_0(d)) [1 - \Lambda(R' \pi_0(d))]} \partial \Lambda(R' \pi_0(d)) R, \\ \mathbb{E}_P[\ell(A, d, r)] &= 0, \mathbb{E}_P[T \ell(A, D, R)^2] < \infty, \end{aligned}$$

where $\Delta^e(d, r)$ is a Gaussian process with continuous paths and covariance function given by $\mathbb{E}_P[\ell(A, d, r) \ell(A, \tilde{d}, \tilde{r})']$. (2) Moreover, there exists $\widetilde{\vartheta}^e : \overline{\mathcal{DR}} \mapsto [0, 1]$ that obeys the same first order representation uniformly over $\overline{\mathcal{DR}}$, is close to $\widehat{\vartheta}^e$ in the sense that $\|\widetilde{\vartheta}^e - \widehat{\vartheta}^e\|_{T, \infty} = o_{\mathbb{P}}(1/\sqrt{n})$ and, with probability approaching one, belongs to a bounded function class Υ such that

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

Thus, Assumption 4 holds for the case $e_i = 1$, and Assumption 8 holds for the case of e_i being drawn from a positive random variable with unit mean and variance as in Assumption 7. Thus, the results of Theorem 1 and 2 apply for the DR estimator of the control variable.

3. COMPUTATION

This section describes the numerical algorithms to compute the CQIV estimator and weighted bootstrap confidence intervals.

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 number this step as 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 $\widehat{V}_i = \widehat{\vartheta}(D_i, W_i, Z_i)$ from (2.11) or (2.12), and construct $\widehat{X}_i = x(D_i, W_i, \widehat{V}_i)$. 1) Select a set of u -quantile uncensored observations $J_0(u) = \{i : \Lambda(\widehat{S}_i' \widehat{\delta}) \geq 1 - u + k_0(u)\}$, where Λ is a known link function, $\widehat{S}_i = s(\widehat{X}_i, C_i)$, s is a vector of transformations, $k_0(u)$ is a cut-off such that $0 < k_0(u) < u$, and $\widehat{\delta} = \arg \max_{\delta \in \mathbb{R}^{\dim(S)}} \sum_{i=1}^n T_i \{1(Y_i > C_i) \log \Lambda(\widehat{S}_i' \delta) + 1(Y_i = C_i) \log[1 - \Lambda(\widehat{S}_i' \delta)]\}$, where $T_i = 1(D_i \in \overline{\mathcal{D}})$. 2) Obtain the 2-step CQIV coefficient estimates: $\widehat{\beta}^0(u) = \arg \min_{\beta \in \mathbb{R}^{\dim(X)}} \sum_{i \in J_0(u)} T_i \rho_u(Y_i - \widehat{X}_i' \beta)$, and update the set of u -quantile uncensored observations, $J_1(u) = \{i : \widehat{X}_i' \widehat{\beta}^0(u) \geq C_i + \varsigma_1(u)\}$. 3) Obtain the 3-step CQIV coefficient estimates $\widehat{\beta}^1(u)$, solving the same minimization program as in step 2 with $J_0(u)$ replaced by $J_1(u)$. 4. (Optional) Update the set of u -quantile uncensored observations J_2 replacing $\widehat{\beta}^0(u)$ by $\widehat{\beta}^1(u)$ in the expression for $J_1(u)$ in step 2, and iterate this and the previous step a bounded number of times.*

Remark 1 (Step 1). We can obtain $J_0(u)$ with a probit, logit, or any other model for the conditional probability of censoring capable of discriminating a subset of u -quantile uncensored observations. For example, we can use a logit model with $s(\widehat{X}_i, C_i)$ including powers or b-splines of the components of (\widehat{X}_i, C_i) and interaction terms. Given the slackness provided by the cut-off $k_0(u)$, the model does not need to be correctly specified. It suffices to select a nontrivial subset of observations with $\widehat{X}_i' \beta_0(u) > C_i$. To choose the value of $k_0(u)$, it is advisable that a constant fraction of observations satisfying $\Lambda(\widehat{S}_i' \widehat{\delta}) > 1 - u$ are excluded from $J_0(u)$ for each quantile. To do so, set $k_0(u)$ as the q_0 th 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 with little sensitivity to values between 5 and 15%).

Remark 2 (Step 2). To choose the cut-off $\varsigma_1(u)$, it is advisable that a constant fraction of observations satisfying $\widehat{X}_i' \widehat{\beta}^0(u) > C_i$ are excluded from $J_1(u)$ for each quantile. To do so,

set $\varsigma_1(u)$ to be the q_1 th 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 with little sensitivity to values between 1 and 5%). In practice, it is desirable that $J_0(u) \subset J_1(u)$. If this is not the case, we recommend altering q_0 , q_1 , or the specification of the regression models. At each quantile, the percentage of observations from the full sample retained in $J_0(u)$, the percentage of observations from the full sample retained in $J_1(u)$, and the percentage of observations from $J_0(u)$ not retained in $J_1(u)$ can be computed as simple robustness diagnostic tests. The estimator $\widehat{\beta}^0(u)$ is consistent but will be less inefficient than the estimator obtained in the subsequent step because it uses a smaller conservative subset of the u -quantile uncensored observations if $q_1 < q_0$.

Remark 3 (Steps 1 and 2). In the notation of Assumption 5, the selector of Step 1 can be expressed as $1(\widehat{S}_i' \widehat{\gamma}(u) \geq \varsigma_0(u))$, where $\widehat{S}_i' \widehat{\gamma}(u) = \widehat{S}_i' \widehat{\delta} - \Lambda^{-1}(1 - u)$ and $\varsigma_0(u) = \Lambda^{-1}(1 - u + k_0(u)) - \Lambda^{-1}(1 - u)$. The selector of Step 2 can be expressed as $1(\widehat{S}_i' \widehat{\gamma}(u) \geq \varsigma_1(u))$, where $\widehat{S}_i = (\widehat{X}_i', C_i)'$ and $\widehat{\gamma}(u) = (\widehat{\beta}^0(u)', -1)'$.

Remark 4 (Steps 1 and 2). The trimming rule is a useful theoretical device that is generally considered to have minor practical importance. In our numerical and empirical examples, the choice of \overline{D} as the observed support of D , i.e. no trimming, works well. We also found that the performance of the estimator is not sensitive to the use of other trimming rules such as $\overline{D} = [\widehat{D}_\tau, \widehat{D}_{1-\tau}]$ where \widehat{D}_τ is the empirical τ -quantile of D for some small τ (e.g, $\tau = .01$).

Remark 5 (Steps 2, 3 and 4). The CQIV algorithm provides a computationally convenient approximation to Powell's censored quantile regression estimator. 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.

Remark 6 (Step 4). Iterating over the 3-step CQIV estimator with fixed cutoff at $\varsigma_1(u)$ does not affect asymptotic efficiency, but it might improve finite-sample properties. In our numerical experiments, however, we find very little or no gain of iterating beyond Step 3 in terms of bias, root mean square error, and value of Powell objective function.

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 $b = 1, \dots, B$, repeat the following steps:* 1) *Draw a set of weights (e_{1b}, \dots, e_{nb}) i.i.d from the standard exponential distribution or another distribution that satisfies Assumption 7.* 2) *Reestimate the control variable in the weighted sample, $\hat{V}_{ib}^e = \hat{v}_b^e(D_i, W_i, Z_i)$, and construct $\hat{X}_{ib}^e = x(D_i, W_i, \hat{V}_{ib}^e)$.* 3) *Estimate the weighted quantile regression: $\hat{\beta}_b^e(u) = \arg \min_{\beta \in \mathbb{R}^{\dim(X)}} \sum_{i \in J_{1b}} e_{ib} T_i \rho_u(Y_i - \beta' \hat{X}_{ib}^e)$, where $J_{1b} = \{i : \hat{\beta}(u)' \hat{X}_{ib}^e \geq C_i + \varsigma_1(u)\}$, and $\hat{\beta}(u)$ is a consistent estimator of $\beta_0(u)$, e.g., the 3-stage CQIV estimator $\hat{\beta}^1(u)$.*

Remark 7 (Step 3). A computationally less expensive alternative is to set $J_{1b} = J_1$ in all the repetitions, where $J_1(u)$ is the subset of selected observations in Step 2 of the CQIV algorithm.

We can construct an asymptotic $(1 - \alpha)$ -confidence interval for a scalar function of the parameter vector $g(\beta_0(u))$ using the percentile method, i.e., $CI_{1-\alpha}[g(\beta_0(u))] = [\hat{g}_{\alpha/2}, \hat{g}_{1-\alpha/2}]$, where \hat{g}_α is the sample α -quantile of $[g(\hat{\beta}_1^e(u)), \dots, g(\hat{\beta}_B^e(u))]$. For example, let $\hat{\beta}_{b,k}^e(u)$ and $\beta_{0,k}(u)$ denote the k th components of $\hat{\beta}_b^e(u)$ and $\beta_0(u)$. Then, the 0.025 and 0.975 quantiles of $(\hat{\beta}_{1,k}^e(u), \dots, \hat{\beta}_{B,k}^e(u))$ form a 95% asymptotic confidence interval for $\beta_{0,k}(u)$.

4. MONTE-CARLO ILLUSTRATION

In this section, we develop a Monte-Carlo numerical example aimed at analyzing the performance of CQIV in finite samples. We first generate data according to two different designs. Next, we compare the performance of CQIV and tobit estimators in terms of bias and root mean squared error. Finally, we discuss the results of sensitivity and diagnostic tests calculated within the simulated data.

4.1. Monte-Carlo Designs. We generate data according to a 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 one of the tobit parametric assumptions.⁴ To facilitate the comparison, in both designs we consider a location model for the latent variable Y^* , where the coefficients of the conditional expectation function and the conditional

⁴The tobit parametric assumptions are classical location models for the first stage and second stage equations and jointly normal unobservables.

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 can be expected to perform as well as possible.

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

$$D = \pi_{00} + \pi_{01}Z + \pi_{02}W + \Phi^{-1}(V), \quad V \sim U(0, 1), \quad (4.1)$$

$$Y^* = \beta_{00} + \beta_{01}D + \beta_{02}W + \Phi^{-1}(\epsilon), \quad \epsilon \sim U(0, 1), \quad (4.2)$$

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

$$Y = \max(Y^*, C) = \max(\beta_{00} + \beta_{01}D + \beta_{02}W + \Phi^{-1}(\epsilon), C). \quad (4.3)$$

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 $U \sim U(0, 1)$. Using this expression, we can combine (4.2) and (4.3) 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 variable 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. Formally, we set $W = \exp[\min(W^*, q_{W^*})]$, where W^* is drawn from a standard normal distribution and q_{W^*} is the 95th sample percentile of W^* , which differs across replication samples. 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 to the 38th sample percentile of Y^* in each replication sample. We report results from 1,000 simulations with $n = 1,000$.

For the design with heteroskedastic first stage, we replace the first stage equation for D in (4.1) by the following equation:

$$D = \pi_{00} + \pi_{01}Z + \pi_{02}W + (\pi_{03} + \pi_{04}W)\Phi^{-1}(V), \quad V \sim U(0, 1) \quad (4.4)$$

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

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

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

4.2. Comparison of CQIV with Tobit. We consider two tobit estimators for comparison. Tobit-iv is the full information maximum likelihood estimator implemented in Stata with the default option of the command `ivtobit`.⁵ Tobit-cmle is the conditional maximum likelihood tobit estimator developed by Smith and Blundell (1986), which uses least squares residuals as a parametric control variable. For CQIV we consider three different methods to estimate the control variable: `cqiv-ols`, which uses least squares to estimate a parametric control variable; `cqiv-qr`, which uses quantile regression to estimate a semiparametric control variable; and `cqiv-dr`, which uses probit distribution regression to estimate a semiparametric control variable.⁶ All the CQIV estimators are computed in three stages using Algorithm 1 with $q_0 = 10$, $q_1 = 3$, no trimming, and a probit model in step 1.

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 $\{.05, .10, \dots, .95\}$ quantiles. For the tobit 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 asymptotically efficient in this design, the CQIV estimators out-perform tobit-iv, and compare well to tobit-cmle. `Cqiv-ols` and `cqiv-qr` display slightly lower rmse than `cqiv-dr` in this design. 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 design with heteroskedastic first stage. 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 and rmse than `cqiv-ols` because it uses a more flexible specification for the control variable. Moreover, at every quantile, `cqiv-qr` outperforms both tobit estimators, which are no longer consistent.

⁵The results reported use the algorithm “difficult” because the default algorithm does not converge in several simulations for the heteroskedastic design. The algorithm “bfgs” and the Newey’s (1987) minimum chi-squared option of the command give similar results to the ones reported.

⁶See appendix for technical details on the computation of the first stage estimators of the control variable.

In summary, CQIV performs well relative to tobit in a model that satisfies the parametric assumptions required for tobit-iv to be asymptotically efficient, and it outperforms tobit in a more flexible model that does not satisfy one of the tobit parametric assumptions.

4.3. Sensitivity and Diagnostic Tests. In Table 1, we analyze the sensitivity of the CQIV estimator to the choice of the quantiles q_0 and q_1 that determine the cut-offs of the selectors. For all the combination of values of $q_0 \in \{5, 10, 15\}$ and $q_1 \in \{1, 2, 5\}$, we report the mean bias and rmse of the 3-step cqiv-qr estimator in the tobit design and the design with heteroskedastic first stage. We find that the performance of the estimator shows very little sensitivity to the choice of quantiles within the range of values considered. In results not reported, we also find very little sensitivity to the choice of quantiles in the value of the Powell objective function computed from the 3-step estimator.

Table 2 reports feasible and unfeasible diagnostic tests for the 2-step, 3-step, and 4-step cqiv-qr estimators obtained by Algorithm 1 with $q_0 = 10$, $q_1 = 3$, and $q_2 = 3$ for both the tobit and nontobit designs. We recommend that applied researchers conduct the feasible tests. The unfeasible tests are those that involve $J^*(u) = \{i : X'_i\beta_0(u) > C_i\}$, the set of u -quantile uncensored observations, that is unobservable in practice. As shown in the table, the percentage of observations in $J(u)$ increases with the quantile. In the table, we compare the composition of $J(u)$ with the compositions of $J_0(u)$ and $J_1(u)$, the subsets of observations selected as u -quantile uncensored in the step 1 and step 2 of the algorithm. We find that $J_0(u)$ and $J_1(u)$ select most of the u -quantile uncensored observations.

The feasible tests are based on calculating the percentage of observations selected in $J_0(u)$ and $J_1(u)$, comparing the composition of the subsets $J_0(u)$ and $J_1(u)$, and calculating the value of the Powell objective function at each step of the algorithm. We find that the percentage of observations retained in $J_0(u)$ and $J_1(u)$ increases with the quantile, as it should given the percentage of observations in J . In applied settings, researchers can diagnose a problem if the number of observations retained in $J_0(u)$ and $J_1(u)$ varies little across quantiles and attempt to address it by making the specifications of the binary choice model in step 1 or the quantile regression model in steps 2 and 3 more flexible. We find that $J_0(u)$ is a strict subset of $J_1(u)$ in the column that reports the intersection of $J_0(u)$ with the complement of $J_1(u)$ ($J_0(u) \cap J_1(u)^c$). In applied settings, researchers can diagnose a problem if many observations from $J_0(u)$ are not included in $J_1(u)$ and attempt to address it by modifying the values of q_0 and q_1 . The value of the Powell objective function decreases between step 2 and step 3 of the algorithm in about 75% of the simulations, whereas it only further decreases with an additional iteration in about 25% of the simulations. In applied settings, researchers can use the relative values of the Powell objective function to assess the gains from iteration.

5. 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 the 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 occurs 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 (Gorman, 1959). 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 quantile regression estimator in (2.11), where $\tau = .01$ and 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 methods 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:

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

where Y is the observed share of total expenditure on alcohol with a mass point 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).⁷ To choose the specification, we rely on recent studies in Engel curve estimation. Thus, following Blundell, Browning, and Crawford (2003) we impose

⁷Demographic 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

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.

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 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 3 reports point and interval estimates of average quantile elasticities as a function of the quantile index u , i.e., $u \mapsto E_P[\partial_d Q_Y(u \mid X)]$. In addition to CQIV with a quantile estimator of the control variable (cqiv), we present results from the censored quantile regression (cqr) estimator of Chernozhukov and Hong (2002), which does not address endogeneity; two-stage quantile regression estimator (qiv) with quantile regression first 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. We also estimate a model for the conditional mean with the tobit-cmle of Smith and Blundell (1986). Given the level of censoring, we focus on conditional quantiles above the .15 quantile.

Fig. 3 shows that accounting for endogeneity and censoring has important consequences for the elasticities. 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 Algorithm 2 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 the tobit estimates of the elasticity.

In Figure 4 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

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.

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 ($u = .5$). 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.

6. CONCLUSION

In this paper, we develop new censored and uncensored quantile instrumental variable estimators that incorporate 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 $1(S'_i \gamma \geq \varsigma)$. It is convenient also to introduce some additional notation, which will be extensively used in the proofs. Let $V_i(\vartheta) := \vartheta(D_i, W_i, Z_i)$, $X_i(\vartheta) := x(D_i, W_i, V_i(\vartheta))$, $S_i(\vartheta) := s(X_i(\vartheta), C_i)$, $\dot{X}_i(\vartheta) := \partial_v x(D_i, W_i, v)|_{v=V_i(\vartheta)}$, and $\dot{S}_i(\vartheta) := \partial_v s(X_i(v), C_i)|_{v=V_i(\vartheta)}$. When the previous functions are evaluated at the true values we use $V_i = V_i(\vartheta_0)$, $X_i = X_i(\vartheta_0)$, $S_i = S_i(\vartheta_0)$, $\dot{X}_i = \dot{X}_i(\vartheta_0)$, and $\dot{S}_i = \dot{S}_i(\vartheta_0)$. Also, let $\varphi_u(z) := [1(z < 0) - u]$. Recall that $A := (Y, D, W, Z, C, X, V)$, $T(d) = 1(d \in \overline{\mathcal{D}})$, and $T = T(D)$. For a function $f : \mathcal{A} \mapsto \mathbb{R}$, we use $\|f\|_{T, \infty} = \sup_{a \in \mathcal{A}} |T(d)f(a)|$; for a K -vector of functions $f : \mathcal{A} \mapsto \mathbb{R}^K$, we use $\|f\|_{T, \infty} = \sup_{a \in \mathcal{A}} \|T(d)f(a)\|_2$. 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

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

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

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

We use the concepts of covering entropy and bracketing entropy in the proofs. The covering entropy $\log N(\epsilon, \mathcal{F}, \|\cdot\|)$ is the logarithm of the minimal number of $\|\cdot\|$ -balls of radius ϵ needed to cover the set of functions \mathcal{F} . The bracketing entropy $\log N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|)$ is the logarithm of the minimal number of ϵ -brackets in $\|\cdot\|$ needed to cover the set of functions \mathcal{F} . An ϵ -bracket $[\ell, u]$ in $\|\cdot\|$ is the set of functions f with $\ell \leq f \leq u$ and $\|u - \ell\| < \epsilon$.

APPENDIX B. PROOF OF THEOREMS 1 AND 2

Throughout this appendix we drop the dependence on u from all the parameters to lighten the notation; for example, β_0 and J signify $\beta_0(u)$ and $J(u)$.

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

By Assumptions 4 and 5, for large enough n :

$$1\{S(\widehat{\vartheta})'\widehat{\gamma} \geq \varsigma\}T \leq 1\{S'\gamma_0 \geq \varsigma - \epsilon_n\}T \leq 1\{S'\gamma_0 \geq \varsigma/2\}T \leq 1\{X'\beta_0 \geq C + \epsilon'\}T,$$

P -a.e., since

$$|S(\widehat{\vartheta})'\widehat{\gamma} - S'\gamma_0|T \leq \epsilon_n := L_S(\|\widehat{\vartheta} - \vartheta_0\|_{T,\infty} + \|\widehat{\gamma} - \gamma_0\|_2) \rightarrow_{\mathbb{P}} 0,$$

where $L_S := (\|\partial_v s\|_{T,\infty} \vee \|s\|_{T,\infty})$ is a finite constant by assumption. Hence, with probability approaching one

$$\widehat{\beta} = \arg \min_{\beta \in \mathbb{R}^{\dim(X)}} \mathbb{E}_n[\rho_u(Y - X(\widehat{\vartheta})'\beta)1(S(\widehat{\vartheta})'\widehat{\gamma} \geq \varsigma)T\chi],$$

where $\chi := 1\{X'\beta_0 \geq C + \epsilon'\}$.

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

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

where

$$\widehat{f}_{\eta, B_\epsilon}(A) := \varphi_u \left\{ Y - X(\widehat{\vartheta})'(\beta_0 + B_\epsilon\eta/\sqrt{n}) \right\} X(\widehat{\vartheta})1\{S(\widehat{\vartheta})'\widehat{\gamma} \geq \varsigma\}T\chi.$$

Let

$$f(A) := \varphi_u \{Y - X'\beta_0\} X 1\{S'\gamma_0 \geq \varsigma\} T.$$

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

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

where relations (1) and (2) follow by Lemma 1 and Lemma 2 with $\widetilde{\beta} = \beta_0 + B_\epsilon\eta/\sqrt{n}$, respectively, using that $\|\widehat{\vartheta} - \widetilde{\vartheta}\|_{T,\infty} = o_{\mathbb{P}}(1/\sqrt{n})$, $\widetilde{\vartheta} \in \Upsilon$, $\|\widetilde{\vartheta} - \vartheta_0\|_{T,\infty} = O_{\mathbb{P}}(1/\sqrt{n})$ and $\|\beta_0 + B_\epsilon\eta/\sqrt{n} - \beta_0\|_2 = O(1/\sqrt{n})$; relation (3) holds by Chebyshev inequality. Since J 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

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

where

$$\widehat{f}(A) := \varphi_u \left\{ Y - X(\widehat{\vartheta})'\widehat{\beta} \right\} X(\widehat{\vartheta}) 1\{S(\widehat{\vartheta})'\widehat{\gamma} \geq \varsigma\} T\chi.$$

Therefore

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

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

Therefore by invertibility of J ,

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

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

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

□

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

By Assumptions 5 and 8, with probability approaching one,

$$\widehat{\beta}^e = \arg \min_{\beta \in \mathbb{R}^{\dim(X)}} \mathbb{E}_n[e\rho_u(Y - X(\widehat{\vartheta}^e)'\beta)1(S(\widehat{\vartheta}^e)'\widehat{\gamma} \geq \varsigma)T\chi],$$

where e is the random variable used in the weighted bootstrap and $\chi = 1(X'\beta_0 \geq C + \epsilon')$. 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

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

where

$$\widehat{f}_{\eta, B_\epsilon}^e(A) := e \cdot \varphi_u \left\{ Y - X(\widehat{\vartheta}^e)'(\beta_0 + B_\epsilon\eta/\sqrt{n}) \right\} X(\widehat{\vartheta}^e)1\{S(\widehat{\vartheta}^e)'\widehat{\gamma} \geq \varsigma\}T\chi.$$

Let

$$f^e(A) := e \cdot \varphi_u \{Y - X'\beta_0\} X1\{S'\gamma_0 \geq \varsigma\}T.$$

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

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

where relations (1) and (2) follow by Lemma 1 and Lemma 2 with $\widetilde{\beta} = \beta_0 + B_\epsilon\eta/\sqrt{n}$, respectively, using that $\|\widehat{\vartheta}^e - \widetilde{\vartheta}^e\|_{T, \infty} = o_{\mathbb{P}}(1/\sqrt{n})$, $\widetilde{\vartheta}^e \in \Upsilon$, $\|\widetilde{\vartheta}^e - \vartheta_0\|_{T, \infty} = O_{\mathbb{P}}(1/\sqrt{n})$ and $\|\beta_0 + B_\epsilon\eta/\sqrt{n} - \beta_0\|_2 = O(1/\sqrt{n})$; relation (3) holds by Chebyshev inequality. Since J 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 $\sqrt{n}(\widehat{\beta}^e - \beta_0) = -J^{-1}\mathbb{G}_n(f^e + g^e) + o_{\mathbb{P}}(1)$ under the unconditional probability \mathbb{P} .

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

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

where

$$\widehat{f}^e(A) := e \cdot \varphi_u \left\{ Y - X(\widehat{\vartheta}^e)'\widehat{\beta}^e \right\} X(\widehat{\vartheta}^e)1\{S(\widehat{\vartheta}^e)'\widehat{\gamma} \geq \varsigma\}T\chi.$$

Therefore

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

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

Therefore by invertibility of J ,

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

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

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

Hence, under \mathbb{P}

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

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 $\epsilon > 0$, $\mathbb{P}^e(\|r_n\|_2 > \epsilon) \rightarrow_{\mathbb{P}} 0$. Indeed, this follows from Markov inequality and by

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

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, \dots, A_n

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

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

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

Conclude that conditional on the data A_1, \dots, A_n

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

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

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

□

B.3. Lemma on Stochastic Equicontinuity.

Lemma 1 (Stochastic equicontinuity). *Let $e \geq 0$ be a positive random variable with $\mathbb{E}_P[e] = 1$, $\text{Var}_P[e] = 1$, and $\mathbb{E}_P[e]^{2+\delta} < \infty$ for some $\delta > 0$, 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 $\chi = 1(X'\beta_0 \geq C + \epsilon')$,*

$$f(A, \vartheta, \beta, \gamma) := e \cdot [1(Y \leq X(\vartheta)'\beta) - u] \cdot X(\vartheta) \cdot 1(S(\vartheta)'\gamma \geq \varsigma) \cdot T \cdot \chi.$$

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 β_0 under the $\|\cdot\|_2$ metric, Υ_0 is the intersection of Υ , defined in Assumption 4, with a small neighborhood of ϑ_0 under the $\|\cdot\|_{T,\infty}$ metric, which are chosen to be small enough so that:

$$|X(\vartheta)'\beta - X'\beta_0|T \leq \epsilon'/2, \text{ } P\text{-a.e. } \forall (\vartheta, \beta) \in \Upsilon_0 \times \mathcal{B},$$

where ϵ' is defined in Assumptions 5. This class is P -Donsker with a square integrable envelope of the form e times a constant.

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

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

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

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

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

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

we have that

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

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

$$\mathcal{I}_1 = \{[1(Y \leq X(\vartheta)' \beta) - u]T\chi : \beta \in \mathcal{B}, \vartheta \in \Upsilon_0\}.$$

For this purpose consider a mesh $\{\vartheta_k\}$ over Υ_0 of $\|\cdot\|_{T,\infty}$ 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

$$[i_1^0, i_1^1] = [\{1(Y \leq X(\vartheta_k)' \beta_l - \kappa\delta) - u\}T\chi, \{1(Y \leq X(\vartheta_k)' \beta_l + \kappa\delta) - u\}T\chi],$$

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

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

$$\begin{aligned} |X(\vartheta)' \beta - X(\vartheta_k)' \beta_l|T &\leq |(X(\vartheta) - X(\vartheta_k))' \beta|T + |X(\vartheta_k)' (\beta - \beta_l)|T \\ &\leq L_X \delta \max_{\beta \in \mathcal{B}} \|\beta\|_2 + L_X \delta \leq \kappa\delta, \end{aligned} \tag{B.5}$$

and the $\|\cdot\|_{P,2}$ -size of this bracket is given by

$$\begin{aligned} \|i_1^0 - i_1^1\|_{P,2} &\leq \sqrt{\mathbb{E}_P[P\{Y \in [X(\vartheta_k)' \beta_l \pm \kappa\delta] \mid D, W, Z, C, \chi = 1\}T]} \\ &\leq \sqrt{\mathbb{E}_P[\sup_{y \in (C + \kappa\delta, \infty)} P\{Y \in [y \pm \kappa\delta] \mid X, C, \chi = 1\}T]} \\ &\leq \sqrt{\|f_Y(\cdot \mid \cdot)\|_{T,\infty} 2\kappa\delta}, \end{aligned}$$

provided that $2\kappa\delta < \epsilon'/2$. In order to derive this bound we use the condition $|X(\vartheta)' \beta - X(\vartheta_k)' \beta_l|T \leq \epsilon'/2$, P -a.e. $\forall (\vartheta, \beta) \in \Upsilon_0 \times \mathcal{B}$, so that conditional on $\chi = 1$ we have that $X(\vartheta)' \beta \geq C + \epsilon'/2$; 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, ∞) .

Hence, counting the number of brackets induced by the mesh created above, we arrive at the following relationship between the bracketing entropy of \mathcal{I}_1 and the covering entropies

of Υ_0 and \mathcal{B} ,

$$\log N_{[]}(\epsilon, \mathcal{I}_1, \|\cdot\|_{P,2}) \lesssim \log N(\epsilon^2, \Upsilon_0, \|\cdot\|_{T,\infty}) + \log N(\epsilon^2, \mathcal{B}, \|\cdot\|_2) \lesssim 1/(\epsilon^2 \log^4 \epsilon) + \log(1/\epsilon),$$

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

Step 2. Similarly to Step 1, it follows that

$$\mathcal{I}_2 = \{X(\vartheta)' \alpha T : \vartheta \in \Upsilon_0, \alpha \in \mathbb{R}^{\dim(X)}, \|\alpha\|_2 \leq 1\}$$

also obeys a similar bracketing entropy bound

$$\log N_{[]}(\epsilon, \|\cdot\|_{P,2}) \lesssim 1/(\epsilon^2 \log^4 \epsilon) + \log(1/\epsilon)$$

with a generic bracket taking the form $[i_2^0, i_2^1] = [\{X(\vartheta_k)' \beta_l - \kappa \delta\}T, \{X(\vartheta_k)' \beta_l + \kappa \delta\}T]$. Hence, this class is also P-Donsker with a constant envelope.

Step 3. Here we bound the bracketing entropy for

$$\mathcal{I}_3 = \{1(S(\vartheta)' \gamma \geq \varsigma)T : \vartheta \in \Upsilon_0, \gamma \in \Gamma\}.$$

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

$$[i_3^0, i_3^1] = [1(S(\vartheta_k)' \gamma_l - \kappa \delta \geq \varsigma)T, 1(S(\vartheta_k)' \gamma_l + \kappa \delta \geq \varsigma)T],$$

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

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

$$\begin{aligned} |S(\vartheta)' \gamma - S(\vartheta_k)' \gamma_l|T &\leq |(S(\vartheta) - S(\vartheta_k))' \gamma|T + |S(\vartheta_k)'(\gamma - \gamma_l)|T \\ &\leq L_S \delta \max_{\gamma \in \Gamma} \|\gamma\|_2 + L_S \delta \leq \kappa \delta, \end{aligned} \tag{B.6}$$

and the $\|\cdot\|_{P,2}$ -size of this bracket is given by

$$\|i_3^0 - i_3^1\|_{P,2} \leq \sqrt{P\{|S(\vartheta_k)' \gamma_l - \varsigma|T \leq 2\kappa \delta\}} \leq \sqrt{\bar{f}_S 2\kappa \delta},$$

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 $\gamma \in \Gamma$.

Hence, counting the number of brackets induced by the mesh created above, we arrive at the following relationship between the bracketing entropy of \mathcal{I}_3 and the covering entropies of Υ_0 and Γ ,

$$\log N_{[]}(\epsilon, \mathcal{I}_3, \|\cdot\|_{P,2}) \lesssim \log N(\epsilon^2, \Upsilon_0, \|\cdot\|_{T,\infty}) + \log N(\epsilon^2, \Gamma, \|\cdot\|_2) \lesssim 1/(\epsilon^2 \log^4 \epsilon) + \log(1/\epsilon)$$

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 . 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).

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), $1(Y \leq X(\vartheta)' \beta) T \chi \rightarrow 1(Y \leq X' \beta_0) T \chi$ everywhere, except for the set $\{A \in \mathcal{A} : Y = X' \beta_0\}$ 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), $|X(\vartheta)' \beta T - X' \beta_0 T| \rightarrow 0$ everywhere;
- (3) in view of the relation such as (B.6), $1(S(\vartheta)' \gamma \geq \varsigma) T \rightarrow 1(S' \gamma_0 \geq \varsigma) T$ 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 $\|\cdot\|_{T,\infty} \vee \|\cdot\|_2 \vee \|\cdot\|_2$ metric in view of Claim (b).

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

$$\begin{aligned} |\mathbb{G}_n[\hat{f} - \tilde{f}]| &\leq |\sqrt{n} \mathbb{E}_n[\hat{f} - \tilde{f}]| + |\sqrt{n} \mathbb{E}_P[\hat{f} - \tilde{f}]| \\ &\lesssim \sqrt{n} \mathbb{E}_n[\hat{\zeta}] + \sqrt{n} \mathbb{E}_P[\hat{\zeta}] \\ &\lesssim \mathbb{G}_n[\hat{\zeta}] + 2\sqrt{n} \mathbb{E}_P[\hat{\zeta}], \end{aligned}$$

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

$$|\hat{f} - \tilde{f}| \lesssim |e| \cdot \|X(\hat{\vartheta}) - X(\tilde{\vartheta})\|_2 \cdot T + \hat{g} + \hat{h} \lesssim \hat{\zeta} := e \cdot L_X \Delta_n + \hat{g} + \hat{h}, \quad \Delta_n \geq \|\hat{\vartheta} - \tilde{\vartheta}\|_{T,\infty}, \quad (\text{B.7})$$

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

$$\hat{g} := e \cdot 1\{|Y - X(\tilde{\vartheta})' \tilde{\beta}| \leq k \Delta_n\} T \chi, \quad \text{and} \quad \hat{h} := e \cdot 1\{|S(\tilde{\vartheta})' \tilde{\gamma} - \varsigma| \leq k \Delta_n\} T,$$

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

Hence it suffices to show that the result follows from

$$\mathbb{G}_n[\hat{\zeta}] = o_{\mathbb{P}}(1), \quad (\text{B.8})$$

and

$$\sqrt{n}\mathbb{E}_P[\widehat{\zeta}] = o_{\mathbb{P}}(1). \quad (\text{B.9})$$

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

$$\mathcal{G} = \{e \cdot 1(|Y - X(\vartheta)' \beta| \leq k) T_{\chi} : \vartheta \in \Upsilon_0, \beta \in \mathcal{B}, k \in [0, \epsilon'/4]\},$$

$$\mathcal{H} = \{e \cdot 1(|S(\vartheta)' \gamma - \varsigma| \leq k) T : \vartheta \in \Upsilon_0, \gamma \in \Gamma, k \in [0, 1]\}.$$

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

$$\log N_{[]}(\epsilon, \mathcal{G}, L_2(P)) \lesssim 1/(\epsilon^2 \log^4 \epsilon) \text{ and } \log N_{[]}(\epsilon, \mathcal{H}, L_2(P)) \lesssim 1/(\epsilon^2 \log^4 \epsilon).$$

Hence these classes are P-Donsker with unit envelopes. Let $g = e \cdot 1\{|Y - X(\vartheta)' \beta| \leq k \Delta_n\} T_{\chi}$, and $h = e \cdot 1\{|S(\vartheta)' \gamma - \varsigma| \leq k \Delta_n\} T$. Note also that if $(\vartheta, \beta, \gamma) \rightarrow (\vartheta_0, \beta_0, \gamma_0)$ in the $\|\cdot\|_{T,\infty} \vee \|\cdot\|_2 \vee \|\cdot\|_2$ metric, then

$$\|g\|_{P,2} \leq \sqrt{\mathbb{E}_P[e^2] \cdot P\{|Y - X(\vartheta)' \beta| T \leq k \Delta_n\}} \leq \sqrt{4\|f_Y(\cdot | \cdot)\|_{T,\infty} k \Delta_n} = o(1), \quad (\text{B.10})$$

$$\|h\|_{P,2} \leq \sqrt{\mathbb{E}_P[e^2] \cdot P\{|S(\vartheta)' \gamma - \varsigma| T \leq k \Delta_n\}} \leq \sqrt{4\bar{f}_S k \Delta_n} = o(1), \quad (\text{B.11})$$

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

Conclude that the relation (B.8) holds by (B.7), (B.10), (B.11), 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, if $(\vartheta, \beta, \gamma) \rightarrow (\vartheta_0, \beta_0, \gamma_0)$ in the $\|\cdot\|_{T,\infty} \vee \|\cdot\|_2 \vee \|\cdot\|_2$ metric,

$$\|e \cdot L_X \Delta_n + g + h\|_{P,2} = o(1) \Rightarrow \mathbb{G}_n[\widehat{\zeta}] = o_{\mathbb{P}}(1).$$

To show (B.9) note that if $(\vartheta, \beta, \gamma) \rightarrow (\vartheta_0, \beta_0, \gamma_0)$ in the $\|\cdot\|_{T,\infty} \vee \|\cdot\|_2 \vee \|\cdot\|_2$ metric,

$$\|e \cdot L_X \Delta_n + g + h\|_{P,1} \leq \mathbb{E}_P|e| \cdot L_X \Delta_n + \|g\|_{P,1} + \|h\|_{P,1} = o(1/\sqrt{n}) \Rightarrow \mathbb{E}_P[\widehat{\zeta}] = o_{\mathbb{P}}(1/\sqrt{n}),$$

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

$$\|g\|_{P,1} \leq \mathbb{E}_P|e| \cdot P\{|Y - X(\vartheta)' \beta| T \leq k \Delta_n\} \leq 2k\|f_Y(\cdot | \cdot)\|_{T,\infty} \Delta_n = o(1/\sqrt{n})$$

$$\|h\|_{P,1} \leq \mathbb{E}_P|e| \cdot P\{|S(\vartheta)' \gamma - \varsigma| T \leq k \Delta_n\} \leq 2k\bar{f}_S \Delta_n = o(1/\sqrt{n}),$$

by the assumption on bounded densities.

B.4. Lemma on Local Expansion.

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

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

we have that

$$\sqrt{n} \mathbb{E}_P \varphi_u \{Y - X(\widehat{\vartheta})' \widetilde{\beta}\} X(\widehat{\vartheta}) 1\{S(\widehat{\vartheta})' \gamma \geq \varsigma\} T \chi = J \widehat{\delta} + \sqrt{n} \mathbb{E}_n[g(A)] + o_{\mathbb{P}}(1),$$

where

$$g(A) = \int B(a) \ell(A, d, r) dP(a, d, r), \quad B(A) := f_Y(X' \beta_0 | X, C) X \dot{X}' \beta_0 1(S' \gamma_0 \geq \varsigma) T.$$

Proof of Lemma 2. With probability approaching one,

$$1\{S(\widehat{\vartheta})' \widehat{\gamma} \geq \varsigma\} T \leq 1\{S' \gamma_0 \geq \varsigma - \epsilon_n\} T \leq 1\{S' \gamma_0 \geq \varsigma/2\} T \leq 1\{X' \beta_0 \geq C + \epsilon'\} T,$$

P -a.e., by Assumptions 4 and 5 since

$$|S(\widehat{\vartheta})' \widehat{\gamma} - S' \gamma_0| T \leq \epsilon_n := L_S(\|\widehat{\vartheta} - \vartheta_0\|_{T, \infty} + \|\widehat{\gamma} - \gamma_0\|_2) \rightarrow_{\mathbb{P}} 0,$$

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

Hence uniformly in X over $\{X' \beta_0 \geq C + \epsilon'\}$,

$$\begin{aligned}& \sqrt{n} \mathbb{E}_P[\varphi_u \{Y - X(\widehat{\vartheta})' \widetilde{\beta}\} | D, W, Z, C] T \\&= f_Y(X(\bar{\vartheta}_X)' \bar{\beta}_X | D, W, Z, C) \{X(\bar{\vartheta}_X)' \widehat{\delta} + \dot{X}(\bar{\vartheta}_X)' \bar{\beta}_X \widehat{\Delta}(D, W, Z)\} T \\&= f_Y(X' \beta_0 | D, W, Z, C) \{X' \widehat{\delta} + \dot{X}' \beta_0 \widehat{\Delta}(D, W, Z)\} T + R_X, \\&= f_Y(X' \beta_0 | X, C) \{X' \widehat{\delta} + \dot{X}' \beta_0 \widehat{\Delta}(D, W, Z)\} T + R_X, \\&\bar{R} = \sup_{\{X: X' \beta_0 \geq C + \epsilon'\}} |R_X| = o_{\mathbb{P}}(1),\end{aligned}$$

where $\bar{\vartheta}_X$ is on the line connecting ϑ_0 and $\widehat{\vartheta}$ and $\bar{\beta}_X$ is on the line connecting β_0 and $\widetilde{\beta}$. The first equality follows by the mean value expansion. The second equality follows by the uniform continuity assumption of $f_Y(\cdot | X, C)$ uniformly in X, C , uniform continuity of $X(\cdot)$ and $\dot{X}(\cdot)$, and by $\|\widehat{\vartheta} - \vartheta_0\|_{T, \infty} \rightarrow_{\mathbb{P}} 0$ and $\|\widetilde{\beta} - \beta_0\|_2 \rightarrow_{\mathbb{P}} 0$. The third equality follows by

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

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

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

$$\begin{aligned} & E_P[\varphi_u\{Y - X(\widehat{\vartheta})'\widetilde{\beta}\}X(\widehat{\vartheta})1(S(\widehat{\vartheta})'\widehat{\gamma} \geq \varsigma)T\chi] \\ &= E_P[f_Y(X'\beta_0|X, C)XX'1\{S(\widehat{\vartheta})'\widehat{\gamma} \geq \varsigma\}T]\widehat{\delta} \\ &+ E_P[f_Y(X'\beta_0|X, C)X\dot{X}'\beta_01\{S(\widehat{\vartheta})'\widehat{\gamma} \geq \varsigma\}T\widehat{\Delta}(D, W, Z)] + O_{\mathbb{P}}(\bar{R}). \end{aligned} \quad (\text{B.12})$$

Furthermore since

$$E_P[1(S'\gamma_0 \geq \varsigma) - 1(S(\widehat{\vartheta})'\widehat{\gamma} \geq \varsigma)|T] \leq E_P[1(S'\gamma_0 \in [\varsigma \pm \epsilon_n])T] \lesssim \bar{f}_S \epsilon_n \rightarrow_{\mathbb{P}} 0,$$

where \bar{f}_S is a constant representing the uniform upper bound on the density of random variable $S'\gamma_0$, the expression (B.12) is equal to

$$J\widehat{\delta} + E_P[f_Y(X'\beta_0|X, C)X\dot{X}'\beta_01(S'\gamma_0 \geq \varsigma)T\widehat{\Delta}(D, W, Z)] + O_{\mathbb{P}}(\bar{f}_S \epsilon_n + \bar{R}).$$

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

$$E_P[f_Y(X'\beta_0|X, C)X\dot{X}'\beta_01(S'\gamma_0 \geq \varsigma)T\widehat{\Delta}(D, W, Z)] = \sqrt{n} \mathbb{E}_n[g(A)] + o_{\mathbb{P}}(1).$$

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 (2013),

$$\sqrt{n}(\widehat{\pi}(v) - \pi_0(v)) = \frac{1}{\sqrt{n}} \sum_{i=1}^n e_i E_P[f_D(R'\pi_0(v) | R)RR']^{-1} [v - 1\{D \leq R'\pi_0(v)\}]R + o_P(1),$$

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

$$\phi_\pi(d, r) = \tau + \int_{\mathcal{T}} 1\{r'\pi(v) \leq d\}dv$$

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

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

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 $\ell^\infty(\overline{\mathcal{DR}})$, for $\widehat{\vartheta}(d, r) = \phi_{\widehat{\pi}}(d, r)$,

$$\begin{aligned} \sqrt{n}(\widehat{\vartheta}(d, r) - \vartheta_0(d, r)) &= -f_D(d | r)r'\frac{1}{\sqrt{n}} \sum_{i=1}^n e_i E_P[f_D(R'\pi_0(\vartheta_0(d, r)) | R)RR']^{-1} \times \\ &\times [\vartheta_0(d, r) - 1\{D \leq R'\pi_0(\vartheta_0(d, r))\}]R + o_{\mathbb{P}}(1). \end{aligned}$$

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

$$\sqrt{n}(\widehat{\pi}(\cdot) - \pi_0(\cdot)) \Rightarrow Z_\pi \text{ in } \ell^\infty(\mathcal{T}), \quad \text{and} \quad r'\sqrt{n}(\widehat{\pi}(\cdot) - \pi_0(\cdot)) \Rightarrow r'Z_\pi \text{ in } \ell^\infty(\mathcal{T} \times \overline{\mathcal{R}}), \quad (\text{C.1})$$

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 covering entropy. Let $C_M^2(\mathcal{T})$ denote the class of functions $f : \mathcal{T} \rightarrow \mathbb{R}$ 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 $\log N(\epsilon, C_M^2(\mathcal{T}), \|\cdot\|_\infty) \lesssim \epsilon^{-1/2}$. Hence also $\log N(\epsilon, \times_{j=1}^{\dim(R)} C_M^2(\mathcal{T}), \|\cdot\|_\infty) \lesssim \epsilon^{-1/2}$. Next construct the set of functions Υ for some small $m > 0$ as:

$$\left\{ \tau + T \int_{\mathcal{T}} 1\{R'\pi(v) \leq D\} dv : \pi = (\pi_1, \dots, \pi_{\dim(R)}) \in \times_{j=1}^{\dim(R)} C_M^2(\mathcal{T}), R'\partial\pi(v) > \frac{m}{1-2\tau} \text{ P-a.e.} \right\}.$$

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

$$\begin{aligned} & T \left| \int_{\mathcal{T}} 1\{R'\pi(v) \leq D\} dv - \int_{\mathcal{T}} 1\{R'\bar{\pi}(v) \leq D\} dv \right| \\ & \leq T \int_{\mathcal{T}} 1\left\{ R'\pi(v) - D \in [-\|R\|_2\delta, \|R\|_2\delta] \right\} dv \lesssim \frac{1}{m} \|TR\|_2 \delta \lesssim \delta, \end{aligned}$$

P -a.e., since the density of $r'\pi(V)$, $V \sim U(\mathcal{T})$, is bounded above by $1/m$. We conclude that

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

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

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

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

where with probability approaching one, $\tilde{\pi} \in \times_{j=1}^{\dim(R)} C_M^2(\mathcal{T})$ and $R'\partial\tilde{\pi}(v) > m/(1-2\tau)$ P -a.e., for some M and some $m > 0$.

We construct $\tilde{\pi}$ by smoothing $\hat{\pi}$ component by component. Let the components of $\hat{\pi}$ be indexed by $1 \leq j \leq \dim(R)$. Before smoothing, we need to extend $\hat{\pi}_j$ outside \mathcal{T} . Start by extending the estimand π_{0j} outside \mathcal{T} onto the ϵ -expansion \mathcal{T}^ϵ 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 $\hat{\pi}_j$ to the outer region by setting $\hat{\pi}_j(v) = \pi_{0j}(v) + \hat{\pi}_j(\tau) - \pi_{0j}(\tau)$ if $v \leq \tau$ and $\hat{\pi}_j(v) = \pi_{0j}(v) + \hat{\pi}_j(1-\tau) - \pi_{0j}(1-\tau)$ if $v \geq 1-\tau$. The extension does not produce a feasible estimator, but it is a useful theoretical device. Note that the extended empirical process $\sqrt{n}(\hat{\pi}_j(v) - \pi_{0j}(v))$ remains to be stochastically equicontinuous by construction. Then we

define $\tilde{\pi}_j$ as the smoothed version of $\hat{\pi}_j$, namely

$$\tilde{\pi}_j(v) = \int_{\mathcal{T}^\epsilon} \hat{\pi}_j(z) [K((z-v)/h)/h] dz, \quad v \in \mathcal{T},$$

where $0 \leq h \leq \epsilon$ is bandwidth such that $\sqrt{nh^3} \rightarrow 0$ and $\sqrt{nh^2} \rightarrow \infty$; $K : \mathbb{R} \rightarrow \mathbb{R}$ 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

$$\begin{aligned} \sqrt{n}(\tilde{\pi}_j(v) - \hat{\pi}_j(v)) &= \int_{\mathcal{T}^\epsilon} \sqrt{n}[\hat{\pi}_j(z) - \pi_{0j}(z) - (\hat{\pi}_j(v) - \pi_{0j}(v))] [K([z-v]/h)/h] dz \\ &\quad + \int_{\mathcal{T}^\epsilon} \sqrt{n}(\pi_{0j}(z) - \pi_{0j}(v)) [K([z-v]/h)/h] dz. \end{aligned}$$

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

$$\omega(2h) = \sup_{|z-v| \leq 2h} |\sqrt{n}[\hat{\pi}_j(z) - \pi_{0j}(z) - (\hat{\pi}_j(v) - \pi_{0j}(v))]| \rightarrow_{\mathbb{P}} 0,$$

by the stochastic equicontinuity of the process $\sqrt{n}(\hat{\pi}_j(\cdot) - \pi_{0j}(\cdot))$ over \mathcal{T}^ϵ . The second term is bounded uniformly in $v \in \mathcal{T}$, up to a constant, by

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

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

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

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

which can be decomposed into two pieces:

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

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

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

This expression is bounded in absolute value by

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

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

Finally, since by Assumption 9(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, $R' \partial \tilde{\pi}(v) > m/(1 - 2\tau)$ P -a.e. for $m := k(1 - 2\tau)/2 > 0$.

Next we construct

$$\tilde{\vartheta}(d, r) = \phi_{\tilde{\pi}}(d, r) = \tau + \int_{\mathcal{T}} 1\{r' \tilde{\pi}(v) \leq d\} dv,$$

if $(d, r) \in \overline{\mathcal{DR}}$, and $\tilde{\vartheta}(d, r) = \tau$ otherwise. Note that by construction $\tilde{\vartheta} \in \Upsilon$ for some M with probability approaching one. It remains to show the first order equivalence with $\hat{\vartheta}$.

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), $\tilde{\vartheta}$ and $\hat{\vartheta}$ have the same first order representation in $\ell^\infty(\overline{\mathcal{DR}})$,

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

i.e., $\sqrt{n}\|\tilde{\vartheta} - \hat{\vartheta}\|_{T, \infty} \rightarrow_{\mathbb{P}} 0$. \square

APPENDIX D. PROOF OF THEOREM 4

Claim (1) follows from the results of Chernozhukov, Fernández-Val and Melly (2013). Also for future reference, note that these results also imply that

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

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 covering entropy. Let $C_M^2(\overline{\mathcal{D}})$ denote the class of functions $f : \overline{\mathcal{D}} \rightarrow \mathbb{R}$ with 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 $\log(\epsilon, C_M^2(\overline{\mathcal{D}}), \|\cdot\|_\infty) \lesssim \epsilon^{-1/2}$. Hence

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

Next construct

$$\Upsilon = \left\{ T\Lambda(R'\pi(D)) : \pi = (\pi_1, \dots, \pi_{\dim(R)}) \in \times_{j=1}^{\dim(R)} C_M^2(\overline{\mathcal{D}}) \right\}.$$

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

$$T |\Lambda(R'\pi(D)) - \Lambda(R'\bar{\pi}(D))| \leq \|\partial\Lambda\|_{T,\infty} \sup_{r \in \bar{\mathcal{R}}} \|r\|_\infty \delta.$$

We conclude that

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

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

We first construct $\hat{\pi}$ and $\tilde{\pi}$ such that,

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

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

We construct $\tilde{\pi}$ by smoothing $\hat{\pi}$ component by component. Before smoothing, we extend the estimand π_{0j} outside $\bar{\mathcal{D}}$, onto the ϵ -expansion $\bar{\mathcal{D}}^\epsilon$ 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 $\hat{\pi}_j$ to the outer region by setting $\hat{\pi}_j(d) = \pi_{0j}(d) + \hat{\pi}_j(\underline{d}) - \pi_{0j}(\underline{d})$ if $d \leq \underline{d}$, and $\hat{\pi}_j(d) = \pi_{0j}(d) + \hat{\pi}_j(\bar{d}) - \pi_{0j}(\bar{d})$ if $d \geq \bar{d}$. The extension does not produce a feasible estimator, but it is a useful theoretical device. Note that the extended process $\sqrt{n}(\hat{\pi}_j(d) - \pi_{0j}(d))$ remains to be stochastically equicontinuous by construction. Then we define the smoothed version of $\hat{\pi}_j$ as

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

where $0 \leq h \leq \epsilon$ is bandwidth such that $\sqrt{n}h^3 \rightarrow 0$ and $\sqrt{n}h^2 \rightarrow \infty$; $K : \mathbb{R} \rightarrow \mathbb{R}$ 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 (Berliner, 1993, and Newey, Hsieh, and Robins, 2004). We then have

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

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

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

by the stochastic equicontinuity of the process $\sqrt{n}(\hat{\pi}_j(\cdot) - \pi_{0j}(\cdot))$ over $\overline{\mathcal{D}}^\epsilon$. The second term is bounded uniformly in $d \in \overline{\mathcal{D}}$, up to a constant, by

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

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

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

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

which can be decomposed into two pieces:

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

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

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

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

We then construct $\tilde{\vartheta}(d, r) = \Lambda(r'\tilde{\pi}(d))$ if $(d, r) \in \overline{\mathcal{DR}}$, and $\tilde{\vartheta}(d, r) = 0$ otherwise. Note that by the preceding arguments $\tilde{\vartheta} \in \Upsilon$ for some M with probability approaching one. Finally, the first order equivalence $\sqrt{n}\|\tilde{\vartheta} - \hat{\vartheta}\|_{T,\infty} \rightarrow_{\mathbb{P}} 0$ follows immediately from (D.2), boundedness of $\|\partial\Lambda\|_{T,\infty}$ and compactness of $\overline{\mathcal{R}}$.

APPENDIX E. COMPUTATION DETAILS FOR FIRST STAGE ESTIMATORS

For the OLS estimator of the control variable in our CQIV estimator, we run an OLS first stage and retain the residuals 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, i.e. we set $\tau = .01$. Next, for each observation, we compute the fraction of the quantile estimates for which the predicted value is less than or equal to the observed value. We then evaluate the standard normal quantile function at this value and retain the result as the estimate of the control variable.

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 $i = j$.

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] Berline, Alain. "Hierarchies of higher order kernels." *Probability Theory and Related Fields*. 1993. 94(4). pp 489-504.
- [4] Blundell, Richard, Chen, Xiaohong, and Kristensen, Dennis. "Semi-nonparametric IV Estimation of Shape-Invariant Engel Curves." *Econometrica*. 2007. 75(6). pp. 1613-1669.
- [5] Blundell, Richard, and Matzkin, Rosa. "Conditions for the Existence of Control Functions in Nonseparable Simultaneous Equations Models." CEMMAP Working Paper 28/10.
- [6] Blundell, Richard, Duncan, Alan, and Pendakur, Krishna. "Semiparametric Estimation and Consumer Demand." *Journal of Applied Econometrics*. 1998. 13(5). pp. 435-461.
- [7] Blundell, Richard, and Powell, James. "Endogeneity in nonparametric and semiparametric regression models." *Econometric Society Monographs* 36. 2003. pp. 312-357.
- [8] Blundell, Richard, and Powell, James. "Endogeneity in Semiparametric Binary Response Models." *Review of Economic Studies* 71. 2004. pp. 581-613.
- [9] Blundell, Richard, and Powell, James. "Censored Regression Quantiles with Endogenous Regressors." *Journal of Econometrics*. 2007. 141. pp. 65-83.
- [10] Buchinsky, Moshe, and Hahn, Jinyong. "A Alternative Estimator for Censored Quantile Regression." *Econometrica*. 1998. 66. pp. 653-671.
- [11] Chamberlain, Gary, and Imbens, Guido. "Nonparametric applications of Bayesian inference." *Journal of Business and Economic Statistics*. 2003. 21(1). pp. 12-18.
- [12] Chen, Songnian, and Khan, Shakeeb. "Semiparametric estimation of a partially linear censored regression model." *Econometric Theory*. 2011. 17(3). pp. 567 - 590.
- [13] 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.
- [14] Chernozhukov, Victor, Fernandez-Val, Ivan, and Galichon, Alfred. "Quantile and Probability Curves without Crossing." *Econometrica*. 2010. 78(3, May) 1093-1125.
- [15] Chernozhukov, Victor, Fernandez-Val, Ivan, and Melly, Blaise. "Inference on Counterfactual Distributions." *Econometrica*. 2013. 81(6, November), pp. 2205-2268.
- [16] Chernozhukov, Victor, Fernandez-Val, Ivan, Han, Sukjin, and Kowalski, Amanda. "CQIV: Stata Module to Perform Censored Quantile Instrumental Variable Regression." <http://EconPapers.repec.org/RePEc:boc:bocode:s457478>.
- [17] Chernozhukov, Victor, and Hansen, Christian. "Instrumental variable quantile regression: A robust inference approach." *Journal of Econometrics*. January 2008. 142(1). pp.379-398.

- [18] 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.
- [19] Chesher, A. "Identification in Nonseparable Models." *Econometrica*, 2003, 71(5), pp. 1405-1441.
- [20] Deaton, Angus and Muelbauer, John. *Economics and consumer behavior*. Cambridge University Press. 1980.
- [21] Fitzenberger, Bernd. "A Guide to Censored Quantile Regressions." In *Handbooks of Statistics: Robust Inference*, Vol. 15, eds. G. S. Maddala and C. R. Rao. Amsterdam: North-Holland. pp. 405-437.
- [22] 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.
- [23] Frisch, R. "Circulation Planning: : Proposal For a National Organization of a Commodity and Service Exchange." *Econometrica*. 1934. 2(3), 258-336.
- [24] Gorman, W.M. "Separable Utility and Aggregation." *Econometrica*. 1959 27(3), 469-481.
- [25] Hahn, J. (1997): "Bayesian Bootstrap of the Quantile Regression Estimator: A Large Sample Study," *International Economic Review* 38, pp. 795-808.
- [26] Hausman, Jerry A. "Specification Tests in Econometrics." *Econometrica*. 1978. 46(6). pp. 1251-71.
- [27] Hausman, Jerry, Newey, Whitney, and Powell, James. "Nonlinear Errors in Variables Estimation of Some Engel Curves." *Journal of Econometrics*. 1995. 65. pp. 203-233.
- [28] Heckman, James J. "Sample Selection Bias as a Specification Error" *Econometrica*. 1979. 47(1), 153-161.
- [29] Hong, Han, and Tamer, Elie. "Inference in Censored Models with Endogenous Regressors." *Econometrica*. 2003. 71(3), 905-932.
- [30] Honoré, Bó, Khan, Shakeeb, and Powell, James L. "Quantile Regression Under Random Censoring." *Journal of Econometrics*. 2002. 109, 67-105.
- [31] Imbens, Guido W., and Newey, Whitney K.. "Identification and Estimation of Triangular Simultaneous Equations Models without Additivity." NBER Technical Working Paper 285. 2002.
- [32] Imbens, Guido W., and Newey, Whitney K.. "Identification and Estimation of Triangular Simultaneous Equations Models without Additivity." *Econometrica*. 2009. 77(5) 1481-1512.
- [33] Jun, Sung Jae. "Local structural quantile effects in a model with a nonseparable control variable." *Journal of Econometrics*. 2009. 151(1) 82-97.
- [34] Khan, Shakeeb, and Powell, James L. "Two-step estimation of semiparametric censored regression models." *Journal of Econometrics*. 2001. 103(1-2) 73-110.
- [35] Khan, Shakeeb, and Tamer, Elie. "Inference on Endogenously Censored Regression Models Using Conditional Moment Inequalities." *Journal of Econometrics*. 2009. 104-119.
- [36] Koenker, Roger. *Quantile Regression*. Cambridge University Press. 2005.
- [37] Koenker, Roger, and Bassett, Gilbert Jr. "Regression Quantiles." *Econometrica*, 1978, 46(1), pp. 33-50.
- [38] 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).
- [39] Kowalski, Amanda E. "Censored Quantile Instrumental Variable Estimates of the Price Elasticity of Expenditure on Medical Care." NBER Working Paper 15085. 2009.
- [40] Lee, Sokbae. "Endogeneity in quantile regression models: A control function approach." *Journal of Econometrics*. 2007. 141, pp. 1131-1158.
- [41] Lewbel, Arthur. "Entry for the New Palgrave Dictionary of Economics, 2nd Edition." Boston College. 2006.

- [42] Ma, Lingjie and Koenker, Roger. "Quantile regression methods for recursive structural equation models." *Journal of Econometrics*. 2006. 134(2). pp. 471-506.
- [43] Ma, Shuangge, and Kosorok, Michael. "Robust semiparametric M-estimation and the weighted bootstrap." *Journal of Multivariate Analysis*. 2005. 96(1). pp. 190-217.
- [44] Matzkin, Rosa L. "Nonparametric Identification." In *Handbook of Econometrics*, Vol. 6B, ed. by J. Heckman and E. Leamer. 2007. Amsterdam : Elsevier.
- [45] Newey, Whitney K. "Efficient Estimation of Limited Dependent Variable Models with Endogenous Explanatory Variables." *Journal of Econometrics*, 1987, 36, pp. 231-250.
- [46] Newey, Whitney K. "The asymptotic variance of semiparametric estimators. " *Econometrica*, 1994. 62 no. 6. 1349-1382.
- [47] Newey, Whitney K. "Two-step series estimation of sample selection models. " *Econometrics Journal*, 2009. 12, pp. 217-229.
- [48] 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.
- [49] Newey, Whitney K., Powell, James L., Vella, Francis. "Nonparametric Estimation of Triangular Simultaneous Equations Models." *Econometrica*. 1999. 67(3), 565-603.
- [50] Portnoy, Stephen. "Censored Regression Quantiles." *Journal of the American Statistical Association*. 2003. 98(464). pp-1001-1012.
- [51] Powell, James L. "Censored Regression Quantiles." *Journal of Econometrics*. 1986. 23. pp-143-155.
- [52] Powell, James L. "Least absolute deviations estimation for the censored regression model." *Journal of Econometrics*, 1984, 25(3), pp. 303-325.
- [53] 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
- [54] Rothe, Christoph. "Semiparametric Estimation of Binary Response Models with Endogenous Regressors." *Journal of Econometrics*, 2009, 153, pp. 51-64.
- [55] 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.
- [56] van der Vaart. *Asymptotic Statistics*. Cambridge University Press. 1998.
- [57] van der Vaart, A.W. and Wellner, Jon A. *Weak convergence and empirical processes*. Springer. 1996.
- [58] Wooldridge, Jeffrey M. *Econometric Analysis of Cross Section and Panel Data*. MIT Press. Cambridge, MA. Second Edition. 2010.

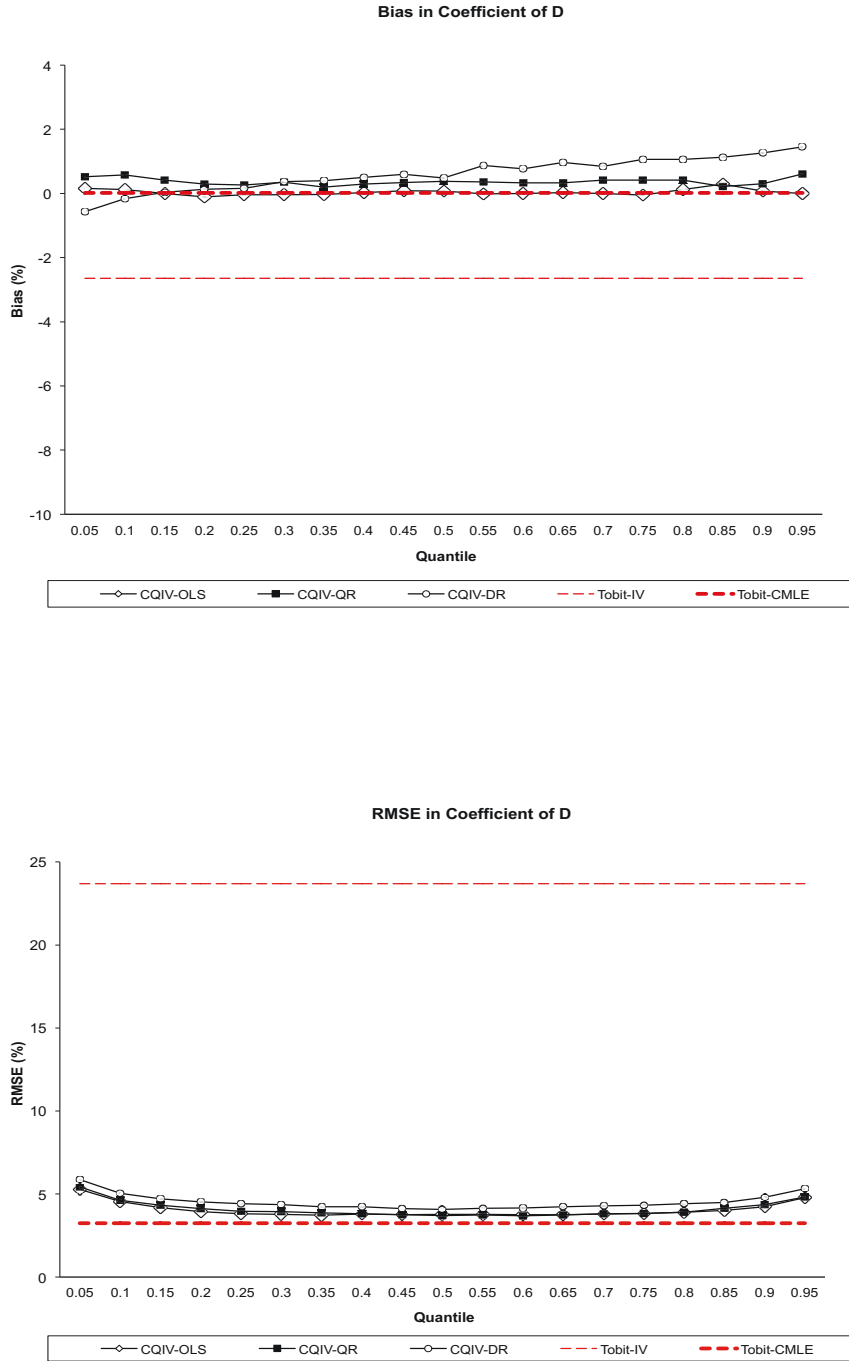


FIGURE 1. Tobit design: Mean bias and rmse of tobit and cqiv estimators. Results obtained from 1,000 samples of size $n = 1,000$.

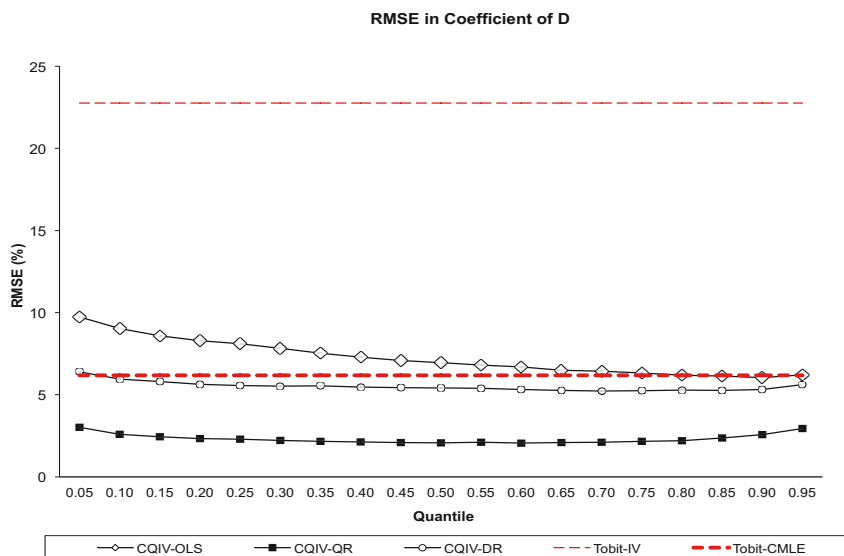
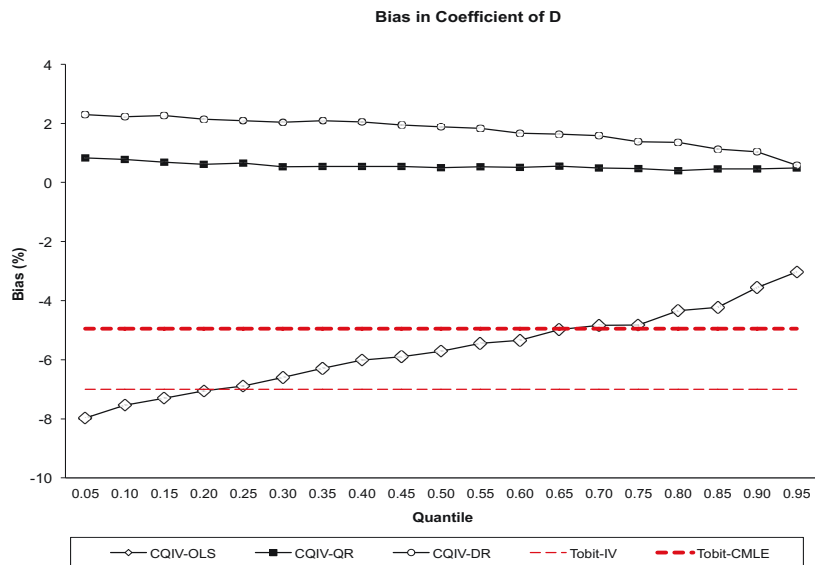


FIGURE 2. Design with heteroskedastic first stage: Mean bias and rmse of tobit and cqiv estimators. Results obtained from 1,000 samples of size $n = 1,000$.

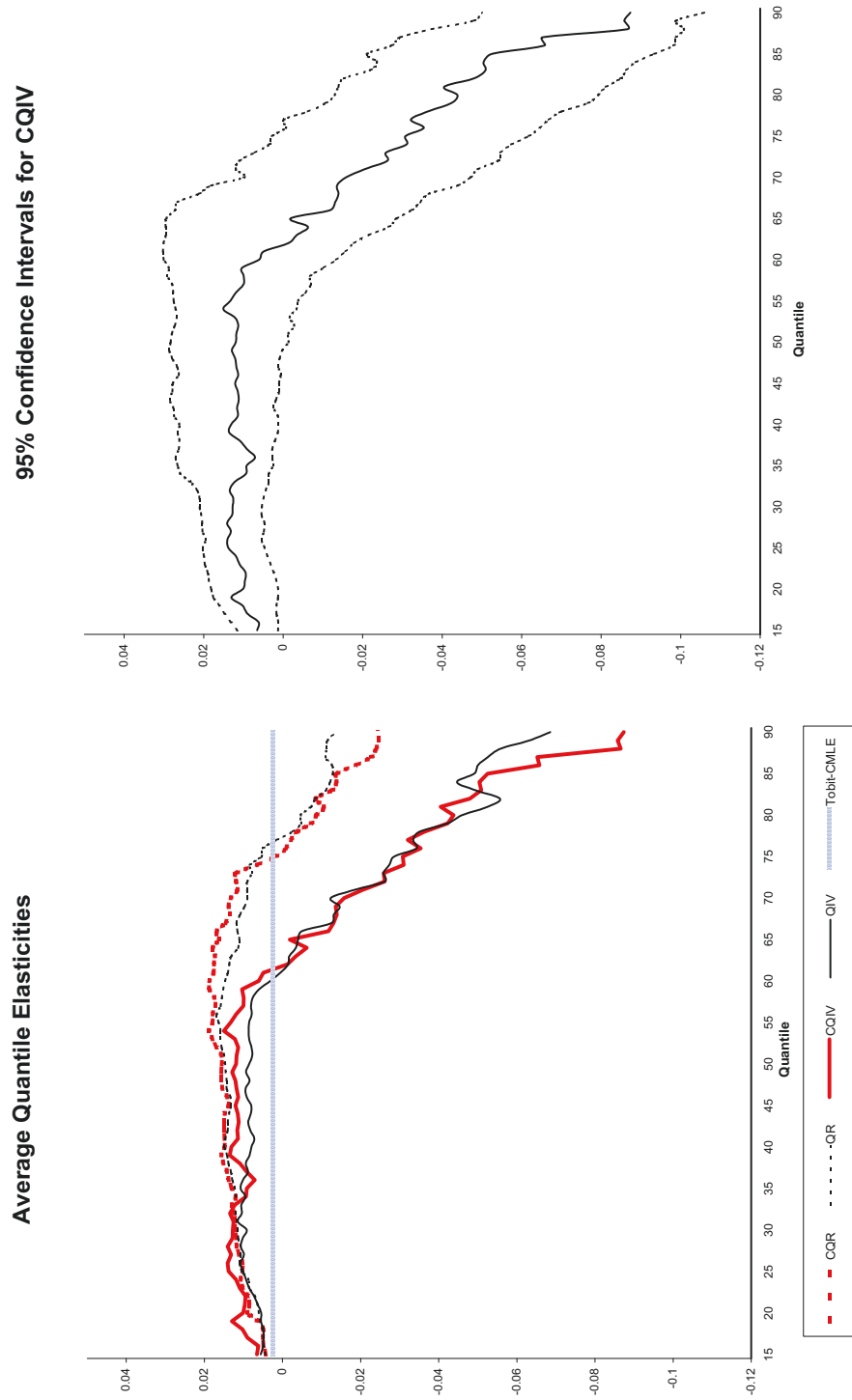


FIGURE 3. Estimates and 95% pointwise confidence intervals for average quantile expenditure elasticities. The intervals are obtained by weighted bootstrap with 200 replications and exponentially distributed weights.

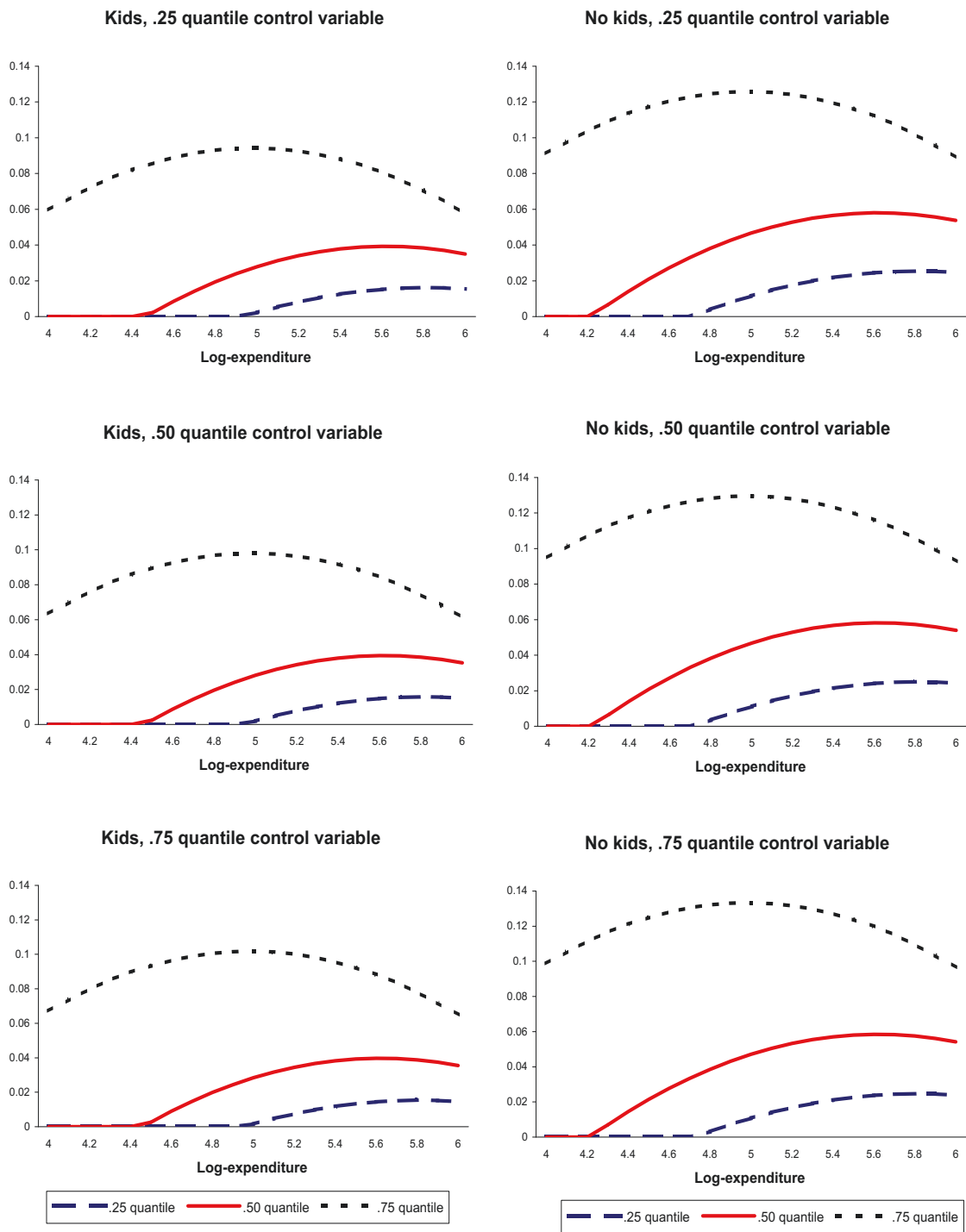


FIGURE 4. Family of Engel curves: each panel plots Engel curves for the three quantiles of alcohol share.

Table 1: Sensitivity of 3-step CQIV-QR to the cut-offs for the selectors

		q ₀ = 5			q ₀ = 10			q ₀ = 15		
		q ₁ = 1	q ₁ = 3	q ₁ = 5	q ₁ = 1	q ₁ = 3	q ₁ = 5	q ₁ = 1	q ₁ = 3	q ₁ = 5
A. Tobit Design										
Bias (%)	0.05	0.24	0.33	0.32	0.24	0.29	0.38	0.25	0.32	0.39
	0.1	0.35	0.36	0.34	0.35	0.38	0.39	0.34	0.37	0.38
	0.25	0.21	0.23	0.23	0.23	0.24	0.25	0.22	0.23	0.24
	0.5	0.29	0.27	0.26	0.26	0.28	0.30	0.27	0.29	0.29
	0.75	0.32	0.34	0.36	0.33	0.34	0.38	0.30	0.33	0.37
	0.9	0.35	0.34	0.32	0.33	0.36	0.36	0.32	0.35	0.36
	0.95	0.46	0.46	0.44	0.45	0.49	0.49	0.42	0.48	0.49
	RMSE (%)	0.05	5.43	5.43	5.47	5.43	5.42	5.54	5.42	5.47
	0.1	4.63	4.61	4.66	4.64	4.61	4.66	4.65	4.64	4.71
	0.25	3.94	3.96	3.99	3.95	3.96	4.00	3.94	3.97	4.00
	0.5	3.70	3.71	3.71	3.70	3.70	3.73	3.70	3.70	3.73
	0.75	3.81	3.83	3.86	3.83	3.84	3.86	3.83	3.82	3.84
	0.9	4.37	4.38	4.37	4.35	4.37	4.36	4.37	4.36	4.35
	0.95	4.79	4.81	4.83	4.80	4.82	4.86	4.80	4.83	4.89
B. Design with Heteroskedastic First Stage										
Bias (%)	0.05	0.65	0.69	0.70	0.67	0.69	0.71	0.70	0.72	0.72
	0.1	0.75	0.75	0.75	0.75	0.74	0.76	0.74	0.74	0.74
	0.25	0.52	0.53	0.55	0.52	0.54	0.55	0.52	0.54	0.55
	0.5	0.42	0.43	0.45	0.42	0.43	0.45	0.43	0.43	0.45
	0.75	0.43	0.45	0.45	0.44	0.45	0.46	0.44	0.44	0.45
	0.9	0.38	0.36	0.39	0.38	0.38	0.41	0.41	0.36	0.39
	0.95	0.41	0.40	0.40	0.44	0.41	0.40	0.46	0.42	0.42
	RMSE (%)	0.05	2.99	3.00	3.04	3.02	3.02	3.06	3.02	3.02
	0.1	2.57	2.59	2.62	2.58	2.59	2.63	2.59	2.61	2.63
	0.25	2.30	2.30	2.32	2.30	2.30	2.34	2.30	2.29	2.32
	0.5	2.07	2.07	2.10	2.07	2.06	2.10	2.07	2.07	2.10
	0.75	2.14	2.15	2.17	2.15	2.16	2.17	2.15	2.16	2.18
	0.9	2.55	2.56	2.58	2.55	2.57	2.57	2.56	2.58	2.57
	0.95	2.93	2.94	2.93	2.95	2.95	2.93	2.93	2.95	2.94

Notes 1,000 simulations.

Table 2: Diagnostic tests for 3-step CQIV-QR

Quantile-uncensored observations							Powell objective function	
u	J* (%)	J ₀ (%)	J ₁ (%)	J ₀ ∩ J ^{*c} (%)	J ₁ ∩ J ^{*c} (%)	J ₀ ∩ J ₁ ^c (%)	P ₁ < P ₀ (%)	P ₂ < P ₁ (%)
A . Tobit design								
0.05	52.4 (0.9)	48.2 (1.0)	51.7 (1.3)	1.7 (0.2)	1.7 (0.4)	0.0 (0.2)	71.9 (45.0)	16.0 (36.7)
0.1	54.5 (0.9)	50.0 (0.9)	53.7 (1.1)	1.6 (0.1)	1.6 (0.3)	0.0 (0.1)	79.9 (40.1)	21.6 (41.2)
0.25	58.1 (0.7)	53.2 (0.5)	57.2 (1.0)	1.5 (0.1)	1.4 (0.2)	0.0 (0.0)	81.5 (38.8)	30.3 (46.0)
0.5	62.0 (0.7)	56.7 (0.3)	61.0 (0.9)	1.4 (0.1)	1.3 (0.2)	0.0 (0.0)	83.8 (36.9)	32.8 (47.0)
0.75	66.0 (0.7)	60.2 (0.5)	64.8 (0.9)	1.3 (0.1)	1.2 (0.2)	0.0 (0.0)	81.4 (38.9)	27.8 (44.8)
0.9	69.4 (0.8)	63.2 (0.8)	68.2 (1.0)	1.2 (0.1)	1.1 (0.2)	0.0 (0.0)	80.8 (39.4)	22.5 (41.8)
0.95	71.4 (0.9)	65.0 (1.0)	70.2 (1.2)	1.2 (0.1)	1.1 (0.2)	0.0 (0.0)	76.3 (42.5)	19.2 (39.4)
B. Design with Heteroskedastic First Stage								
0.05	53.9 (0.9)	49.6 (0.9)	53.4 (1.2)	1.7 (0.1)	1.8 (0.3)	0.0 (0.1)	77.5 (41.8)	12.6 (33.2)
0.1	55.7 (0.8)	51.1 (0.7)	55.1 (1.1)	1.7 (0.1)	1.6 (0.3)	0.0 (0.1)	80.7 (39.5)	20.0 (40.0)
0.25	58.7 (0.7)	53.8 (0.5)	58.0 (0.9)	1.6 (0.1)	1.5 (0.2)	0.0 (0.0)	82.4 (38.1)	26.9 (44.4)
0.5	62.0 (0.6)	56.7 (0.3)	61.2 (0.8)	1.5 (0.1)	1.4 (0.1)	0.0 (0.0)	84.1 (36.6)	27.3 (44.6)
0.75	65.3 (0.7)	59.7 (0.5)	64.4 (0.8)	1.4 (0.1)	1.4 (0.1)	0.0 (0.0)	81.7 (38.7)	26.6 (44.2)
0.9	68.3 (0.8)	62.3 (0.7)	67.3 (1.0)	1.4 (0.1)	1.3 (0.2)	0.0 (0.0)	82.0 (38.4)	22.7 (41.9)
0.95	70.0 (0.9)	63.8 (0.9)	69.0 (1.1)	1.3 (0.1)	1.3 (0.2)	0.0 (0.0)	79.0 (40.8)	15.7 (36.4)

Notes 1,000 simulations. The entries of the table are simulation means, with standard deviations in parentheses. J* is the set of quantile-uncensored observations; J₀ is the set of observations selected in step 1; J₁ is the set of observations selected in step 2, P₀ is the value of Powell objective function for the 2-step estimator, P₁ is the value of the Powell objective function for the 3-step estimator, and P₂ is the value of the objective function for the 4-step estimator. The cut-offs for the selectors are q₀ = 10, q₁ = 3, and q₂ = 3.