Smoothed instrumental variables quantile regression, with estimation of quantile Euler equations*

Luciano de Castro† Antonio F. Galvao‡ David M. Kaplan§

July 10, 2017

Abstract

This paper develops theory for feasible estimation and testing of finite-dimensional parameters identified by general conditional quantile restrictions. This includes instrumental variables nonlinear quantile regression as a special case, under much weaker assumptions than previously seen in the literature. More specifically, we consider a set of unconditional moments implied by the conditional quantile restrictions and provide conditions for local identification. Since estimators based on the sample moments are generally impossible to compute numerically in practice, we study a feasible estimator based on *smoothed* sample moments. We establish consistency and asymptotic normality under general conditions that allow for weakly dependent data and nonlinear structural models, and we explore options for testing general nonlinear hypotheses. Simulations with iid and time series data illustrate the finite-sample properties of the estimators and tests. Our in-depth empirical application concerns the consumption Euler equation derived from quantile utility maximization. Advantages of the quantile Euler equation include robustness to fat tails, decoupling of risk attitude from the elasticity of intertemporal substitution, and log-linearization without any approximation error. For the four countries we examine, the quantile estimates of discount factor and elasticity of intertemporal substitution are economically reasonable for a range of quantiles just above the median, even when two-stage least squares estimates are not reasonable.

Keywords: instrumental variables, nonlinear quantile regression, quantile utility maximization.
JEL classification codes: C31, C32, C36

---

*The authors would like to express their appreciation to seminar participants at the University of Illinois at Urbana–Champaign, the Pennsylvania State University, and the 2017 North American Summer Meeting of the Econometric Society, for helpful comments, references, and discussions. Code is provided for all methods, simulations, and applications.

†Department of Economics, University of Iowa. E-mail: `decastro.luciano@gmail.com`
‡Department of Economics, University of Arizona. E-mail: `agalvao@email.arizona.edu`
§Department of Economics, University of Missouri. E-mail: `kaplandm@missouri.edu`

1 Introduction

Since the seminal work of Koenker and Bassett (1978), quantile regression (QR) has attracted considerable interest in statistics and econometrics. QR estimates conditional quantile functions that provide insight into heterogeneous effects of policy variables. This is especially valuable for program evaluation studies, where these methods help analyze how treatments or social programs affect the outcome's distribution. Nevertheless, endogeneity has been a pervasive concern in economics due to simultaneous causality, omitted variables, measurement error, self-selection, and estimation of equilibrium conditions, among other causes. Extending the standard QR, Chernozhukov and Hansen (2005, 2006, 2008) present results on identification, estimation, and inference for an instrumental variables QR (IVQR) model that allows for endogenous regressors.[1] However, computational difficulties have limited practical estimators to linear models with iid data (discussed below).

Under much weaker conditions than prior IVQR papers, we develop theory for a feasible smoothed estimator, including results on identification, estimation, and inference.[2] We consider a set of unconditional moments implied by a general parametric conditional quantile restriction. For identification, we provide sufficient conditions for local identification based on these moments. For estimation, since unsmoothed sample moments are generally intractable, we study a smoothed estimator that computes quickly and may have improved precision (Kaplan and Sun, 2017). Unlike prior IVQR estimation papers, we allow for weakly dependent data and nonlinear structural models when establishing consistency and asymptotic normality. For inference, we implement Wald and bootstrap approaches that allow weakly dependent data, unlike other subvector inference methods in the literature. In addition to their immediate value, our estimator and theoretical results are a helpful step toward full generalized method of moments (GMM) estimation and inference (including overidentification testing), which we pursue in a separate paper.

Our in-depth empirical study estimates a quantile Euler equation using aggregate time series data. This equation is derived from a quantile utility maximization model. The

[1]We refer to Chernozhukov, Hansen, and Wüthrich (2017) for an overview of IVQR. They discuss alternative, complementary QR models with endogeneity in Section 1.2.5.

[2]The methods developed in this paper are also related to those for semiparametric and nonparametric models. Identification, estimation, and inference of general (non-smooth) conditional moment restriction models have received much attention in the econometrics literature, as in Newey and McFadden (1994, §7), Chen, Linton, and van Keilegom (2003), Chen and Pouzo (2009, 2012), and Chen and Liao (2015), for example. However, theoretical results are only for unsmoothed estimators that are often not computationally feasible in practice.

quantile utility maximization model is an interesting alternative to the standard expected utility model because it is robust to fat tails and allows heterogeneity through the quantiles (partially decoupling the elasticity of intertemporal substitution from risk attitude), and the resulting Euler equation does not suffer from any approximation error when log-linearized.[3] Quantile preferences were first studied by Manski (1988) and were axiomatized by Chambers (2009) and Rostek (2010). De Castro and Galvao (2017) use quantile preferences in a dynamic economic setting and provide a comprehensive analysis of a dynamic rational quantile model. They derive the policy function (Euler equation) as a nonlinear conditional quantile restriction. Structural parameters from the quantile utility problem can be estimated by our smoothed IVQR method. Using a standard economic model of intertemporal allocation of consumption, we estimate the elasticity of intertemporal substitution (EIS) coefficient. From this, we employ a variation of the standard economy model of Lucas (1978) where the economic agents decide on the intertemporal consumption and savings (assets to hold) over an infinity horizon economy, maximizing the discounted present value of the stream of quantile utilities, subject to a linear budget constraint. The decision generates a policy function, which is used to estimate the parameters of interest for a given utility function. Numerous papers have estimated the EIS, e.g., Hansen and Singleton (1983), Hall (1988), Campbell and Mankiw (1989), Ogaki and Reinhart (1998), and Yogo (2004). For the four countries we study, the SIVQR estimates of the discount factor and EIS are economically reasonable for a range of quantiles above the median, including cases where the 2SLS estimates are not reasonable.

For IVQR estimation of the model in Chernozhukov and Hansen (2005), the literature lacks results for feasible estimators allowing nonlinear structural models and/or dependent data.[4] The following are iid sampling assumptions: Condition (i) on p. 310 in Chernozhukov and Hong (2003), Assumption 2.R1 in Chernozhukov and Hansen (2006), and Assumption 1 in Kaplan and Sun (2017). A nonlinear structural model is allowed by the computationally demanding Markov Chain Monte Carlo estimator in Chernozhukov and Hong (2003, Ex. 3, p. 297ff.), but linear-in-parameters models are required in (3.4) in Chernozhukov and Hansen (2006) and Assumption 1 in Kaplan and Sun (2017). Additionally, Chernozhukov and Hansen (2006) note, “The computational advantages of our estimator rapidly diminish as the number of endogenous variables increases” (p. 501). Even if only one observed variable

[3]Heavy tails in consumption data have been documented recently by Toda and Walsh (2015, 2017).

[4]For nonlinear QR (no IV), see Powell (1994, §2.2), Oberhofer and Haupt (2016), and references therein.

is endogenous, this restriction limits the use of interactions and transformations (like polynomial terms). Chen and Lee (2017) propose an estimator for linear-in-parameters IVQR models using mixed integer quadratic programming, but computation is very slow: with only four parameters and $n = 100$ observations, their Table 1 shows average computation times for IV median regression exceeding five minutes. From a Bayesian perspective, Lancaster and Jun (2010) allow nonlinear models but only iid data (and require Markov Chain Monte Carlo computation). We relax both iid sampling and linearity in our formal results, while maintaining the computational simplicity, speed, and scalability of the method in Kaplan and Sun (2017).

For IVQR inference, too, the literature generally lacks results for dependent data, although most methods are robust to weak identification (unlike ours). For inference on the *full* parameter vector, iid sampling is not required by either a χ^2 test as in Kaplan and Sun (2017) or the simulation-based method of Chernozhukov, Hansen, and Jansson (2009), but these may be very conservative for subvector inference (using projection), and the latter still restricts sampling (their A6, p. 95) in a way that is violated in our time series simulation DGPs in Section 4, for example. Other approaches require iid sampling, as in Assumption R1 of Chernozhukov and Hansen (2008, p. 384), p. 122 of Jun (2008, "the data. . . are assumed to be iid"), and Assumption 10(i) of Andrews and Mikusheva (2016, p. 1597). Our methods fill the gap of time series IVQR subvector inference when instruments are strong, as in our empirical application.

Historically, the idea of GMM with smoothed IVQR moment conditions was proposed first in unpublished notes by MaCurdy and Hong (1999), mentioned later in (also unpublished) MaCurdy and Timmins (2001, §2.4) and the handbook chapter by MaCurdy (2007, §5).[5] Whang (2006) and Otsu (2008) use moment smoothing for empirical likelihood QR. The closely related idea of smoothing non-differentiable *objective functions* goes back to Amemiya (1982, §3), if not earlier, and such smoothing has been employed in the smoothed maximum score estimator of Horowitz (1992) as well as for QR by Horowitz (1998), Galvao and Kato (2016), and Fernandes, Guerre, and Horta (2017), among others. Kaplan and Sun (2017, §2.2) argue that smoothing the moment conditions (instead of objective function) is better even for QR, in terms of simplicity and bias.

Section 2 presents our estimator, with identification and asymptotic properties. Inference

[5] Among others, Buchinsky (1998, §III.A) discusses QR (but not IVQR) as GMM.

is described in Section 3. Section 4 contains simulation results. In Section 5 we illustrate the new approach empirically. Section 6 suggests directions for future research. The appendix collects all proofs.

We conclude this introduction with some remarks about the notation. Random variables and vectors are uppercase (Y, X, etc.), while non-random values are lowercase (y, x); for vector/matrix multiplication, all vectors are treated as column vectors. Also, $\mathbb{1}\{\cdot\}$ is the indicator function, $\mathrm{E}(\cdot)$ expectation, $\mathrm{Q}_\tau(\cdot)$ the τ-quantile, $\mathrm{P}(\cdot)$ probability, and $\mathrm{N}(\mu,\sigma^2)$ the normal distribution. For vectors, $\|\cdot\|$ is the Euclidean norm. Acronyms used include those for central limit theorem (CLT), continuous mapping theorem (CMT), elasticity of intertemporal substitution (EIS), [smoothed] instrumental variables quantile regression ([S]IVQR), mean value theorem (MVT), probability density function (PDF), uniform law of large numbers (ULLN), and weak law of large numbers (WLLN).

2 Identification and large-sample properties

For a given quantile index $\tau \in (0,1)$, we consider estimating the parameter vector $\beta_{0\tau} \in \mathcal{B} \subseteq \mathbb{R}^{d_\beta}$ using the moment condition

$$0 = \mathrm{E}\bigl\{Z_i\bigl[\mathbb{1}\bigl\{\Lambda\bigl(Y_i, X_i, \beta_{0\tau}\bigr) \le 0\bigr\} - \tau\bigr]\bigr\}, \tag{2.1}$$

with endogenous vector $Y_i \in \mathcal{Y} \subseteq \mathbb{R}^{d_Y}$, full instrument vector $Z_i \in \mathcal{Z} \subseteq \mathbb{R}^{d_Z}$ that contains $X_i \in \mathcal{X} \subseteq \mathbb{R}^{d_X}$ as a subset, and known function $\Lambda(\cdot)$; $\mathbb{1}\{\cdot\}$ is the indicator function. Kaplan and Sun (2017) consider a special case of (2.1) with $\Lambda\bigl(Y, X, \beta\bigr) = Y_1 - Y_{-1}^\top \beta_1 - X^\top \beta_2$, where $Y_{-1} = (Y_2, \ldots, Y_{d_Y})$.

We take (2.1) with the given Z_i as our starting point for estimation and asymptotic results in Sections 2.3 and 2.4. For determining the efficient set of instruments (given a *conditional* quantile restriction), we refer to Newey (2004) or Newey and Powell (1990), for example. Alternatively, Otsu (2008) achieves asymptotic efficiency via conditional empirical likelihood for nonlinear QR with iid data, but not IVQR.

For this paper, we restrict attention to the just-identified case where $d_Z = d_\beta$. Computation is more robust in the just-identified case because it is easy to check whether $\hat{\beta}$ indeed solves the sample moment conditions, whereas it is practically impossible to know whether a local minimum is the global minimum.[6] In the overidentified case, one can always transform

[6]Footnote 5 in Chernozhukov and Hong (2003) notes that the smoothing (similar to ours) in the unpublished MaCurdy and Timmins (2001) "does not eliminate non-convexities and local optima."

the original moment conditions into a set of d_β moment conditions by pre-multiplying Z_i by a $d_\beta \times d_Z$ matrix to get $\tilde{Z}_i \in \mathbb{R}^{d_\beta}$, which replaces Z_i in the moment conditions. See Kaplan and Sun (2017, §2.1) for additional discussion of the overidentified case, including efficiency considerations and practical issues (estimation error) in nonparametrically estimating optimal instruments in this model. Although it may be possible to improve efficiency while maintaining computational robustness in the overidentified case, we leave a thorough study as future research.

Before discussing estimation and asymptotics, we discuss identification in Section 2.1.

2.1 Structural motivation and identification

The parameter $\beta_{0\tau}$ is "locally identified" if there exists a neighborhood of $\beta_{0\tau}$ within which only $\beta_{0\tau}$ satisfies (2.1). Given (2.1), local identification of $\beta_{0\tau}$ holds if the partial derivative matrix of the right-hand side of (2.1) with respect to the β argument is full rank; see Chen, Chernozhukov, Lee, and Newey (2014, p. 787), for example. This full rank condition is formally stated below in Assumption A9(ii). The following proposition states the local identification result.

Proposition 2.1. *Given* (2.1) *and (the full rank) Assumption A9(ii),* $\beta_{0\tau}$ *is locally identified.*

Global identification is notoriously more difficult to establish and depends on the application; see additional remarks on Assumption A3 below.

Usually the unconditional moments in (2.1) come from a conditional quantile restriction implied by a structural model. We discuss two examples.

One example of a structural model is a random coefficients model similar to Chernozhukov and Hansen (2005, 2006).[7] By the law of iterated expectations, a sufficient condition for our moment condition in (2.1) is the conditional quantile restriction

$$\tau = \mathrm{E}\big[\mathbb{1}\big\{\Lambda\big(Y, X, \beta_{0\tau}\big) \le 0\big\} \mid Z\big] = \mathrm{P}\big[\Lambda\big(Y, X, \beta_{0\tau}\big) \le 0 \mid Z\big]. \tag{2.2}$$

If $Y = (\tilde{Y}, D)$, where $\tilde{Y}$ is an outcome and D is an endogenous "treatment" regressor, (2.2) can be derived from a structural random coefficient model as in Theorem 1 of Chernozhukov and Hansen (2005). That is, given $X = x$, with unobserved $U_d \sim \mathrm{Unif}(0,1)$, the potential outcomes are $\tilde{Y}_d = q\big(d, x, \beta_0(U_d)\big)$ for each possible $D = d$; $q(\cdot)$ is the (known) structural

[7] Our Y is equivalent to their Y and D combined; our Z combines their X and Z.

function, and $\Lambda(Y, X, \beta) = \tilde{Y} - q(D, X, \beta)$, so $\Lambda(Y, X, \beta) \leq 0$ is equivalent to $\tilde{Y} \leq q(D, X, \beta)$. Chernozhukov and Hansen (2005) show how this model combines with other assumptions to yield (a.s.) the conditional quantile restriction $\mathrm{P}\{\tilde{Y} \leq q(D, X, \beta_{0\tau}) \mid Z\} = \tau$, which is a special case of (2.2). For identification, Chernozhukov and Hansen (2005) consider a general nonparametric model (see also Chen et al., 2014), but for estimation, Chernozhukov and Hansen (2006) assume $q(\cdot)$ is linear-in-parameters in their equation (3.4).

The unconditional moments in (2.1) may hold even if condition (2.2) does not, since (2.2) is stronger than (2.1), i.e., (2.2) implies (2.1). If the conditional model (2.2) is misspecified, then $\beta_{0\tau}$ is a pseudo-true population parameter satisfying (2.1) whose interpretation depends on the economic model. For example, Angrist, Chernozhukov, and Fernández-Val (2006) interpret the linear QR parameter when the conditional quantile function is misspecified.

Another example of a structural model applies to our empirical application in Section 5. Under certain assumptions, if individuals maximize the τ-quantile of utility instead of expected utility, then the resulting consumption Euler equation can be written in the form

$$\mathrm{Q}_\tau[\Lambda(Y_t, X_t, \beta_{0\tau}) \mid \Omega_t] = 0, \tag{2.3}$$

where $\beta_{0\tau}$ is the unknown parameter vector (describing preferences), Ω_t is the information set at time t, (Y_t, X_t) are observable variables that may not all yet be realized (e.g., C_{t+1}), $\Lambda(\cdot)$ is a function known up to the finite-dimensional parameter $\beta_{0\tau}$, and the conditional quantile operator $\mathrm{Q}_\tau[W_t \mid \Omega_t]$ denotes the conditional τ-quantile of W_t given Ω_t. This is not simply a nonlinear quantile regression of a scalar outcome Y_t on regressor vector X_t. Using the isoelastic utility function, the quantile Euler equation in (5.17) in Section 5 is

$$\mathrm{Q}_\tau\left[\beta_{0\tau}(1 + r_{t+1})(C_{t+1}/C_t)^{-\gamma_{0\tau}} - 1 \mid \Omega_t\right] = 0.$$

2.2 Estimator and assumptions

We introduce some notation and the smoothed IVQR (SIVQR) estimator, followed by assumptions. Let the population map $M : \mathcal{B} \times \mathcal{T} \mapsto \mathbb{R}^{d_Z}$ be

$$M(\beta, \tau) \equiv \mathrm{E}[g_i^u(\beta, \tau)], \tag{2.4}$$

$$g_i^u(\beta, \tau) \equiv g^u(Y_i, X_i, Z_i, \beta, \tau) \equiv Z_i[\mathbb{1}\{\Lambda(Y_i, X_i, \beta) \leq 0\} - \tau], \tag{2.5}$$

where superscript "u" denotes "unsmoothed." The population moment condition (2.1) is

$$0 = M(\beta_{0\tau}, \tau). \tag{2.6}$$

For the corresponding sample moments, the population expectation $\mathrm{E}(\cdot)$ is replaced by the sample expectation $\hat{\mathrm{E}}(\cdot)$, i.e., the sample average. Analogous to (2.4), using (2.5), the unsmoothed *sample* moment map is

$$\hat{M}_n^u(\beta,\tau) \equiv \hat{\mathrm{E}}[g^u(Y,X,Z,\beta,\tau)] \equiv \frac{1}{n}\sum_{i=1}^{n} g_i^u(\beta,\tau). \tag{2.7}$$

Even if population system (2.6) has a unique solution, the unsmoothed system $\hat{M}_n^u(\beta,\tau)=0$ may have zero or multiple solutions.[8] Although this issue can be worked around theoretically, smoothing addresses it directly.

The SIVQR estimator we consider in this paper is the solution to the system of smoothed sample moments, shown in (2.9). With smoothing (no "u" superscript), the sample analogs of (2.4)–(2.6) are

$$\begin{aligned} g_{ni}(\beta,\tau) &\equiv g_n(Y_i,X_i,Z_i,\beta,\tau) \equiv Z_i\big[\tilde{I}\big(-\Lambda(Y_i,X_i,\beta)/h_n\big)-\tau\big], \\ \hat{M}_n(\beta,\tau) &\equiv \frac{1}{n}\sum_{i=1}^{n} g_{ni}(\beta,\tau), \end{aligned} \tag{2.8}$$

$$\hat{M}_n(\hat{\beta}_\tau,\tau) = 0, \tag{2.9}$$

where h_n is a bandwidth (sequence) and $\tilde{I}(\cdot)$ is a smoothed version of the indicator function $\mathbb{1}\{\cdot>0\}$. As a mnemonic, I may stand for "indicator-like function" or "integral of kernel." An example function $\tilde{I}(\cdot)$ is shown in Figure 1; it is the integral of a fourth-order polynomial kernel and has been used by Horowitz (1998), Whang (2006), and Kaplan and Sun (2017). The double subscript on g_{ni} is a reminder that we have a triangular array setup because g_{ni} depends on the bandwidth sequence h_n in addition to (Y_i,X_i,Z_i). The right-hand side of (2.9) can be relaxed to $o_p(n^{-1/2})$.

Thanks to smoothing, (2.9) always has a solution. Computationally, the solution is easy to find if $\Lambda(\cdot)$ is differentiable in β (unless h_n is extremely close to zero), and it is easy to check whether some $\hat{\beta}_\tau$ returned by a numerical solver is indeed the solution: just plug it in to check if (2.9) indeed holds. In contrast, it is generally impossible to know whether a local minimum is a global minimum.

The price of smoothing is finite-sample bias in the sample moments: $\mathrm{E}[\hat{M}_n(\beta_{0\tau},\tau)]\neq 0$, whereas for the unsmoothed version $\mathrm{E}[\hat{M}_n^u(\beta_{0\tau},\tau)]=0$. However, Kaplan and Sun (2017)

[8] For example, let $X=Z=1$, $\Lambda(Y,b)=Y-b$, and $n=2$, with observations Y_1 and Y_2. If $\tau=1/2$, then the equation is solved by any $Y_1\le b<Y_2$. If $\tau\neq 1/2$, then it is not solved by any b.

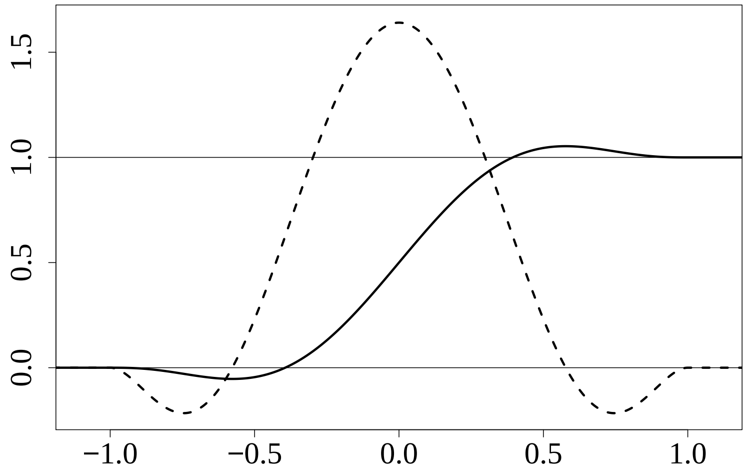

Figure 1: Graph of $\tilde{I}(u) = 0.5 + \frac{105}{64}\left(u - \frac{5}{3}u^3 + \frac{7}{5}u^5 - \frac{3}{7}u^7\right)$ (solid line) and its derivative (dashed), which is a fourth-order kernel.

suggest that an even larger decrease in variance results in an overall (higher-order) improvement in mean squared error in the linear/iid setting. Under much weaker assumptions, the following sections establish consistency and asymptotic normality of the SIVQR estimator in (2.9).

Different subsets of the following assumptions are used for different results.

Assumption A1. *For each observation i among n in the sample, endogenous vector $Y_i \in \mathcal{Y} \subseteq \mathbb{R}^{d_Y}$ and instrument vector $Z_i \in \mathcal{Z} \subseteq \mathbb{R}^{d_Z}$; a subset of Z_i is $X_i \in \mathcal{X} \subseteq \mathbb{R}^{d_X}$, with $d_X \leq d_Z$. The sequence $\{Y_i, Z_i\}$ is strictly stationary and weakly dependent.*

Assumption A2. *The function $\Lambda : \mathcal{Y} \times \mathcal{X} \times \mathcal{B} \mapsto \mathbb{R}$ is known and has (at least) one continuous derivative in its $\mathcal{B}$ argument for all $y \in \mathcal{Y}$ and $x \in \mathcal{X}$.*

Assumption A3. *The parameter space $\mathcal{B} \in \mathbb{R}^{d_\beta}$ is compact; $d_\beta = d_Z$. Given $\tau \in (0,1)$, the population parameter $\beta_{0\tau}$ is in the interior of $\mathcal{B}$ and uniquely satisfies the moment condition*

$$0 = \mathrm{E}\big[Z_i\big(\mathbb{1}\big\{\Lambda\big(Y_i, X_i, \beta_{0\tau}\big) \leq 0\big\} - \tau\big)\big]. \tag{2.10}$$

Assumption A4. *The matrix $\mathrm{E}\big(Z_i Z_i^\top\big)$ is positive definite (and finite).*

Assumption A5. *The function $\tilde{I}(\cdot)$ satisfies $\tilde{I}(u) = 0$ for $u \leq -1$, $\tilde{I}(u) = 1$ for $u \geq 1$, and $-1 \leq \tilde{I}(u) \leq 2$ for $-1 < u < 1$. The derivative $\tilde{I}'(\cdot)$ is a symmetric, bounded kernel function of order $r \geq 2$, so $\int_{-1}^{1} \tilde{I}'(u)\,du = 1$, $\int_{-1}^{1} u^k \tilde{I}'(u)\,du = 0$ for $k = 1, \ldots, r-1$, and $\int_{-1}^{1} |u^r \tilde{I}'(u)|\,du < \infty$ but $\int_{-1}^{1} u^r \tilde{I}'(u)\,du \neq 0$.*

Assumption A6. *The bandwidth sequence* h_n *satisfies* $h_n = o(n^{-1/(2r)})$.

Assumption A7. *Given any* $\beta \in \mathcal{B}$ *and almost all* $Z_i = z$ *(i.e., up to a set of zero probability), the conditional distribution of* $\Lambda(Y_i, X_i, \beta)$ *given* $Z_i = z$ *is continuous in a neighborhood of zero.*

Assumption A8. *For a fixed* $\tau \in (0,1)$*, using the definition in* (2.8)*,*

$$\sup_{\beta\in\mathcal{B}} \Big| \hat{M}_n(\beta,\tau) - \mathrm{E}[\hat{M}_n(\beta,\tau)] \Big| = o_p(1). \tag{2.11}$$

Assumption A9. *Let* $\Lambda_i \equiv \Lambda\big(Y_i, X_i, \beta_{0\tau}\big)$ *and* $D_i \equiv \nabla_\beta \Lambda\big(Y_i, X_i, \beta_{0\tau}\big)$*, using the notation*

$$\nabla_\beta \Lambda(y,x,\beta_0) \equiv \frac{\partial}{\partial\beta}\Lambda(y,x,\beta)\bigg|_{\beta=\beta_0}, \tag{2.12}$$

for the $d_\beta \times 1$ *partial derivative vector. Let* $f_{\Lambda|Z}(\cdot \mid z)$ *denote the conditional PDF of* Λ_i *given* $Z_i = z$*, and let* $f_{\Lambda|Z,D}(\cdot \mid z, d)$ *denote the conditional PDF of* Λ_i *given* $Z_i = z$ *and* $D_i = d$*. (i) For almost all* z *and* d*,* $f_{\Lambda|Z}(\cdot \mid z)$ *and* $f_{\Lambda|Z,D}(\cdot \mid z, d)$ *are at least* r *times continuously differentiable in a neighborhood of zero, where the value of* r *is from A5. For almost all* $z \in \mathcal{Z}$ *and* u *in a neighborhood of zero, there exists a dominating function* $C(\cdot)$ *such that* $\Big| f^{(r)}_{\Lambda|Z}(u \mid z)\Big| \le C(z)$ *and* $\mathrm{E}[C(Z)|Z|] < \infty$*. (ii) The matrix*

$$G \equiv \frac{\partial}{\partial\beta^\top}\mathrm{E}\big[Z_i \mathbb{1}\{\Lambda(Y_i,X_i,\beta) \le 0\}\big]\bigg|_{\beta=\beta_{0\tau}} = -\mathrm{E}\big\{Z_i D_i^\top f_{\Lambda|Z,D}(0 \mid Z_i, D_i)\big\} \tag{2.13}$$

is nonsingular.

Assumption A10. *A pointwise CLT applies to give*

$$\sqrt{n}\big\{\hat{M}_n(\beta_{0\tau},\tau) - \mathrm{E}[\hat{M}_n(\beta_{0\tau},\tau)]\big\} \xrightarrow{d} \mathrm{N}(0, \Sigma_{M\tau}). \tag{2.14}$$

Assumption A11. *Let* $Z_i^{(k)}$ *denote the* k*th element of* Z_i*, and similarly* $\beta^{(k)}$*. Let* G_{kj} *denote the row* k*, column* j *element of* G *(from A9). Assume*

$$-\frac{1}{nh_n}\sum_{i=1}^{n} \tilde{I}'\big(-\Lambda(Y_i,X_i,\tilde{\beta}_{\tau,k})/h_n\big) Z_i^{(k)} \frac{\partial}{\partial\beta^{(j)}}\Lambda(Y_i,X_i,\beta)\bigg|_{\beta=\tilde{\beta}_{\tau,k}} \xrightarrow{p} G_{kj}. \tag{2.15}$$

for each $k = 1,\ldots,d_\beta$ *and* $j = 1,\ldots,d_\beta$*, where each* $\tilde{\beta}_{\tau,k}$ *lies between* $\beta_{0\tau}$ *and* $\hat{\beta}_\tau$ *(defined in A3 and* (2.9)*, respectively).*

For transparency, A1 includes sampling assumptions that help establish the high-level assumptions A8, A10, and A11; see Appendix B. Assumption A2 is stronger than a nonparametric model but more general than the linear-in-parameters model of Chernozhukov and Hansen (2006, eqn. (3.4)) and Kaplan and Sun (2017, Assumption 1). Assumption A3 is a standard assumption in settings like GMM. Since global identification depends on the application (e.g., the $\Lambda(\cdot)$), it has a tradition of being assumed going back to Hansen (1982, Thm. 2.1(iii), p. 1035), but it may not be trivial to establish in some cases, especially with highly nonlinear $\Lambda(\cdot)$.[9] Assumption A4 matches Assumption 2(ii) in Kaplan and Sun (2017); it is relatively weak, imposing only a finite second moment on Z_i (and, unlike 2SLS, no moment restrictions on Y_i or X_i). Assumption A5 is essentially Assumption 4(i,ii) of Kaplan and Sun (2017).

Assumption A6 ensures that the asymptotic effect of smoothing is negligible. For consistency, this only requires $h_n \to 0$. For asymptotic normality, the stated $h_n = o(n^{-1/(2r)})$ is needed to prove Lemma 2.4, although this is still relatively mild: for example, with $r = 4$ (as in Figure 1), it requires $h_n \to 0$ faster than $n^{-1/8}$. To establish A11 under weak dependence, it seems $nh_n^2 \to \infty$ is also necessary; with iid sampling and a linear-in-parameters structural model, Kaplan and Sun (2017, proof of Lemma A.2) only require $nh_n \to \infty$ for this result. Note that the optimal bandwidth rate from Kaplan and Sun (2017) for a linear IVQR model with iid sampling is $h_n \asymp n^{-1/(2r-1)}$, which satisfies all aforementioned restrictions for any $r \geq 2$. Appendix C.1 contains details for practical bandwidth selection.

Assumption A7 can be checked easily in most cases. For example, if $\Lambda(Y_i, X_i, \beta) = \tilde{Y}_i - (D_i, X_i^\top)\beta$ and $Z^\top = (X^\top, \tilde{Z})$, then A7 is satisfied if the outcome Y has a continuous distribution conditional on almost all $(X^\top, \tilde{Z}) = (x^\top, \tilde{z})$. Assumption A8 generally requires some restriction on dependence and moments, but it is much weaker than the iid sampling assumption of Chernozhukov and Hansen (2006, Assumption 2.R1) or Kaplan and Sun (2017, Assumption 1). Assumption A9 is used for the asymptotic normality result; it generalizes parts of Assumptions 3 and 7 in Kaplan and Sun (2017) to our nonlinear model. The nonsingularity of G is also sufficient for local (but not global) identification (e.g., Chen et al., 2014, p. 787). The CLT in A10 is a high-level assumption, similar to condition (iv) in Theorem 7.2 of Newey and McFadden (1994), for example; like A8, it requires some restriction on dependence and moments but does not require iid sampling. Examples of more primitive sufficient

[9] In their Handbook chapter, Newey and McFadden (1994) write of “the difficulty of specifying primitive identification conditions for GMM” (p. 2120).

conditions for A8 and A10 are given later in Appendix B. Assumption A11 is actually a generalization of the consistency of Powell's estimator for the asymptotic covariance matrix of the usual quantile regression estimator, as detailed in Section 2.4. Assumption A11 embodies the stochastic equicontinuity that is often separately assumed, as in Theorem 7.2(v) in Newey and McFadden (1994), for example.

2.3 Consistency

To establish consistency, we use Theorem 5.9 in van der Vaart (1998), showing the two required conditions are satisfied here. One condition is an identification condition. The other requires uniform (in $\beta \in \mathcal{B}$) convergence in probability of the sample maps $\hat{M}_n(\beta, \tau)$ to the population map $M(\beta, \tau)$. No iid sampling assumption is required; the second assumption may be established under weak dependence.

A detailed example of primitive conditions for the high-level uniform weak law of large numbers assumed in Assumption A8 is given in Appendix B.1.

In addition to A8, we must show that the sequence of (non-random) maps $\mathrm{E}[\hat{M}_n(\beta, \tau)]$ converges to the desired population map $M(\beta, \tau)$, as in Lemma 2.2.

Lemma 2.2. *Under Assumptions A1–A7, for a fixed τ, using definitions in* (2.4) *and* (2.8),

$$\sup_{\beta \in \mathcal{B}} \left| \mathrm{E}[\hat{M}_n(\beta, \tau)] - M(\beta, \tau) \right| = o(1). \tag{2.16}$$

Lemma 2.2 is intuitive. Without smoothing, $M(\cdot) = \mathrm{E}[\hat{M}_n^u(\cdot)]$ for all n. With smoothing, we should expect this to hold asymptotically if the smoothing is asymptotically negligible. The next result establishes consistency of the proposed SIVQR estimator.

Theorem 2.3. *Under Assumptions A1–A8, the estimator from* (2.9) *is consistent:*

$$\hat{\beta}_\tau - \beta_{0\tau} = o_p(1). \tag{2.17}$$

2.4 Asymptotic normality

To establish asymptotic normality, smoothing facilitates the usual approach of expanding the sample moments around $\beta_{0\tau}$ because the smoothed sample moments are differentiable. Recall from (2.9) that $0 = \hat{M}_n(\hat{\beta}_\tau, \tau)$. Define

$$\nabla_{\beta^\top} \hat{M}_n(\beta_{0\tau}, \tau) \equiv \frac{\partial}{\partial \beta^\top} \hat{M}_n(\beta, \tau) \bigg|_{\beta = \beta_{0\tau}}. \tag{2.18}$$

Let $\hat{M}_n^{(k)}(\beta, \tau)$ refer to the kth element in the vector $\hat{M}_n(\beta, \tau)$, so $\nabla_{\beta^\top} \hat{M}_n^{(k)}(\beta_{0\tau}, \tau)$ is a row vector and $\nabla_\beta \hat{M}_n^{(k)}(\beta_{0\tau}, \tau)$ is a column vector. Define

$$\dot{M}_n(\tau) \equiv \left(\nabla_\beta \hat{M}_n^{(1)}(\tilde{\beta}_{\tau,1}, \tau), \ldots, \nabla_\beta \hat{M}_n^{(d_\beta)}(\tilde{\beta}_{\tau,d_\beta}, \tau)\right)^\top, \tag{2.19}$$

a $d_\beta \times d_\beta$ matrix with its first row equal to that of $\nabla_{\beta^\top} \hat{M}_n(\tilde{\beta}_{\tau,1}, \tau)$, its second row equal to that of $\nabla_{\beta^\top} \hat{M}_n(\tilde{\beta}_{\tau,2}, \tau)$, etc., where each vector $\tilde{\beta}_{\tau,k}$ lies on the line segment between $\beta_{0\tau}$ and $\hat{\beta}_\tau$. Due to smoothing, we can take a derivative (for any n) to get a mean value expansion, and then rearrange:

$$0 = \hat{M}_n(\beta_{0\tau}, \tau) + \dot{M}_n(\tau)(\hat{\beta}_\tau - \beta_{0\tau}), \tag{2.20}$$

$$\sqrt{n}(\hat{\beta}_\tau - \beta_{0\tau}) = -[\dot{M}_n(\tau)]^{-1} \sqrt{n} \hat{M}_n(\beta_{0\tau}, \tau). \tag{2.21}$$

From A11, after plugging in definitions, $\dot{M}_n(\tau) \xrightarrow{p} G$; applying the continuous mapping theorem, $-[\dot{M}_n(\tau)]^{-1} \xrightarrow{p} -G^{-1}$. Using A10, the rest of the right-hand side of (2.21) has an asymptotic normal distribution. Equation (2.21) also implies the asymptotically linear (influence function) representation

$$\sqrt{n}(\hat{\beta}_\tau - \beta_{0\tau}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n [-\dot{M}_n(\tau)]^{-1} g_{ni}(\beta_{0\tau}, \tau) = -\frac{1}{\sqrt{n}} \sum_{i=1}^n G^{-1} g_{ni}(\beta_{0\tau}, \tau) + o_p(1). \tag{2.22}$$

The following lemma relies on A10 and a proof that the asymptotic "bias" is negligible, i.e., $\sqrt{n}\, \mathrm{E}[\hat{M}_n(\beta_{0\tau}, \tau)] \to 0$.

Lemma 2.4. *Under Assumptions A1–A6, A9, and A10,*

$$\sqrt{n} \hat{M}_n(\beta_{0\tau}, \tau) \xrightarrow{d} \mathrm{N}(0, \Sigma_{M\tau}), \quad \Sigma_{M\tau} = \lim_{n \to \infty} \mathrm{Var}\left(n^{-1/2} \sum_{i=1}^n g_{ni}(\beta_{0\tau}, \tau)\right). \tag{2.23}$$

With iid data and the conditional quantile restriction $\mathrm{P}\big(\Lambda(Y_i, X_i, \beta_{0\tau}) \le 0 \mid Z_i\big) = \tau$*, then* $\Sigma_{M\tau} = \tau(1-\tau)\, \mathrm{E}(Z_i Z_i^\top)$*.*

Given Lemma 2.4, we now establish the asymptotic normality of the SIVQR estimator.

Theorem 2.5. *Under Assumptions A1–A11,*

$$\sqrt{n}(\hat{\beta}_\tau - \beta_{0\tau}) \xrightarrow{d} \mathrm{N}\big(0, G^{-1} \Sigma_{M\tau} [G^\top]^{-1}\big),$$

where G *and* $\Sigma_{M\tau}$ *are defined in* (2.13) *and* (2.23)*, respectively.*

3 Inference

In this section, we turn our attention to inference in the smoothed quantile model. We discuss a Wald-type test and bootstrap procedures.

We consider testing the general r-dimensional null hypothesis

$$H_0 : \underbrace{a(\beta_{0\tau})}_{r\times 1} = 0. \tag{3.1}$$

One approach is to apply an existing method for full vector inference, and then use projection (e.g., Chernozhukov et al., 2009, end of §2.3). The primary advantage of this approach is that such methods are robust to weak (and even set) identification of $\beta_{0\tau}$; one even provides exact finite-sample inference under certain conditions. The primary disadvantage is that projection may be very conservative, especially if d_β is much larger than r. One such method is a χ^2 test of $H_0 : \beta_{0\tau} = b$: if $\sqrt{n}\hat{M}_n^u(\beta_{0\tau}, \tau) \xrightarrow{d} \mathrm{N}(0, \Sigma)$, then $n\hat{M}_n^u(\beta_{0\tau}, \tau)^\top \hat{\Sigma}^{-1} \hat{M}_n^u(\beta_{0\tau}, \tau) \xrightarrow{d} \chi^2_d$ given a consistent estimator $\hat{\Sigma} \xrightarrow{p} \Sigma$, where $d = d_\beta = d_Z$. Another option is the simulation-based method in Chernozhukov et al. (2009); it restricts sampling dependence more and requires a correctly specified *conditional* quantile model, in return for exact finite-sample (instead of asymptotic) properties.

There are also weak identification-robust methods (unlike ours) for testing null hypotheses that specify the value of the coefficient on each endogenous regressor, but they all assume iid sampling. These include Chernozhukov and Hansen (2008), Jun (2008), and Andrews and Mikusheva (2016), as compared by Chernozhukov et al. (2017, §1.3.3), who advocate the latter method.

Notationally, let the Jacobian of $a(\cdot)$ evaluated at the true $\beta_{0\tau}$ be

$$\underbrace{A}_{r\times d_\beta} \equiv \frac{\partial}{\partial \beta^\top} a(\beta)\big|_{\beta=\beta_{0\tau}}. \tag{3.2}$$

Assume that $a(\cdot)$ is continuously differentiable at $\beta_{0\tau}$ and that A has rank r, $r \leq d_\beta$. Let

$$\hat{A} \equiv \frac{\partial}{\partial \beta^\top} a(\beta)\big|_{\beta=\hat{\beta}_\tau} \xrightarrow{p} A, \tag{3.3}$$

where $\hat{\beta}_\tau$ is from (2.9).

As a special case, (3.1) includes linear hypotheses of the form $A\beta_{0\tau} = c$ when $a(\beta) = A\beta - c$. The full vector null hypothesis $H_0 : \beta_{0\tau} = b$ is another special case, when $a(\beta) = \beta - b$.

3.1 Wald test

In our setting, the advantage of the Wald test (over LM or LR-type tests) is that it does not require computation of a constrained estimator, only the unconstrained estimator. Recall that our unconstrained estimator in (2.9) solves the sample analog of a just-identified system of moment equations because this appears to be more computationally robust than minimizing a quadratic form of sample moments in an overidentified system. A constrained estimator, however, cannot set the sample moments to zero due to the constraints, so it must solve a minimization problem and may fail to find the global minimum. Comparing the Wald test to LM or LR-type tests, there may be a trade-off between this advantage of computational robustness and the usual disadvantages of the need to estimate the asymptotic covariance and the lack of invariance to reparameterization for certain null hypotheses.

Given Theorem 2.5, the statistic and its asymptotic distribution under H_0 are standard, similar to W_{1n} and Theorem 9.2 in Newey and McFadden (1994, p. 2222). From the delta method and Theorem 2.5,

$$\sqrt{n}[a(\hat{\beta}_\tau) - \overbrace{a(\beta_{0\tau})}^{=0 \text{ under (3.1)}}] \xrightarrow{d} \mathrm{N}\big(0, A\,G^{-1}\Sigma_{M\tau}[G^\top]^{-1}A^\top\big).$$

Therefore, the Wald statistic is asymptotically chi-square distributed as

$$\hat{W} \equiv na(\hat{\beta}_\tau)^\top \Big[\hat{A}\,\hat{G}^{-1}\hat{\Sigma}_{M\tau}(\hat{G}^\top)^{-1}\hat{A}^\top\Big]^{-1} a(\hat{\beta}_\tau) \xrightarrow{d} \chi^2_r, \tag{3.4}$$

as long as consistent estimators are available for $\hat{A} \xrightarrow{p} A$, $\hat{\Sigma}_{M\tau} \xrightarrow{p} \Sigma_{M\tau}$, and $\hat{G} \xrightarrow{p} G$. The Wald test rejects H_0 in (3.1) when $\hat{W}$ exceeds the $(1-\alpha)$-quantile of the χ^2_r distribution; given (3.4), its asymptotic size is α.

The result $\hat{A} \xrightarrow{p} A$ follows from consistency of $\hat{\beta}_\tau$ and continuity of the derivative of $a(\cdot)$.

The result $\hat{\Sigma}_{M\tau} \xrightarrow{p} \Sigma_{M\tau}$ holds under iid sampling for the estimator

$$\hat{\Sigma}_{M\tau} = n^{-1}\sum_{i=1}^{n} g_{ni}(\hat{\beta}_\tau, \tau) g_{ni}(\hat{\beta}_\tau, \tau)^\top. \tag{3.5}$$

With weakly dependent data, a long-run variance estimator like that in Newey and West (1987) and Andrews (1991b) can be used (as implemented in our code).

In some cases, the result $\hat{G} \xrightarrow{p} G$ follows from Kato (2012) with only slight modification. In our notation, a plug-in estimator based on (2.15) is

$$\hat{G} \equiv -\frac{1}{nh_n}\sum_{i=1}^{n} \tilde{I}'\big(-\Lambda(Y_i, X_i, \hat{\beta}_\tau)/h_n\big) Z_i D_i(\hat{\beta}_\tau)^\top, \quad D_i(\hat{\beta}_\tau) \equiv \frac{\partial}{\partial\beta}\Lambda(Y_i, X_i, \beta)\Big|_{\beta=\hat{\beta}_\tau}. \tag{3.6}$$

Kato (2012) studies the special case of $Z_i = D_i(\beta) = X_i$, i.e., quantile regression. If $D_i(\beta)$ does not depend on β, then the results of Kato (2012) go through with only minor modification, and with the finite moment assumptions for X_i applied to Z_i and D_i.

3.2 Bootstrap

The earliest paper using a smoothed quantile regression estimator was by Horowitz (1998) for the purpose of establishing higher-order accuracy of bootstrap χ^2 tests and two-sided t-tests, where the Studentized statistic (not just the estimator) is computed in each bootstrap replication. We conjecture that smoothing has a similar effect on bootstrap inference here, too. Like the Wald test, this approach requires only the unconstrained estimator, as well as an estimator of the asymptotic covariance matrix. Adapting the procedure in Section 3 of Horowitz (1998) to our setting, for $b = 1, \ldots, B$ bootstrap samples, the Wald statistic $\hat{W}^{(b)}$ in (3.4) is computed, with $a(\hat{\beta}_\tau^{(b)}) - a(\hat{\beta}_\tau)$ replacing $a(\hat{\beta}_\tau) - a(\beta_{0\tau}) = a(\hat{\beta}_\tau)$, as well as $\hat{A}^{(b)}$, $\hat{G}^{(b)}$, and $\hat{\Sigma}^{(b)}$ replacing $\hat{A}$, $\hat{G}$, and $\hat{\Sigma}$. That is, the procedure approximates the Wald statistic's distribution in the "bootstrap world," where the "population" distribution is the empirical distribution, using simulation since the exact analytic form is intractable. The null is rejected if the original sample's $\hat{W}$ exceeds the $(1-\alpha)$-quantile of the bootstrap distribution, i.e., of the B values of $\hat{W}^{(b)}$.

As usual, the bootstrap samples may be generated differently depending on the structure of the data. In the iid case, as in Horowitz (1998), one may use a pairs bootstrap: for each $b = 1, \ldots, B$, resampling $(Y_i^{(b)}, X_i^{(b)}, Z_i^{(b)})$ (for $i = 1, \ldots, n$) from the n original values (Y_i, X_i, Z_i) with probability $1/n$ each, with replacement. With stationary time series data, we suggest the stationary block bootstrap of Politis and Romano (1994). With clustered data, where there is independence between clusters but arbitrary dependence within clusters, one may resample clusters instead of individuals as in Cameron, Gelbach, and Miller (2008); subsequent papers offer more robust methods in some cases. All these are implemented in our code.

Alternatively, one may simply bootstrap the estimator directly. The advantage is that estimators of Σ and G are not needed. In small samples, especially with dependent data, such estimators may be poor enough to hurt accuracy, even though the Studentized bootstrap probably (we conjecture) has higher-order asymptotic accuracy.

4 Monte Carlo simulations

This section reports Monte Carlo simulation results to illustrate the finite-sample performance of the proposed methods. Code to replicate results is available on the third author's website.[10] The following DGPs are used.

DGP 1 With iid sampling, there is a randomized treatment offer (instrument) $Z_i = 1$ with probability $1/2$ and $Z_i = 0$ otherwise; $U_i \sim \text{Unif}(0,1)$ (the unobservable); $D_i = 1$ if i is treated and $D_i = 0$ otherwise, with $\text{P}(D_i = 1 \mid Z_i, U_i) = Z_i \min\{1, (4/3)U_i\}$ so that there is imperfect compliance and endogenous self-selection into treatment; and $Y_i = \beta(U_i) + D_i\gamma(U_i)$, where the function $\beta(\tau) = 60 + Q(\tau)$ with $Q(\cdot)$ the quantile function of the χ^2_3 distribution, and $\gamma(\tau) = 100(\tau - 0.5)$, so there is heterogeneity of the quantile treatment effects.

DGP 2 This DGP is a stationary time series regression with measurement error. The latent explanatory variable is Z_t, with $Z_0 \sim \text{N}(0,1)$, $Z_t = \rho_Z Z_{t-1} + \sqrt{1-\rho_Z^2}\nu_t$, and $\nu_t \overset{iid}{\sim} \text{N}(0,1)$, so $\text{Var}(Z_t) = 1$ for all t; we use $\rho_Z = 0.5$. The measurement error is $\eta_t \overset{iid}{\sim} \text{N}(0,1)$, and $X_t = Z_t + \eta_t$ is the observed (mismeasured) explanatory variable. Since the η_t are independent, the lagged X_{t-1} provides a valid instrument. The outcome is $Y_t = \gamma Z_t + \epsilon_t$ with $\gamma = 1$ and ϵ_t unobserved. Finally, $\epsilon_t = \rho_\epsilon \epsilon_{t-1} + \sqrt{1-\rho_\epsilon^2}V_t$, $\epsilon_0 \sim \text{N}(0,1)$, $V_t \overset{iid}{\sim} \text{N}(0,1)$, so the marginal distribution is $\epsilon_t \sim \text{N}(0,1)$ for all t, and the series $\{\epsilon_t\}$, $\{\eta_t\}$, and $\{\nu_t\}$ are mutually independent. Letting $\beta(\tau)$ be the τ-quantile of $\epsilon_t - \gamma\eta_t$, since $Y_t = \gamma Z_t + \epsilon_t = \gamma X_t + \epsilon_t - \gamma\eta_t$, we have the quantile restrictions $\text{P}(Y_t - \gamma X_t - \beta(\tau) \le 0) = \text{P}(\epsilon_t - \gamma\eta_t \le \beta(\tau)) = \tau$, and $\text{P}(\epsilon_t - \gamma\eta_t \le \beta(\tau) \mid X_{t-1}) = \text{P}(\epsilon_t - \gamma\eta_t \le \beta(\tau)) = \tau$ since $X_{t-1} \perp\!\!\!\perp \eta_t, \epsilon_t$, so the IVQR intercept and slope parameters are $\beta(\tau)$ and $\gamma = 1$, respectively.

DGP 3 This DGP is identical to DGP 2 except that $\epsilon_t = \rho_\epsilon \epsilon_{t-1} + (1-\rho_\epsilon)V_t$, $\epsilon_0 \sim \text{Cauchy}$, $V_t \overset{iid}{\sim} \text{Cauchy}$, so the marginal distribution is $\epsilon_t \sim \text{Cauchy}$ for all t.

To measure estimator precision, instead of the usual root mean squared error (RMSE), we report "robust RMSE" with median bias replacing bias and interquartile range (divided by 1.349) replacing standard deviation. The primary reason is that (in many cases) the usual IV estimator does not even possess a first moment in finite samples, let alone finite variance (e.g., Kinal, 1980).[11] If the sampling distribution is normal, then robust RMSE equals RMSE.

Table 1 shows the precision of our smoothed estimator from (2.9) ("SIVQR") with the

[10] It is written in R (R Core Team, 2013) and uses the solver `pracma::newtonsys` (Borchers, 2015).

[11] If one really cares about finite-sample RMSE per se, then OLS should be preferred to IV in the many cases where the IV RMSE is infinite but the OLS RMSE is finite.

Table 1: Simulated precision of estimators of γ_τ.

			Robust RMSE			Median Bias		
DGP	τ	n	SIVQR	QR	IV	SIVQR	QR	IV
1	0.25	20	26.03	31.85	41.53	17.71	27.49	39.97
		50	17.74	27.69	40.42	10.38	25.69	39.88
		200	11.59	25.52	40.00	4.17	24.98	39.86
		500	9.11	25.26	39.98	1.09	25.03	39.93
	0.50	20	20.61	22.68	18.75	8.85	17.97	14.97
		50	13.42	20.45	16.29	3.58	18.36	14.88
		200	8.14	20.55	15.23	0.57	20.01	14.86
		500	5.32	20.10	15.08	0.34	19.91	14.93
2	0.25	20	1.16	0.61	1.17	−0.40	−0.54	−0.34
		50	0.86	0.55	0.73	−0.09	−0.52	−0.04
		200	0.41	0.51	0.30	0.01	−0.50	0.02
		500	0.25	0.51	0.20	0.01	−0.50	0.00
	0.50	20	1.17	0.60	1.17	−0.41	−0.54	−0.34
		50	0.80	0.55	0.73	−0.11	−0.52	−0.04
		200	0.37	0.51	0.30	0.00	−0.50	0.02
		500	0.23	0.50	0.20	0.00	−0.50	0.00
3	0.25	20	2.33	0.75	3.74	−0.35	−0.53	−0.30
		50	1.70	0.62	3.03	−0.04	−0.52	−0.13
		200	0.75	0.53	2.29	0.01	−0.51	−0.03
		500	0.46	0.52	2.28	−0.01	−0.51	−0.01
	0.50	20	2.13	0.67	3.74	−0.37	−0.52	−0.30
		50	1.36	0.56	3.03	−0.11	−0.51	−0.13
		200	0.62	0.52	2.29	0.02	−0.50	−0.03
		500	0.34	0.51	2.28	0.00	−0.51	−0.01

1000 replications. "SIVQR" is the estimator in (2.9); "QR" is quantile regression (no IV); "IV" is the usual (mean) IV estimator.

smallest possible smoothing bandwidth (for simplicity), as well as the usual quantile regression ("QR") estimator (ignoring endogeneity) and the usual (mean) IV estimator. The robust RMSE values are plotted in Figures 2 and 3.

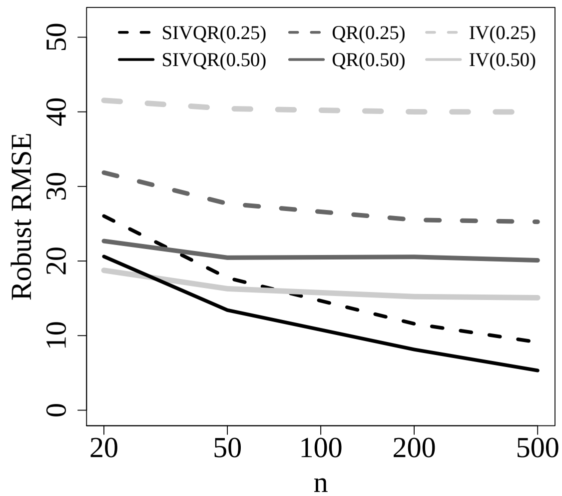

Figure 2: Simulated robust RMSE, DGP 1, 1000 replications.

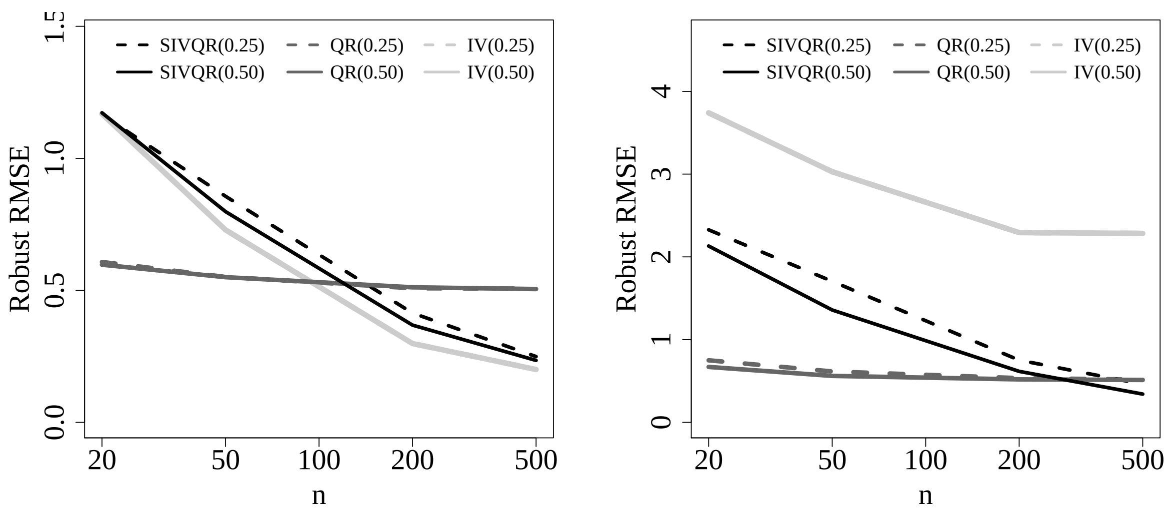

Figure 3: Simulated robust RMSE, DGP 2 (left) and DGP 3 (right), 1000 replications.

For all DGPs, the smoothed IVQR estimator's robust RMSE declines toward zero as n increases. In contrast, the QR estimator's robust RMSE never goes to zero due to endogeneity, and the IV estimator's robust RMSE fails to go to zero in DGP 1 where there is heterogeneity across quantiles. This reflects the theoretical result that only our estimator is consistent for γ_τ when there is endogeneity and heterogeneity.

Figures 2 and 3 also show important finite-sample differences not captured by first-order asymptotics. With $n = 20$, for DGP 2 or 3, the lowest robust RMSE is actually that of QR: despite its (median) bias being the largest due to ignoring the endogeneity, its dispersion is so much smaller than the other estimators' dispersions that its overall robust RMSE is the smallest. This advantage persists to $n = 50$, but eventually n is large enough for the (median) bias to dominate. Although the usual IV estimator is consistent in DGPs 2 and 3 due to the constant slope at all τ, it is only modestly more precise than SIVQR in the Gaussian innovation variant, and it has much larger robust RMSE with Cauchy innovations. This illustrates the efficiency motivation for using a median instead of mean estimator when heavy-tailed distributions are suspected, even with endogeneity.

For testing $H_0 : \gamma_\tau = \gamma_0$ against $H_1 : \gamma_\tau \neq \gamma_0$, we consider three methods: the Wald test, basic bootstrap, and Studentized bootstrap. A comprehensive simulation study of the many methods in the literature would be valuable but is beyond the scope of this paper; we only attempt to check whether or not these three new methods appear reasonable, especially in the time series case where the literature is sparse. Each method uses the rule-of-thumb bandwidth from Kato (2012, p. 264), appropriately modified. The Wald test uses the statistic in (3.4) with the χ^2_1 critical value. For the iid DGP, (3.5) is used for $\hat{\Sigma}_{M\tau}$. For the time series DGPs, the Newey and West (1987) estimator is used (i.e., a Bartlett kernel); specifically, we use equation (2.5) of Andrews (1991b) with the Bartlett kernel in (2.7) and the corresponding data-dependent bandwidth in (6.2), using (6.4). In either case, we use (3.6) for $\hat{G}$. The Studentized bootstrap computes the Wald statistic this same way in each bootstrap draw. The basic bootstrap only computes the estimator in each draw. For the time series DGPs, our stationary bootstrap (whether Studentized or not) follows the resampling algorithm just before Proposition 1 in Politis and Romano (1994, p. 1304) and sets the tuning parameter value to $p = n^{-1/3}$, the rate suggested in Politis and Romano (1994, p. 1306). Total simulation replications, bootstrap draws, and other details are found in Table 2.

Table 2 shows the simulated size of the tests for various DGPs, n, and τ. The Wald test's size appears to approach the nominal α asymptotically, but it has considerable size distortion with $n = 100$ and even $n = 1000$ in some cases. This is probably due to the slow convergence rate of the nonparametric estimator of the asymptotic covariance matrix, which involves a conditional PDF. The asymptotic χ^2_1 distribution of the Wald statistic essentially ignores this covariance estimation error, replacing the estimator with its plim and treating the $o_p(1)$

Table 2: Simulated size of two-sided tests of $H_0 : \gamma_\tau = \gamma_0$.

DGP	τ	n	α	Wald	BS-t	BS
1	0.25	100	0.10	0.411	0.094	0.196
		1000	0.10	0.415	0.067	0.115
		10,000	0.10	0.226	n/a	n/a
	0.50	100	0.10	0.550	0.074	0.120
		1000	0.10	0.342	0.041	0.101
		10,000	0.10	0.134	n/a	n/a
2	0.25	100	0.10	0.269	0.098	0.040
		1000	0.10	0.154	0.107	0.088
		10,000	0.10	0.116	n/a	n/a
	0.50	100	0.10	0.246	0.103	0.040
		1000	0.10	0.147	0.119	0.090
		10,000	0.10	0.101	n/a	n/a
3	0.25	100	0.10	0.389	0.059	0.022
		1000	0.10	0.179	0.075	0.059
		10,000	0.10	0.116	n/a	n/a
	0.50	100	0.10	0.296	0.067	0.019
		1000	0.10	0.159	0.110	0.079
		10,000	0.10	0.092	n/a	n/a

Nominal α shown in table, 1000 replications. "Wald" uses (3.4) directly, while "BS-t" bootstraps the Wald statistic and "BS" bootstraps the estimator, both as described in Section 3.2 and using 299 bootstrap draws each. Bootstrap results for $n = 10{,}000$ not available ("n/a") due to computation time.

remainder as asymptotically negligible. In contrast to the Wald test's size distortion, both bootstraps have tolerable size in almost all cases, even with $n = 100$: among the 24 values, the highest size was 0.196 (compared to $\alpha = 0.1$), and next-worst was only 0.120. Using more bootstrap draws may improve the bootstraps further. Studentization appears to be helpful: the Studentized bootstrap's size is closer than the basic bootstrap's size to the nominal α in 8 of 12 cases shown in the table.

5 Application: quantile Euler equation

This section illustrates the usefulness of the proposed methods through an empirical example: the estimation of a quantile Euler equation. We apply the proposed methodology to an economic model of intertemporal allocation of consumption and estimate the elasticity of intertemporal substitution (EIS). The EIS is a parameter of central importance in macroeconomics and finance. We refer to Campbell (2003), Cochrane (2005), and Ljungqvist and Sargent (2012), and the references therein, for a comprehensive overview.

There is a large empirical literature that attempts to estimate the EIS; among others, Hansen and Singleton (1983), Hall (1988), Campbell and Mankiw (1989), Campbell and Viceira (1999), Campbell (2003), and Yogo (2004). The majority of the literature relies on the traditional expected utility framework to estimate the EIS. The purpose of this application is to estimate and make inference on the EIS for selected developed countries in Campbell's (2003) data set using the quantile utility maximization model.

Section 5.1 describes in detail the model that leads to the quantile Euler equation, establishing parallels with the standard expected utility Euler equation. Section 5.2 describes the estimation procedure, and Section 5.3 discusses log-linearization for the quantile model. Section 5.4 discusses an interpretation of the parameters in question. In Section 5.5 we review the data, and finally Section 5.6 presents the empirical results.

5.1 Description of the economic model

De Castro and Galvao (2017) employ a variation of the standard economy model of Lucas (1978). The economic agents decide on the intertemporal consumption and savings (assets to hold) over an infinity horizon economy, subject to a linear budget constraint. The decision generates an intertemporal policy function, which is used to estimate the parameters of interest for a given utility function. Their work is related to that of Giovannetti (2013), who works with a similar model but restricts the analysis to two periods, whereas de Castro and Galvao (2017) consider an infinite horizon.

The specific model is as follows. Let C_t denote the amount of consumption good that the individual consumes in period t. At the beginning of period t, the consumer has x_t units of the risky asset, which pays dividend d_t. The price of the consumption good is normalized to one, while the price of the risky asset in period t is $p(d_t)$. Then, the consumer decides its consumption C_t and how many units of the risky asset x_{t+1} to save for the next period, subject to the budget constraint

$$C_t + p(d_t)x_{t+1} \leq [d_t + p(d_t)]x_t \tag{5.1}$$

and positivity restriction

$$C_t, x_{t+1} \geq 0. \tag{5.2}$$

In equilibrium, we have that $x_t^* = 1, \forall t, k$.

So far, the model is exactly the same as the standard Lucas' model, but the objective function will differ. In the standard model, the consumer maximizes

$$\sum_{t=0}^{\infty} \mathrm{E}\big[\beta^t U(C_t) \mid \Omega_0\big] \tag{5.3}$$

subject to (5.1) and (5.2), where $\beta \in (0,1)$ is the discount factor, $U : \mathbb{R}_+ \to \mathbb{R}$ is the utility function, and Ω_0 is the information set at time $t = 0$. In de Castro and Galvao (2017), the consumer instead maximizes the objective function

$$\sum_{t=0}^{\infty} \mathrm{Q}_\tau^t\big[\beta^t U(C_t) \mid \Omega_0\big], \tag{5.4}$$

again subject to (5.1) and (5.2), where (5.4) is (an abuse of) notation defined below in (5.8).

For the expected utility choice problem, dynamic consistency and the principle of optimality for (5.3) imply that at time $s \geq 1$, the consumer chooses $\{C_t, x_t\}_{t\geq s}$ to maximize

$$\sum_{t=s}^{\infty} \mathrm{E}\big[\beta^{t-s} U(C_t) \mid \Omega_s\big] \tag{5.5}$$

subject to (5.1) and (5.2). The connection between problems (5.3) and (5.5) is made explicit by the linearity of the expectation operator and the law of iterated expectations:

$$\begin{aligned} &\sum_{t=0}^{\infty} \mathrm{E}\big[\beta^t U(C_t) \mid \Omega_0\big] \\ &= U(C_0) + \mathrm{E}\Big[\beta U(C_1) + \mathrm{E}\Big[\beta^2 U(C_2) + \mathrm{E}\big[\beta^3 U(C_3) + \cdots \mid \Omega_2\big] \;\Big|\; \Omega_1\Big] \;\Big|\; \Omega_0\Big]. \end{aligned} \tag{5.6}$$

The structure in the right-hand side of (5.6) reflects the associated value function,

$$v(x_t, d_t) = \max_{x_{t+1}\geq 0} \big\{ U\big([d_t + p(d_t)]x_t - p(d_t)x_{t+1}\big) + \beta\, \mathrm{E}[v(x_{t+1}, d_{t+1}) \mid \Omega_t] \big\}. \tag{5.7}$$

Unfortunately, linearity and the law of iterated expectations do not hold for the τ-quantile operator, Q_τ. Thus, in order to preserve dynamic consistency and the principle of optimality, we need to maintain the structure developed in (5.6). The (abuse of) notation introduced in (5.4) is defined as follows:

$$\sum_{t=0}^{\infty} \mathrm{Q}_\tau^t \left[\beta^t U(C_t) \mid \Omega_0\right]$$

$$\equiv U(C_0) + \mathrm{Q}_\tau\left[\beta U(C_1) + \mathrm{Q}_\tau\left[\beta^2 U(C_2) + \mathrm{Q}_\tau\left[\beta^3 U(C_3) + \cdots \mid \Omega_2\right] \;\middle|\; \Omega_1\right] \;\middle|\; \Omega_0\right]. \tag{5.8}$$

De Castro and Galvao (2017) show that the limit above exists and is well defined. Moreover, they show that the quantile preferences are dynamically consistent, the principle of optimality holds, and the corresponding dynamic problem yields a value function, via a fixed-point argument. They further provide conditions so that the value function is differentiable and concave. The value function for the quantile problem is the same as (5.7) but with Q_τ replacing E:

$$v(x_t, d_t) = \max_{x_{t+1}\geq 0} \big\{ U\big([d_t + p(d_t)]x_t - p(d_t)x_{t+1}\big) + \beta\, \mathrm{Q}_\tau[v(x_{t+1}, d_{t+1}) \mid \Omega_t] \big\}. \tag{5.9}$$

In addition, de Castro and Galvao (2017) derive the corresponding Euler equation, using the fact that in equilibrium, the holdings are $x_t = 1$ for all t:

$$-\,p(d_t)U'(C_t) + \beta\, \mathrm{Q}_\tau[U'(C_{t+1})(z_{t+1} + p(z_{t+1})) \mid \Omega_t] = 0. \tag{5.10}$$

Defining the asset's return by

$$1 + r_{t+1} \equiv \frac{z_{t+1} + p(z_{t+1})}{p(d_t)}, \tag{5.11}$$

the Euler equation in (5.10) simplifies to

$$\mathrm{Q}_\tau\left[\beta(1 + r_{t+1})\frac{U'(C_{t+1})}{U'(C_t)} \;\middle|\; \Omega_t\right] = 1. \tag{5.12}$$

Equation (5.12) is in the form of (2.3), which is within our econometric framework.

The quantile Euler equation in (5.12) looks similar to the standard Euler equation from expected utility maximization,

$$\mathrm{E}\left[\beta(1+r_{t+1})\frac{U'(C_{t+1})}{U'(C_t)} \;\middle|\; \Omega_t\right] = 1. \tag{5.13}$$

The expressions inside the conditional quantile and conditional expectation are identical.

For obtaining the mentioned results, de Castro and Galvao (2017) assume the following.

Assumption A12. *(i) The dividends assume values in $\mathcal{Z} \subseteq \mathbb{R}$, which is a bounded interval, and $\mathcal{X} = [0, \bar{x}]$ for some $\bar{x} > 1$;*

(ii) $\{d_t\}$ is a Markov process with PDF $f : \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}_+$, which is continuous, symmetric ($f(a,b) = f(b,a)$), $f(d_t, d_{t+1}) > 0$ for all $(d_t, d_{t+1}) \in \mathcal{Z} \times \mathcal{Z}$, and satisfies the property: if $h : \mathcal{Z} \to \mathbb{R}$ is weakly increasing and $z \leq z'$, then

$$\int_{\mathcal{Z}} h(\alpha) f(\alpha \mid z)\, d\alpha \leq \int_{\mathcal{Z}} h(\alpha) f(\alpha \mid z')\, d\alpha; \tag{5.14}$$

(iii) $U : \mathbb{R}_+ \to \mathbb{R}$ is given by $U(c) = \frac{1}{1-\gamma} c^{1-\gamma}$, for $\gamma > 0$;

(iv) $z \mapsto z + p(z)$ is C^1 and non-decreasing, with $\frac{d}{dz} z[\ln(z + p(z))] \geq \gamma$.

Assumptions A12(i)–(iii) are standard in economic applications. In Assumption A12(iv), it is natural to expect that the price $p(z)$ is non-decreasing with the dividend z, and $z + p(z)$ being non-decreasing is an even weaker requirement.

5.2 Estimation

We follow a large body of the literature (e.g., Campbell, 2003) and use the isoelastic utility function

$$U(C_t) = \frac{1}{1-\gamma} C_t^{1-\gamma}, \quad \gamma > 0. \tag{5.15}$$

The ratio of marginal utilities is

$$\frac{U'(C_{t+1})}{U'(C_t)} = \left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}. \tag{5.16}$$

From (5.12) and (5.16), the Euler equation is thus

$$\mathrm{Q}_\tau\left[\beta_\tau(1+r_{t+1})(C_{t+1}/C_t)^{-\gamma_\tau} - 1 \mid \Omega_t\right] = 0. \tag{5.17}$$

The quantile Euler equation in (5.17) is a conditional quantile restriction with finite-dimensional parameter vector $(\beta_\tau, \gamma_\tau)$. Thus, we may estimate the parameters with SIVQR.

5.3 Log-linearization

One benefit of the quantile Euler equation is that it may be log-linearized with no approximation error, unlike the standard Euler equation. One may rewrite the Euler equation in (5.17) as

$$\mathrm{Q}_\tau[\epsilon_{t+1} \mid \Omega_t] = 1, \quad \epsilon_{t+1} \equiv \beta(1+r_{t+1})(C_{t+1}/C_t)^{-\gamma}. \tag{5.18}$$

For general random variable W, $\mathrm{Q}_\tau[\ln(W)] = \ln(\mathrm{Q}_\tau[W])$ since $\ln(\cdot)$ is strictly increasing. In contrast, $\mathrm{E}[\ln(W)] \le \ln(\mathrm{E}[W])$ by Jensen's inequality. Continuing from (5.18), taking logs and rearranging,

$$\begin{aligned} \ln(\epsilon_{t+1}) &= \ln(\beta) + \ln(1+r_{t+1}) - \gamma \ln(C_{t+1}/C_t), \\ \ln(C_{t+1}/C_t) &= \gamma^{-1}\ln(\beta) + \gamma^{-1}\ln(1+r_{t+1}) - \gamma^{-1}\ln(\epsilon_{t+1}). \end{aligned} \tag{5.19}$$

Assuming $\gamma > 0$, then $-\gamma^{-1}\ln(\epsilon)$ is strictly *decreasing* in ϵ. This switches the quantile index since $\mathrm{Q}_\tau(W) = 0$ is equivalent to $\mathrm{Q}_{1-\tau}(-W) = 0$, so

$$\begin{aligned} 0 = \ln(1) &= \mathrm{Q}_\tau[\ln(\epsilon_{t+1}) \mid \Omega_t] = \mathrm{Q}_{1-\tau}[-\gamma^{-1}\ln(\epsilon_{t+1}) \mid \Omega_t] \\ &= \mathrm{Q}_{1-\tau}[\ln(C_{t+1}/C_t) - \gamma^{-1}\ln(\beta) - \gamma^{-1}\ln(1+r_{t+1}) \mid \Omega_t]. \end{aligned}$$

Thus, $\ln(\beta)/\gamma$ and $1/\gamma$ are the intercept and slope (respectively) of the $1-\tau$ IV quantile regression of $\ln(C_{t+1}/C_t)$ on a constant and $\ln(1+r_{t+1})$, where instruments are variables in Ω_t.

Because log-linearization of a quantile Euler equation has no approximation error, the nonlinear and log-linear SIVQR methods "should" give the same results. In our application, this is generally true, but there are some exceptions. Nonlinear and log-linear point estimates usually match up to 3+ significant figures. The one exception (in our application) is when the log-linear estimate has $\hat{\gamma}_\tau < 0$ (so $\mathrm{Q}_{1-\tau}$ above should instead remain as Q_τ); the nonlinear estimator often has difficulty finding the solution in such cases and requires more smoothing to find a solution, which sometimes has a positive EIS. For hypothesis testing, the results are also similar, but the log-linear tests tend to produce somewhat smaller p-values.

5.4 Interpretation

The parameters of interest in (5.17) are β_τ and γ_τ. The former is the usual discount factor. The parameter $1/\gamma_\tau$ is the standard measure of EIS implicit in the CRRA utility function in

(5.15). The EIS is a measure of responsiveness of the consumption growth rate to the real interest rate. As in Hall (1988), in a model with uncertainty, the interpretation is similar, and a high value of EIS means that when the real interest rate is expected to be high, the consumer will actively defer consumption to the later period.

The interpretation of $1/\gamma_\tau$ as the EIS remains valid for the quantile maximization model.[12] Most directly, this can be seen in equation (5.19), where $1/\gamma$ is the derivative of $\ln(C_{t+1}/C_t)$ with respect to $\ln(1 + r_{t+1})$, holding ϵ_{t+1} constant.

5.5 Data

We use data originally from Campbell (2003) and provided by Yogo (2004).[13] It consists of aggregate level quarterly data for the United States (US), United Kingdom (UK), Netherlands, and Sweden. The sample period for the US is 1947Q3–1998Q4, UK is 1970Q3–1999Q1, Netherlands is 1977Q3–1998Q4, and Sweden is 1970Q3–1999Q2. Consumption is measured at the beginning of the period, consisting of nondurables plus services for the US and total consumption for the other countries, in real, per capita terms. The real interest rate deflates a proxy for the nominal short-term rate by the consumer price index. Instruments include lags of log real consumption growth, nominal interest rate, inflation, and a log dividend-price ratio for equities. For a complete description of the data, see Campbell (2003).

5.6 Results

Other than using quantiles, our estimation generally follows Table 2 of Yogo (2004, p. 805). Yogo (2004) uses 2SLS to estimate the (structural) log-linearized model $\ln(C_{t+1}/C_t) = \delta_0 + \delta_1 \ln(1 + r_{t+1}) + u_{t+1}$, where r_{t+1} is the real interest rate, instrumenting for $\ln(1 + r_{t+1})$ with twice lagged measures of nominal interest rate, inflation, consumption growth, and log dividend-price ratio (per his Table 2 footnote), where $\delta_1 = 1/\gamma$ is the EIS and $\delta_0 = \ln(\beta)/\gamma$. Yogo (2004) emphasizes that these are strong instruments that predict the real interest rate well.[14] Since the log-linear and nonlinear quantile models are equivalent when $\gamma_\tau > 0$, we report estimates from the original (nonlinear) quantile Euler equation in (5.17); the log-linear

[12] Hall's (1988) argument that γ fundamentally represents the EIS rather than risk aversion applies here, too.

[13] https://sites.google.com/site/motohiroyogo/research/EIS_Data.zip

[14] In contrast, when trying to estimate the EIS by 2SLS of $\ln(1 + r_{t+1})$ on $\ln(C_{t+1}/C_t)$, or replacing the real interest rate with a real stock index return, the instruments are weak because it is difficult to predict consumption growth or stock returns. We focus only on the specification with strong instruments.

estimates usually match at least two significant digits.

Table 3 shows the quantile Euler equation estimates for β_τ and γ_τ using SIVQR for the deciles $\tau = 0.1, \ldots, 0.9$. For comparison, Table 3 also reports the 2SLS estimates, which match those in Yogo (2004) (except that he reports $1/\hat{\gamma}$ instead of $\hat{\gamma}$ and does not report $\hat{\beta}$). Figures 4 and 5 show plots of $\hat{\beta}_\tau$ and $1/\hat{\gamma}_\tau$.

Table 3: SIVQR estimates of β_τ and γ_τ.

	US		UK		Netherlands		Sweden	
τ	$\hat{\beta}_\tau$	$\hat{\gamma}_\tau$	$\hat{\beta}_\tau$	$\hat{\gamma}_\tau$	$\hat{\beta}_\tau$	$\hat{\gamma}_\tau$	$\hat{\beta}_\tau$	$\hat{\gamma}_\tau$
0.10	3.18*	22.0*	1.14	8.2*	1.02	−3.0	n/a	n/a
0.20	1.64	20.5*	1.11	7.5*	1.00	−2.9*	n/a	n/a
0.30	1.30	20.4*	1.13	10.8*	0.99	−10.8*	n/a	n/a
0.40	1.15	17.0*	1.07	8.4*	1.03	−8.6*	n/a	n/a
0.50	0.79	−43.9*	1.09	14.6*	1.07	15.2*	0.93	−32.4*
0.60	1.10	23.2*	1.01	4.4*	1.12	38.0*	1.01	28.1*
0.70	1.01	4.9*	1.00	2.8	0.98	25.5*	0.98	8.9*
0.80	1.01	5.2*	0.98	3.9*	0.90	19.1*	0.98	4.7*
0.90	0.98	10.8*	0.95	6.1*	0.87	−6.0*	n/a	n/a
2SLS	1.08	16.7*	1.03	6.0*	0.96	−6.8*	0.27	−544.4*

Using quarterly data from Yogo (2004). Two instruments: constant, and the linear projection of the $t+1$ real interest rate onto a constant and twice lagged nominal interest rate, inflation, log consumption growth, and log dividend-price ratio, the same excluded instruments used in Table 2 of Yogo (2004). The smoothing bandwidth was (usually) $h = 0.0001$. Asterisks indicate statistically significant difference from 1 at a 10% level (two-sided basic stationary bootstrap).

Table 3 and Figure 4 show SIVQR estimates of β_τ and γ_τ to be economically plausible over a range of τ for both the US and UK. For the US, 2SLS has $\hat{\beta} = 1.08$ and $\hat{\gamma} = 16.7$; over $\tau \in \{0.6, 0.7, 0.8, 0.9\}$, the ranges of quantile estimates are $\hat{\beta}_\tau \in [0.98, 1.10]$ and $\hat{\gamma}_\tau \in [4.9, 23.2]$. For the UK, 2SLS has $\hat{\beta} = 1.03$ and $\hat{\gamma} = 6.0$; over $\tau \in \{0.6, 0.7, 0.8, 0.9\}$, $\hat{\beta}_\tau \in [0.95, 1.01]$ and $\hat{\gamma}_\tau \in [2.8, 6.1]$. For both the US and UK with these τ (and almost all other τ, too), $H_0 : \beta_\tau = 1$ is never rejected, and $H_0 : \gamma_\tau = 1$ is rejected all but one time (10% level, two-sided basic stationary bootstrap). In sum, the quantile estimates have similarities with the 2SLS estimates in some cases but economically significant differences in others, partly depending on τ.

Table 3 and Figure 5 show examples where the 2SLS estimates are not economically plausible but some quantile estimates are. For the Netherlands, 2SLS has $\hat{\gamma} = -6.8$, contradicting the common presumption $\gamma \geq 1$ and failing the sanity check of $\gamma > 0$; in contrast,

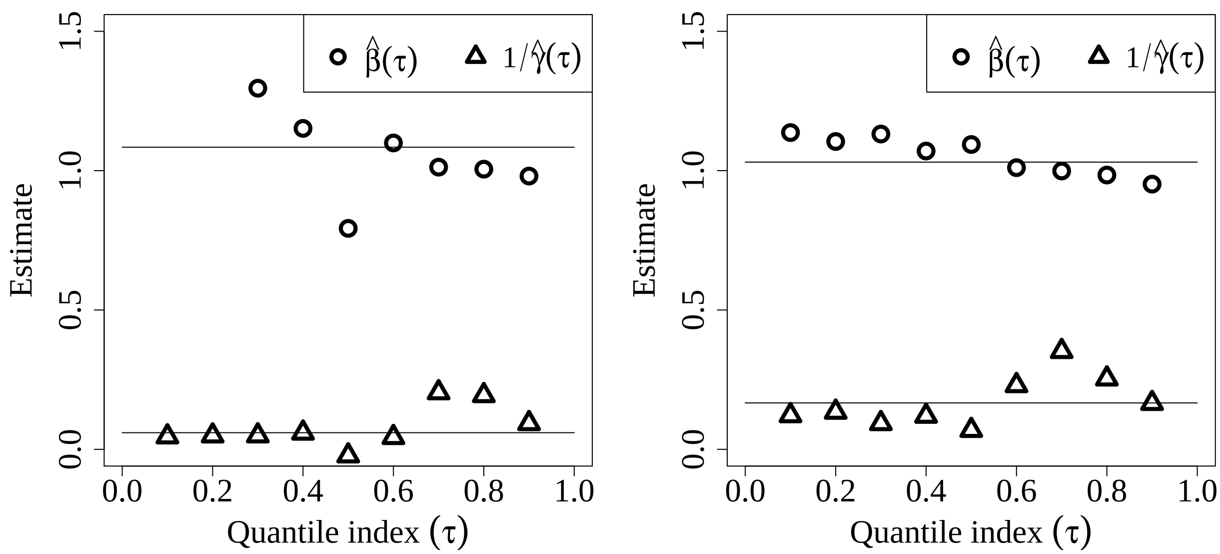

Figure 4: Estimates from Table 3 for US (left) and UK (right). Horizontal lines are 2SLS estimates.

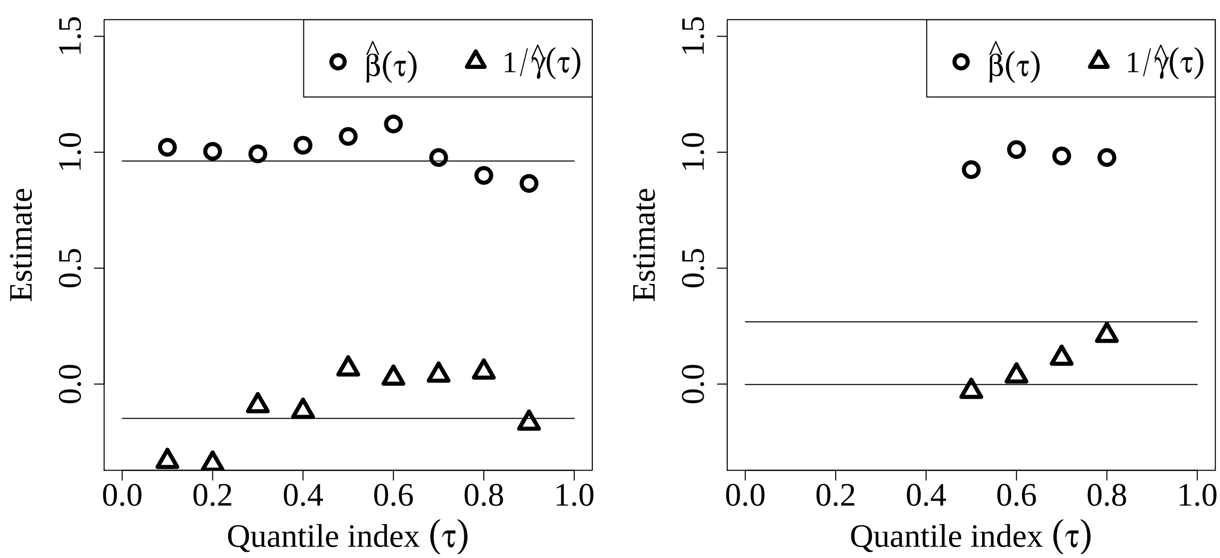

Figure 5: Estimates from Table 3 for the Netherlands (left) and Sweden (right). Horizontal lines are 2SLS estimates.

over $\tau \in \{0.5, 0.6, 0.7, 0.8\}$, $\hat{\gamma}_\tau \in [15.2, 38.0]$, and the corresponding $\hat{\beta}_\tau \in [0.90, 1.12]$ with none statistically different from one. For Sweden, 2SLS has $\hat{\beta} = 0.27$ (implausibly low) and $\hat{\gamma} = -544.4$ (implausibly negative); in contrast, over $\tau \in \{0.6, 0.7, 0.8\}$, $\hat{\beta}_\tau \in [0.98, 1.01]$ and $\hat{\gamma}_\tau \in [4.7, 28.1]$. Of course, this does not *always* happen; we omit less interesting results for countries where neither 2SLS nor SIVQR estimate plausible values.

Across all four countries, the values $\tau \in \{0.7, 0.8\}$ consistently yield plausible estimates, though perhaps this is mere coincidence. Among the eight values (four countries, two τ), all $\hat{\gamma}_\tau \in [2.8, 25.5]$, and all but one $\hat{\beta}_\tau \in [0.98, 1.01]$.

In all, the application illustrates that the quantile utility maximization model and the proposed SIVQR methods serve as important tools to study economic behavior.

6 Conclusion

For finite-dimensional parameters defined by general quantile-type restrictions, we have developed asymptotic theory for a feasible estimator using smoothed sample moments, allowing for weakly dependent data and nonlinear models. This includes nonlinear instrumental variables quantile regression as a special case, as well as the quantile Euler equations in our empirical example.

The empirical results suggest that the combination of quantile utility maximization and our smoothed estimation approach can provide a useful, economically meaningful alternative to estimation based on expected utility. A bonus feature is the ability to log-linearize the quantile Euler equation without any approximation error, unlike the standard Euler equation. Future work may apply our methods to household panel data or carefully consider how to determine τ.

There is much more to explore econometrically, too: smoothed quantile GMM (work in progress), efficiency improvements from optimal instruments or averaging estimators like in Hansen (2017) or Cheng, Liao, and Shi (2016), optimal smoothing bandwidth choice, semiparametric or nonparametric moment conditions (i.e., involving infinite-dimensional parameters), formally establishing A11 when D_i depends on the unknown parameter, extending the “one-step” Theorem 3.5 in Newey and McFadden (1994) to account for non-smoothness, and derivation of results uniform in τ, among other topics. In addition, an overidentification test can be constructed by modifying Section 9.5 in Newey and McFadden (1994), viewing the SIVQR estimator as a GMM estimator under exact identification. Although the most familiar GMM overidentification test is the J-test based on the (overidentified) GMM criterion with the (estimated) efficient weight matrix, Newey and McFadden (1994) describe a more general test that only requires a $\sqrt{n}$-consistent estimator of the parameter vector. Under the assumptions of Theorem 2.5, our estimator $\hat{\beta}_\tau$ is such an estimator.

A Proofs

Proof of Proposition 2.1. See Appendix A.1 of Chen et al. (2014). □

Proof of Lemma 2.2. Noting that $\left|Z[\tilde{I}(\cdot) - \mathbb{1}\{\cdot\}]\right| \le |Z|$ (i.e., $|Z|$ is a dominating function) and applying the dominated convergence theorem (since Z has finite expectation by A4), since $h_n \to 0$ by A6,

$$\begin{aligned}
&\lim_{h_n\to 0}\sup_{\beta\in\mathcal{B}}\left\|\mathrm{E}\left[\hat{M}_n(\beta,\tau)\right] - \mathrm{E}[Z(\mathbb{1}\{\Lambda(Y,X,\beta)\le 0\}-\tau)]\right\| \\
&= \lim_{h_n\to 0}\max_{\beta\in\mathcal{B}}\left\|\mathrm{E}\left\{Z\left[\tilde{I}\left(\frac{-\Lambda(Y,X,\beta)}{h_n}\right) - \mathbb{1}\{\Lambda(Y,X,\beta)\le 0\}\right]\right\}\right\| \\
&= \lim_{h_n\to 0}\left\|\mathrm{E}\left\{Z\left[\tilde{I}\left(\frac{-\Lambda(Y,X,\beta_n^*)}{h_n}\right) - \mathbb{1}\{\Lambda(Y,X,\beta_n^*)\le 0\}\right]\right\}\right\| \\
&= \left\|\mathrm{E}\left\{\lim_{h_n\to 0} Z\left[\tilde{I}\left(\frac{-\Lambda(Y,X,\beta_n^*)}{h_n}\right) - \mathbb{1}\{\Lambda(Y,X,\beta_n^*)\le 0\}\right]\right\}\right\| \\
&= 0
\end{aligned} \tag{A.1}$$

as long as there is no probability mass at $\Lambda(Y,X,\beta)=0$ for any $\beta\in\mathcal{B}$ and almost all Z, which is indeed true by Assumption A7. The notation β_n^* denotes the value attaining the maximum, which exists since $\mathcal{B}$ is compact by A3. □

Proof of Theorem 2.3. We show that the conditions of Theorem 5.9 in van der Vaart (1998) are satisfied. Alternatively, one could apply Theorem 2.1 in Newey and McFadden (1994, p. 2121), where $\hat{\beta}$ maximizes $\hat{Q}_n(\beta) \equiv -\|\hat{M}_n(\beta,\tau)\|$ with $\hat{Q}_n(\hat{\beta}) = 0$.

Combining results from A8 and Lemma 2.2 and the triangle inequality,

$$\begin{aligned}
&\sup_{\beta\in\mathcal{B}}\left|\hat{M}_n(\beta,\tau) - M(\beta,\tau)\right| \\
&= \sup_{\beta\in\mathcal{B}}\left|\hat{M}_n(\beta,\tau) - \mathrm{E}\big[\hat{M}_n(\beta,\tau)\big] + \mathrm{E}\big[\hat{M}_n(\beta,\tau)\big] - M(\beta,\tau)\right| \\
&\le \overbrace{\sup_{\beta\in\mathcal{B}}\left|\hat{M}_n(\beta,\tau) - \mathrm{E}\big[\hat{M}_n(\beta,\tau)\big]\right|}^{=o_p(1)\text{ by A8}} + \overbrace{\sup_{\beta\in\mathcal{B}}\left|\mathrm{E}\big[\hat{M}_n(\beta,\tau)\big] - M(\beta,\tau)\right|}^{=o_p(1)\text{ by Lemma 2.2}}
\end{aligned}$$

$$= o_p(1) + o_p(1) = o_p(1). \tag{A.2}$$

This satisfies the first condition of Theorem 5.9 in van der Vaart (1998, p. 46), or (combined with the continuity of $\|\cdot\|$) condition (iv) in Theorem 2.1 of Newey and McFadden (1994).

For the second condition of Theorem 5.9 in van der Vaart (1998), since $\mathcal{B}$ is a compact subset of Euclidean space, so is the set

$$\{\beta : \|\beta - \beta_{0\tau}\| \geq \epsilon, \beta \in \mathcal{B}\}$$

for any $\epsilon > 0$. Writing out

$$\begin{aligned} M(\beta, \tau) &= \mathrm{E}\{Z_i[\mathbb{1}\{\Lambda(Y_i, X_i, \beta) \leq 0\} - \tau]\} \\ &= \mathrm{E}(\mathrm{E}\{Z_i[\mathbb{1}\{\Lambda(Y_i, X_i, \beta) \leq 0\} - \tau] \mid Z_i\}) \\ &= \mathrm{E}\big\{Z_i\big[\mathrm{P}\big(\Lambda(Y_i, X_i, \beta) \leq 0 \mid Z_i\big) - \tau\big]\big\}, \end{aligned}$$

we see that the function $M(\beta, \tau)$ is continuous in β given A2 and A7. Note that A2 alone is not sufficient: it implies $\lim_{\delta \to 0} \Lambda(Y_i, X_i, \beta + \delta) \to \Lambda(Y_i, X_i, \beta)$ (for any realization $\omega \in \Omega$ in the implicit underlying probability space), but $\mathbb{1}\{\cdot \leq 0\}$ is not a continuous function. Specifically, it is discontinuous at zero, so the continuous mapping theorem only guarantees convergence (for $\omega \in \Omega$) where $\Lambda(Y_i, X_i, \beta) \neq 0$. Assumption A7 assumes this is a zero probability event (conditional on almost all Z_i), so $\mathbb{1}\{\Lambda(Y_i, X_i, \beta + \delta) \leq 0\}$ still converges almost surely to $\mathbb{1}\{\Lambda(Y_i, X_i, \beta) \leq 0\}$ as $\delta \to 0$ (i.e., the set of $\omega \in \Omega$ for which it does not converge has measure zero). Altogether, by A2 and A7, the bounded convergence theorem, and the continuous mapping theorem, writing $\Lambda_i \equiv \Lambda(Y_i, X_i, \beta)$,

$$\begin{aligned} &\lim_{\delta \to 0} \mathrm{P}\big(\Lambda(Y_i, X_i, \beta + \delta) \leq 0 \mid Z_i\big) \\ &= \lim_{\delta \to 0} \mathrm{E}\big(\mathbb{1}\{\Lambda(Y_i, X_i, \beta + \delta) \leq 0\} \mid Z_i\big) \\ &= \mathrm{E}\big(\lim_{\delta \to 0} \mathbb{1}\{\Lambda(Y_i, X_i, \beta + \delta) \leq 0\} \mid Z_i\big) \\ &= \mathrm{E}\big(\mathbb{1}\{\Lambda(Y_i, X_i, \beta) \leq 0\} \mid Z_i, \Lambda_i \neq 0\big) \overbrace{\mathrm{P}(\Lambda_i \neq 0 \mid Z_i)}^{=1 \text{ a.s., by A7}} \\ &\quad + \mathrm{E}\big(\lim_{\delta \to 0} \mathbb{1}\{\Lambda(Y_i, X_i, \beta + \delta) \leq 0\} \mid Z_i, \Lambda_i = 0\big) \overbrace{\mathrm{P}(\Lambda_i = 0 \mid Z_i)}^{=0 \text{ a.s., by A7}} \\ &= \mathrm{E}\big(\mathbb{1}\{\Lambda(Y_i, X_i, \beta) \leq 0\} \mid Z_i\big) \end{aligned}$$

almost surely.

Since a continuous function on a compact set attains a minimum, letting β^* denote the minimizer,

$$\inf_{\beta:\|\beta-\beta_{0\tau}\|\geq\epsilon} \|M(\beta,\tau)\| = \min_{\beta:\|\beta-\beta_{0\tau}\|\geq\epsilon} \|M(\beta,\tau)\| = \|M(\beta^*,\tau)\| > 0 \tag{A.3}$$

by A3, which says that for any $\beta \neq \beta_{0\tau}$, $M(\beta,\tau) \neq 0$, so $\|M(\beta^*,\tau)\| > 0$ (since $\|\cdot\|$ is a norm). Alternatively, for the conditions in Theorem 2.1 of Newey and McFadden (1994), (i) and (ii) are directly assumed in our A3, and (iii) is satisfied by the continuity of $M(\cdot,\tau)$ (as shown above).

Consistency follows by Theorem 5.9 in van der Vaart (1998) or Theorem 2.1 in Newey and McFadden (1994). □

Proof of Lemma 2.4. Decomposing into a mean-zero term and a "bias" term,

$$\sqrt{n}\hat{M}_n(\beta_{0\tau},\tau) = \overbrace{\sqrt{n}\{\hat{M}_n(\beta_{0\tau},\tau) - \mathrm{E}[\hat{M}_n(\beta_{0\tau},\tau)]\}}^{\xrightarrow{d}\mathrm{N}(0,\Sigma_{M\tau}) \text{ by A10}} + \overbrace{\sqrt{n}\,\mathrm{E}[\hat{M}_n(\beta_{0\tau},\tau)]}^{\text{want to show } o_p(1)}.$$

With iid data, Kaplan and Sun (2017, Thm. 1) show $\Sigma_{M\tau} = \tau(1-\tau)\,\mathrm{E}(Z_i Z_i^\top)$. The remainder of the proof shows that the second term is indeed $o_p(1)$, actually $o(1)$.

Let $\Lambda_i \equiv \Lambda(Y_i, X_i, \beta_{0\tau})$, with marginal PDF $f_\Lambda(\cdot)$ and conditional PDF $f_{\Lambda|Z}(\cdot \mid z)$ given $Z_i = z$. Given strict stationarity of the data, using the definitions in (2.8), assuming the support of Λ_i given $Z_i = z$ is the interval $[\Lambda_L(z), \Lambda_H(z)]$ with $\Lambda_L(z) \leq -h_n \leq h_n \leq \Lambda_H(z)$,

$$\begin{aligned}
&\mathrm{E}[\hat{M}_n(\beta_{0\tau},\tau)] = \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} g_n(Y_i, X_i, Z_i, \beta_{0\tau}, \tau)\right] = \mathrm{E}[g_n(Y_i, X_i, Z_i, \beta_{0\tau}, \tau)] \\
&= \mathrm{E}\Big\{Z_i[\tilde{I}(-\Lambda_i/h_n) - \tau]\Big\} \\
&= \mathrm{E}\Big\{Z_i\,\mathrm{E}[\tilde{I}(-\Lambda_i/h_n) - \tau \mid Z_i]\Big\} \\
&= \mathrm{E}\Big\{Z_i \overbrace{\int_{\Lambda_L(Z_i)}^{\Lambda_H(Z_i)} [\tilde{I}(-L/h_n) - \tau]\, dF_{\Lambda|Z}(L \mid Z_i)}^{\text{integrate by parts}}\Big\} \\
&= \mathrm{E}\Big\{Z_i\Big[\overbrace{\Big(\tilde{I}(-L/h_n) - \tau\Big) F_{\Lambda|Z}(L \mid Z_i)\Big|_{\Lambda_L(Z_i)}^{\Lambda_H(Z_i)}}^{=-\tau:\text{ use A5 and } \Lambda_H(Z_i)\geq h_n} \\
&\qquad - \int_{\Lambda_L(Z_i)}^{\Lambda_H(Z_i)} F_{\Lambda|Z}(L \mid Z_i)\ \overbrace{\tilde{I}'(-L/h_n)}^{=0 \text{ for } L\notin[-h_n,h_n]}\ (-h_n^{-1})\, dL\Big]\Big\}
\end{aligned}$$

$$
\begin{aligned}
&= \mathrm{E}\Big\{ Z_i \Big[-\tau + \overbrace{h_n^{-1} \int_{-h_n}^{h_n} F_{\Lambda|Z}(L \mid Z_i) \tilde{I}'(-L/h_n)\, dL}^{\text{change of variables to } v=-L/h_n} \Big] \Big\} \\
&= \mathrm{E}\Big\{ Z_i \Big[-\tau + \int_{-1}^{1} F_{\Lambda|Z}(-h_n v \mid Z_i) \tilde{I}'(v)\, dv \Big] \Big\} \\
&= \mathrm{E}\Big\{ Z_i \Big[-\tau + \int_{-1}^{1} \Big(\sum_{k=0}^{r} F_{\Lambda|Z}^{(k)}(0 \mid Z_i) \frac{(-h_n)^k v^k}{k!} \Big) \tilde{I}'(v)\, dv \Big] \Big\} \\
&\quad + \mathrm{E}\Big\{ Z_i \int_{-1}^{1} \overbrace{f_{\Lambda|Z}^{(r)}(-\tilde{h} v \mid Z_i)}^{\tilde{h}\in[0,h_n] \text{ (from MVT)}} \frac{(-h_n)^{r+1} v^{r+1}}{(r+1)!} \tilde{I}'(v)\, dv \Big\} \\
&= \mathrm{E}\Big\{ Z_i \Big[-\tau + \sum_{k=0}^{r} F_{\Lambda|Z}^{(k)}(0 \mid Z_i) \frac{(-h_n)^k}{k!} \overbrace{\int_{-1}^{1} v^k \tilde{I}'(v)\, dv}^{=0 \text{ for } 1 \le k \le r-1 \text{ by A5}} \Big] \Big\} \\
&\quad + \overbrace{O(h_n^{r+1})\, \mathrm{E}\Big\{ Z_i \int_{-1}^{1} \overbrace{f_{\Lambda|Z}^{(r)}(-\tilde{h} v \mid Z_i)}^{\text{bounded by A9}} v^{r+1} \tilde{I}'(v)\, dv \Big\}}^{O(1) \text{ by A5 and A9}} \\
&= \mathrm{E}\Big\{ Z_i \Big[-\tau + F_{\Lambda|Z}(0 \mid Z_i) + f_{\Lambda|Z}^{(r-1)}(0 \mid Z_i) \frac{(-h_n)^r}{r!} \int_{-1}^{1} v^r \tilde{I}'(v)\, dv \Big] \Big\} + O(h_n^{r+1}) \\
&= \mathrm{E}\{ Z_i[-\tau + \mathrm{E}(\mathbb{1}\{\Lambda_i \le 0\} \mid Z_i)]\} + \frac{\overbrace{(-h_n)^r}^{r \text{ is even}}}{r!} \Big[\int_{-1}^{1} v^r \tilde{I}'(v)\, dv \Big] \mathrm{E}\Big[Z_i f_{\Lambda|Z}^{(r-1)}(0 \mid Z_i) \Big] + O(h_n^{r+1}) \\
&= \overbrace{\mathrm{E}\{\mathrm{E}[Z_i(\mathbb{1}\{\Lambda_i \le 0\} - \tau) \mid Z_i]\}}^{=\mathrm{E}[Z_i(\mathbb{1}\{\Lambda_i \le 0\}-\tau)]=0 \text{ by A3}} + \frac{h_n^r}{r!} \Big[\int_{-1}^{1} v^r \tilde{I}'(v)\, dv \Big] \mathrm{E}\Big[Z_i f_{\Lambda|Z}^{(r-1)}(0 \mid Z_i) \Big] + O(h^{r+1}) \\
&= \frac{h_n^r}{r!} \Big[\int_{-1}^{1} v^r \tilde{I}'(v)\, dv \Big] \mathrm{E}\Big[Z_i f_{\Lambda|Z}^{(r-1)}(0 \mid Z_i) \Big] + O(h_n^{r+1}) = O(h_n^r).
\end{aligned}
$$

Thus, the result follows if $\sqrt{n} h_n^r = o(1)$, i.e., $h_n = o(n^{-1/(2r)})$ as in A6. □

Proof of Theorem 2.5. First, apply the continuous mapping theorem (CMT), using the non-singularity of G assumed in A9 and the result $\dot{M}_n(\tau) \xrightarrow{p} G$ in A11 to get $[\dot{M}_n(\tau)]^{-1} \xrightarrow{p} G^{-1}$. Using the CMT again, combine this with the results in (2.21) and Lemma 2.4:

$$
\begin{aligned}
\sqrt{n}(\hat{\beta}_\tau - \beta_{0\tau}) &= - \overbrace{[\dot{M}_n(\tau)]^{-1}}^{\text{use A11 and CMT}} \overbrace{\sqrt{n} \hat{M}_n(\beta_{0\tau}, \tau)}^{\text{use Lemma 2.4}} \\
&\xrightarrow{d} -G^{-1} \mathrm{N}(0, \Sigma_{M\tau}) \stackrel{d}{=} \mathrm{N}\big(0, G^{-1} \Sigma_{M\tau} [G^\top]^{-1}\big).
\end{aligned}
$$

□

B Primitive conditions for high-level assumptions

The following subsections discuss primitive conditions for the high-level Assumptions A8, A10, and A11.

B.1 Assumption A8

Assumption A8 is a high-level ULLN-type assumption. Intuitively, it holds under weak enough dependence and a moment restriction on Z_i. Howver, it is not trivial since most ULLNs assume a constant function $g(\cdot)$ instead of a function indexed by n. We provide an example of sufficient lower-level assumptions in Lemma B.1.

Lemma B.1. *Let Assumptions A1–A5 and A7 hold. Additionally, assume the following. (i)* $\big(\mathcal{B}, d(\cdot)\big)$ *is a metric space. (ii) Defining open balls* $B(\beta,\rho) \equiv \{\tilde{\beta} \in \mathcal{B} : d(\beta,\tilde{\beta}) < \rho\}$,

$$\begin{aligned} g_n^*(Y_i, X_i, Z_i, \beta, \tau, \rho) &\equiv \sup\Big\{g_n(Y_i, X_i, Z_i, \tilde{\beta}, \tau) : \tilde{\beta} \in B(\beta,\rho)\Big\}, \\ g_{*n}(Y_i, X_i, Z_i, \beta, \tau, \rho) &\equiv \inf\Big\{g_n(Y_i, X_i, Z_i, \tilde{\beta}, \tau) : \tilde{\beta} \in B(\beta,\rho)\Big\} \end{aligned} \tag{B.1}$$

are random variables for all i, $\beta \in \mathcal{B}$, and sufficiently small ρ (which may depend on β), where the sup *and* inf *are taken separately for each element of the vector. (iii) A pointwise WLLN holds for the random vectors in* (B.1)*, for each $\beta \in \mathcal{B}$. (iv) The data are strictly stationary. Then, for a fixed $\tau \in (0,1)$, using the definition in* (2.8),

$$\sup_{\beta\in\mathcal{B}}\Big|\hat{M}_n(\beta,\tau) - \mathrm{E}[\hat{M}_n(\beta,\tau)]\Big| = o_p(1).$$

Proof. We show that the theorem in Andrews (1987) applies. The theorem concerns a uniform law of large numbers (ULLN) for a sample average of functions of the data. By Comment 6 in Andrews (1987), both the data and the functions may be indexed by both i and n. In our case, the function $g_n(\cdot)$ is not indexed by i but must be indexed by n since it depends on the sequence of bandwidths, h_n. We continue to index the observations only by i but note that triangular arrays are permitted by Andrews (1987). Since Andrews (1987) presumes a scalar-valued function, we write $g_n(\cdot)$; since the dimension of $g_n(\cdot)$ is fixed and finite, uniformity extends immediately to the vector.

Assumption A1 in Andrews (1987) is simply that $\mathcal{B}$ is compact, which is in our A3. (More recent work shows that "compact" can be replaced by "totally bounded" under a metric; see Andrews (1992) and Pötscher and Prucha (1994).)

Assumption A2(a) in Andrews (1987) is a technical measurability assumption; this is assumption (ii) in the statement of Lemma B.1.

Assumption A2(b) in Andrews (1987) is assumption (iii) in the statement of the lemma. There are many WLLNs for weakly dependent triangular arrays, where dependence is quantified and restricted in various ways; for example, see Theorem 2 in Andrews (1988) and the theorems in de Jong (1998). With iid sampling, sufficient primitive conditions for a WLLN are already in our A4 and A5, respectively: a) $\mathrm{E}\big(\|Z_i\|^2\big) < \infty$, and b) $\tilde{I}(\cdot)$ is bounded. From A5, $-2 \leq \tilde{I}(\cdot) - \tau \leq 2$, so we have the dominating function $|g_n(Y_i, X_i, Z_i, \beta, \tau)| \leq 2|Z_i|$. Consequently,

$$|g_n^*(Y_i, X_i, Z_i, \beta, \tau, \rho)| \leq 2|Z_i|, \quad |g_{*n}(Y_i, X_i, Z_i, \beta, \tau, \rho)| \leq 2|Z_i|.$$

If the data are iid, then $g_n^*(Y_i, X_i, Z_i, \beta, \tau, \rho)$ is a row-wise iid triangular array. Thus, a sufficient condition for a WLLN is $\sup_n \mathrm{E}\big[\|g_n^*\|^2\big] < \infty$ (as can be shown with Markov's inequality). This condition holds since $\sup_n \mathrm{E}\big[\|g_n^*\|^2\big] \leq \mathrm{E}\big[\|2Z_i\|^2\big] < \infty$ by A4. An extension to independent but not identical sampling follows from a Lindeberg condition for Z_i. A pointwise WLLN continues to hold with dependence, too, as long as the dependence is not too strong.

Assumption A3 in Andrews (1987) in our notation is

$$\lim_{\rho \to 0} \sup_{n \geq 1} \left| \frac{1}{n} \sum_{i=1}^{n} \{\mathrm{E}[g_n^*(Y_i, X_i, Z_i, \beta, \tau, \rho)] - \mathrm{E}[g_n(Y_i, X_i, Z_i, \beta, \tau)]\} \right| = 0, \tag{B.2}$$

and similarly when replacing g_n^* with g_{*n}. Since g_n varies with n but not i, the strict stationarity in assumption (iv) in Lemma B.1 implies the summands do not vary with i, which simplifies (B.2) to be

$$\lim_{\rho \to 0} \sup_{n \geq 1} |\Delta_n| = 0, \ \Delta_n \equiv \mathrm{E}[g_n^*(Y_i, X_i, Z_i, \beta, \tau, \rho)] - \mathrm{E}[g_n(Y_i, X_i, Z_i, \beta, \tau)], \ \Delta_\infty \equiv \lim_{n \to \infty} \Delta_n. \tag{B.3}$$

Strict stationarity is not necessary, though, as long as (B.2) still holds.

A necessary condition for (B.3) is pointwise convergence $\lim_{\rho \to 0} \Delta_n = 0$ for each fixed n. By A3 and A5, $g_{ni}(\beta, \tau)$ is continuous and even differentiable in β. Additionally, as noted above, $2|Z_i|$ is a dominating function with finite expectation (by A4), so the dominated convergence theorem gives

$$\begin{aligned}\lim_{\rho\to 0}\Delta_n &= \lim_{\rho\to 0}\mathrm{E}[g_n^*(Y_i,X_i,Z_i,\beta,\tau,\rho)-g_n(Y_i,X_i,Z_i,\beta,\tau)]\\ &= \mathrm{E}\Big\{\lim_{\rho\to 0}[g_n^*(Y_i,X_i,Z_i,\beta,\tau,\rho)-g_n(Y_i,X_i,Z_i,\beta,\tau)]\Big\}\\ &= \mathrm{E}\{0\}=0,\end{aligned}$$

and similarly for g_{*n}, for any n.

For Δ_∞, as $n\to\infty$, we can again move the limit inside expectations by the dominated convergence theorem, so

$$\begin{aligned}\lim_{n\to\infty}\mathrm{E}[g_n(Y_i,X_i,Z_i,\beta,\tau)] &= \mathrm{E}\Big\{\lim_{n\to\infty}g_n(Y_i,X_i,Z_i,\beta,\tau)\Big\}\\ &= \mathrm{E}\{Z_i[\mathbb{1}\{\Lambda(Y_i,X_i,\beta)\le 0\}-\tau]\}\\ &= \mathrm{E}\{\mathrm{E}[Z_i(\mathbb{1}\{\Lambda(Y_i,X_i,\beta)\le 0\}-\tau)\mid Z_i]\}\\ &= \mathrm{E}\{Z_i[\mathrm{E}(\mathbb{1}\{\Lambda(Y_i,X_i,\beta)\le 0\}\mid Z_i)-\tau]\}\\ &= \mathrm{E}\{Z_i[\mathrm{P}(\Lambda(Y_i,X_i,\beta)\le 0\mid Z_i)-\tau]\}.\end{aligned}$$

Technically, since $\tilde{I}(0/h_n)=0.5$ for any $h_n>0$, the function $\tilde{I}(\cdot/h_n)\to\mathbb{1}\{\cdot\ge 0\}-0.5\,\mathbb{1}\{\cdot=0\}$ as $n\to\infty$, so we have

$$\begin{aligned}&\mathrm{E}\{Z_i[\mathrm{E}(\mathbb{1}\{\Lambda(Y_i,X_i,\beta)\le 0\}-0.5\,\mathbb{1}\{\Lambda(Y_i,X_i,\beta)=0\}\mid Z_i)-\tau]\}\\ &=\mathrm{E}\left\{Z_i\left[\mathrm{P}(\Lambda(Y_i,X_i,\beta)\le 0\mid Z_i)-0.5\overbrace{\mathrm{P}(\Lambda(Y_i,X_i,\beta)=0\mid Z_i)}^{=0\text{ a.s. by A7}}-\tau\right]\right\}.\end{aligned}$$

That is, by A7, the 0.5 adjustment corresponds to a zero probability event that does not affect the overall expectation. For g_n^*, similarly,

$$\lim_{n\to\infty}\mathrm{E}[g_n^*(Y_i,X_i,Z_i,\beta,\tau,\rho)]=\mathrm{E}\left\{\sup_{\tilde{\beta}\in B(\beta,\rho)}Z_i\Big[\mathrm{P}\Big(\Lambda(Y_i,X_i,\tilde{\beta})\le 0\mid Z_i\Big)-\tau\Big]\right\}.$$

Consequently,

$$\Delta_\infty=\mathrm{E}\left\{Z_i[\mathrm{P}(\Lambda(Y_i,X_i,\beta)\le 0\mid Z_i)-\tau]-\sup_{\tilde{\beta}\in B(\beta,\rho)}Z_i\Big[\mathrm{P}\Big(\Lambda(Y_i,X_i,\tilde{\beta})\le 0\mid Z_i\Big)-\tau\Big]\right\}.$$

For this to have a limit of zero as $\rho\to 0$ again requires continuity in ρ, but the necessary and sufficient conditions are different than for fixed n. Sufficient conditions here are found in A2 and A7: $\Lambda(\cdot)$ is continuous in β, and for any $\beta \in \mathcal{B}$ and almost all $z \in \mathcal{Z}$, the conditional distribution of $\Lambda(Y_i, X_i, \beta)$ given $Z_i = z$ is continuous in a neighborhood of zero. For example, if $\Lambda(Y_i, X_i, \beta) = Y_i - X_i^\top \beta$, then it is sufficient that Y_i has a continuous distribution given almost all $Z_i = z$.

Given $\lim_{\rho\to 0} \Delta_n = 0$ for any $n < \infty$ and $n = \infty$, the conclusion $\lim_{\rho\to 0} \sup_{n\geq 1}|\Delta_n| = 0$ follows because the supremum is attained: $\sup_{n\geq 1}|\Delta_n| = |\Delta_k|$ for some $k \geq 1$ or $k = \infty$. If instead $\lim_{\rho\to 0} \Delta_n = 0$ only for $n \geq 1$, and not with $\lim_{n\to\infty}$, then it would be possible for all $\lim_{\rho\to 0} \Delta_n = 0$ pointwise but $\lim_{\rho\to 0} \sup_{n\geq 1} \Delta_n \neq 0$; for example if $\Delta_n = (1 - 1/n)^{1/\rho}$ then all $\lim_{\rho\to 0} \Delta_n = 0$ but $\sup_{n\geq 1} \Delta_n = 1$ for any ρ, so $\lim_{\rho\to 0} \sup_{n\geq 1} \Delta_n = 1$. This is why the calculations for Δ_∞ are necessary.

Having verified A1, A2, and A3 in Andrews (1987), his theorem applies, yielding the desired ULLN. □

B.2 Assumption A10

To establish Assumption A10, with iid data, the Lindeberg–Feller CLT can be applied as in the proof of Theorem 1 in Kaplan and Sun (2017). More generally, A10 can hold under weak dependence. For example, Theorem 3.13 in Wooldridge (1986, Ch. 2), as reproduced in Proposition 1 of Andrews (1991a), is a CLT for near epoch dependent triangular arrays that holds under some moment and dependence restrictions. The moment restriction, condition (ii), in our notation is $\mathrm{E}\{\|g_{ni}(\beta_{0\tau}, \tau)\|^{2+\epsilon}\} < \infty$ for some $\epsilon > 0$. Since $|g_{ni}(\beta_{0\tau}, \tau)| < 2|Z_i|$, if the underlying Z_i are strictly stationary then $\mathrm{E}(\|Z_i\|^{2+\epsilon}) < \infty$ is sufficient; triangular array data are allowed if $\sup_{i\leq n, n\geq 1} \mathrm{E}(\|Z_{ni}\|^{2+\epsilon}) < \infty$ for some $\epsilon > 0$.

B.3 Assumption A11

Unfortunately, for multiple reasons, Assumption A11 cannot be deduced simply by applying a result like Lemma 4.3 in Newey and McFadden (1994, p. 2156). Fortunately, it is closely related to the well-studied result of consistency of the kernel estimator for the quantile regression asymptotic covariance matrix. Since the argument is the same for each row in the matrix, we consider row k. Plugging in definitions,

$$\dot{M}_n^{(k,\cdot)}(\tau) = \nabla_{\beta^\top} \hat{M}_n^{(k)}(\tilde{\beta}_\tau, \tau) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \beta^\top} g_{ni}^{(k)}(\beta, \tau)\Big|_{\beta=\tilde{\beta}_\tau}$$

$$= \frac{1}{n}\sum_{i=1}^{n} Z_i^{(k)} \tilde{I}'\bigl(-\Lambda(Y_i, X_i, \tilde{\beta}_\tau)/h_n\bigr)(-h_n^{-1}) D_i(\tilde{\beta}_\tau)^\top, \tag{B.4}$$

$$D_i(b) \equiv \frac{\partial}{\partial\beta}\Lambda(Y_i, X_i, \beta)\bigg|_{\beta=b}. \tag{B.5}$$

By A5, $\tilde{I}'(\cdot)$ is a kernel function. The RHS of (B.4) is closely related to the kernel estimator of the usual quantile regression estimator's asymptotic covariance matrix initially proposed under censoring by Powell (1984, eqn. (5.6)) and without censoring in Powell (1991), but with two differences: 1) we have $\tilde{\beta}$ instead of $\hat{\beta}$, 2) we have a more general model. As a special case of our model, the "usual" quantile regression model would have $Z_i = X_i$ and $\Lambda(Y_i, X_i, \beta) = Y_i - X_i^\top\beta$, so $\nabla_\beta\Lambda(Y_i, X_i, \tilde{\beta}) = -X_i$, and (B.4) simplifies to

$$\frac{1}{n}\sum_{i=1}^{n} X_i^{(k)} \tilde{I}'\bigl((X_i^\top\tilde{\beta}_\tau - Y_i)/h_n\bigr)(-h_n^{-1})(-X_i^\top) = \frac{1}{nh_n}\sum_{i=1}^{n}\tilde{I}'\left(\frac{Y_i - X_i^\top\tilde{\beta}_\tau}{h_n}\right) X_i^{(k)} X_i^\top, \tag{B.6}$$

using the symmetry $\tilde{I}'(-u) = \tilde{I}'(u)$ from A5. Since $\tilde{\beta}_\tau$ lies between $\beta_{0\tau}$ and $\hat{\beta}_\tau$, proofs using $\hat{\beta}$ still hold since $\sqrt{n}$-consistency of $\hat{\beta}$ implies $\sqrt{n}$-consistency of $\tilde{\beta}$.

For (B.6), Kato (2012) shows consistency (i.e., our A11) with both iid and weakly dependent data. In fact, he shows the stronger result of asymptotic normality, so some of his assumptions may be weakened if only consistency is required; for example, his $h_n\sqrt{n}/\log(n) \to \infty$ in Assumption 13 can be (slightly) weakened to $h_n\sqrt{n} \to \infty$, and not as many moments of X_i are required. Specifically, Kato (2012) considers strictly stationary β-mixing data, and the mixing coefficients, moments of X_i, and bandwidth rate are jointly restricted in his Assumptions 9, 10, and 13.

For the more general (B.4), similar conditions are sufficient if $D_i(\beta) = D_i$, a random variable depending on Y_i and X_i but not the argument β. This occurs if (and only if) the residual function $\Lambda(\cdot)$ is linear-in-parameters. Then, D_i replaces one of the X_i in Kato (2012), while Z_i replaces the other. The most notable restriction is on moments of D_i (which implies a certain number of finite moments for Y_i and X_i) in addition to Z_i (which is already in A4). Many economic variables are bounded or reasonably have infinite moments (e.g., a normal distribution), in which case such moment assumptions are not binding. If $D_i(b)$ does depend on its argument, then an extension of the argument itself in Kato (2012) is necessary.

In (B.6), X_i plays the roles of both the derivative of $\Lambda(Y_i, X_i, \beta)$ and the instrument vector, so the PDF in G is just conditional on X_i; more generally, both the instrument vector and derivative must be conditioned on. This can be seen by computing the expectation of

(B.4) in a similar manner to the proof of Lemma 2.4. After replacing $\tilde{\beta}_\tau = \beta_{0\tau} + O_p(n^{-1/2})$ and dropping the remainder, letting $D_i \equiv \nabla_\beta \Lambda(Y_i, X_i, \beta_{0\tau})$ and $\Lambda_i \equiv \Lambda(Y_i, X_i, \beta_{0\tau})$,

$$
\begin{aligned}
\mathrm{E}[\dot{M}_n^{(k,\cdot)}(\tau)] &\doteq \frac{1}{n}\sum_{i=1}^{n} \overbrace{\mathrm{E}\Big\{ Z_i^{(k)} \tilde{I}'\big(-\Lambda(Y_i, X_i, \beta_{0\tau})/h_n\big)(-h_n^{-1}) \frac{\partial}{\partial \beta^\top}\Lambda(Y_i, X_i, \beta)\Big|_{\beta=\tilde{\beta}_\tau} \Big\}}^{\text{same for all } i \text{ by A1}} \\
&= \mathrm{E}\Big[Z_i^{(k)} D_i^\top (-h_n^{-1}) \tilde{I}'\big(-\Lambda_i/h_n\big)\Big] \\
&= \mathrm{E}\Big\{ Z_i^{(k)} D_i^\top \,\mathrm{E}\Big[(-h_n^{-1})\tilde{I}'\big(-\Lambda_i/h_n\big) \mid Z_i, D_i\Big]\Big\} \\
&= \mathrm{E}\Big\{ Z_i^{(k)} D_i^\top \int (-h_n^{-1})\tilde{I}'\big(-L/h_n\big) f_{\Lambda|Z,D}(L \mid Z_i, D_i)\, dL \Big\} \\
&= \mathrm{E}\Big\{ Z_i^{(k)} D_i^\top \int_{-1}^{1} -\tilde{I}'(v) f_{\Lambda|Z,D}(-h_n v \mid Z_i, D_i)\, dv \Big\} \\
&= -\,\mathrm{E}\Big\{ Z_i^{(k)} D_i^\top \int_{-1}^{1} \tilde{I}'(v)\big[f_{\Lambda|Z,D}(0 \mid Z_i, D_i) \\
&\qquad\qquad - f'_{\Lambda|Z,D}(-h_n v \mid Z_i, D_i) h_n v + \cdots \big]\, dv \Big\} \\
&= -\,\mathrm{E}\Big[Z_i^{(k)} D_i^\top f_{\Lambda|Z,D}(0 \mid Z_i, D_i)\Big] \\
&\quad - h_n^r\, \mathrm{E}\Big\{ Z_i^{(k)} D_i^\top \int_{-1}^{1} \tilde{I}'(v)(v^r/r!) f^{(r)}_{\Lambda|Z,D}(-\tilde{h} v \mid Z_i, D_i)\, dv \Big\} \\
&= -\,\mathrm{E}\Big[Z_i^{(k)} D_i^\top f_{\Lambda|Z,D}(0 \mid Z_i, D_i)\Big] + O(h_n^r).
\end{aligned}
$$

C Computational details

C.1 Bandwidth selection

For consistency, only $h_n \to 0$ is required, so just picking the smallest possible bandwidth seems reasonable and performs well in simulations.

For the estimator to attain asymptotic normality, with dependent data, it seems $nh_n^2 \to \infty$ is required for Assumption A11, which implies h_n must be larger than $n^{-1/2}$. The smallest possible bandwidth is too small. Additionally, in the linear IVQR case with iid sampling, Kaplan and Sun (2017) find the (approximate) MSE-optimal bandwidth rate to be $n^{-1/(2r-1)}$, where r is the order of the kernel function $\tilde{I}'(\cdot)$, like $r = 4$ for the function used in the code. To try to get the rate $n^{-1/7}$, we first find the smallest possible bandwidth h_0 and the corresponding number of smoothed observations n_0, i.e., for how many i does $-h \leq \Lambda(Y_i, X_i, \hat{\beta}_\tau) \leq h$. Then, we take $h = h_0(n^{6/7}/n_0)$. This is ad hoc, but appears sufficient for at least Studentized bootstrap tests to be accurate in samples of even $n = 100$.

Alternatively, one could experiment with a variation of the AMSE-optimal bandwidth (for estimating G) proposed in Kato (2012). With a linear model, this is easily done by replacing one of the X in the rule-of-thumb formula in Kato (2012) with our Z, leaving the other X as the regressor vector that in our notation includes the endogenous regressors in Y and exogenous regressors in X. This is what we used for our simulations, where it seemed to work better than the more ad hoc procedure above.

References

Amemiya, T. (1982): "Two Stage Least Absolute Deviations Estimators," *Econometrica*, 50, 689–711.

Andrews, D. W. K. (1987): "Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers," *Econometrica*, 55, 1465–1471.

——— (1988): "Laws of Large Numbers for Dependent Non-identically Distributed Random Variables," *Econometric Theory*, 4, 458–467.

——— (1991a): "An Empirical Process Central Limit Theorem for Dependent Non-identically Distributed Random Variables," *Journal of Multivariate Analysis*, 38, 187–203.

——— (1991b): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

——— (1992): "Generic Uniform Convergence," *Econometric Theory*, 8, 241–257.

Andrews, I. and A. Mikusheva (2016): "Conditional Inference With a Functional Nuisance Parameter," *Econometrica*, 84, 1571–1612.

Angrist, J., V. Chernozhukov, and I. Fernández-Val (2006): "Quantile Regression Under Misspecification, with an Application to the U.S. Wage Structure," *Econometrica*, 74, 539–563.

Borchers, H. W. (2015): *pracma: Practical Numerical Math Functions*, R package version 1.8.3.

Buchinsky, M. (1998): "Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research," *Journal of Human Resources*, 33, 88–126.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008): "Bootstrap-Based Improvements for Inference with Clustered Errors," *Review of Economics and Statistics*, 90, 414–427.

Campbell, J. Y. (2003): "Consumption-Based Asset Pricing," in *Handbook of the Economics of Finance: Financial Markets and Asset Pricing*, ed. by G. M. Constantinides, M. Harris, and R. M. Stulz, North Holland, vol. 1, Part B, chap. 13, 803–887.

Campbell, J. Y. and N. G. Mankiw (1989): "Consumption, Income, and Interest Rates: Reinterpreting the Time Series Evidence," in *NBER Macroeconomics Annual 1989*, ed. by O. J. Blanchard and S. Fischer, Cambridge, MA: MIT Press, 185–216.

Campbell, J. Y. and L. M. Viceira (1999): "Consumption and Portfolio Decisions

When Expected Returns Are Time Varying," *Quarterly Journal of Economics*, 114, 433–495.

Chambers, C. P. (2009): "An Axiomatization of Quantiles on the Domain of Distribution Functions," *Mathematical Finance*, 19, 335–342.

Chen, L.-Y. and S. Lee (2017): "Exact Computation of GMM Estimators for Instrumental Variable Quantile Regression Models," Working paper, available at https://arxiv.org/abs/1703.09382.

Chen, X., V. Chernozhukov, S. Lee, and W. K. Newey (2014): "Local Identification of Nonparametric and Semiparametric Models," *Econometrica*, 82, 785–809.

Chen, X. and Z. Liao (2015): "Sieve Semiparametric Two-Step GMM under Weak Dependence," *Journal of Econometrics*, 189, 163–186.

Chen, X., O. Linton, and I. van Keilegom (2003): "Estimation of Semiparametric Models When the Criterion Function is not Smooth," *Econometrica*, 71, 1591–1608.

Chen, X. and D. Pouzo (2009): "Efficient Estimation of Semiparametric Conditional Moment Models with Possibly Nonsmooth Residuals," *Journal of Econometrics*, 152, 46–60.

——— (2012): "Estimation of Nonparametric Conditional Moment Models with Possibly Nonsmooth Moments," *Econometrica*, 80, 277–322.

Cheng, X., Z. Liao, and R. Shi (2016): "An Averaging GMM Estimator Robust to Misspecification," Working paper, available at http://www.econ.ucla.edu/liao/papers_pdf/ChengLiaoShi_AvgGMM.pdf.

Chernozhukov, V. and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245–261.

——— (2006): "Instrumental Quantile Regression Inference for Structural and Treatment Effect Models," *Journal of Econometrics*, 132, 491–525.

——— (2008): "Instrumental Variable Quantile Regression: A Robust Inference Approach," *Journal of Econometrics*, 142, 379–398.

Chernozhukov, V., C. Hansen, and M. Jansson (2009): "Finite Sample Inference for Quantile Regression Models," *Journal of Econometrics*, 152, 93–103.

Chernozhukov, V., C. Hansen, and K. Wüthrich (2017): "Instrumental Variable Quantile Regression," in *Handbook of Quantile Regression*, ed. by R. Koenker, V. Chernozhukov, X. He, and L. Peng, CRC/Chapman-Hall, forthcoming.

Chernozhukov, V. and H. Hong (2003): "An MCMC Approach to Classical Estimation," *Journal of Econometrics*, 115, 293–346.

Cochrane, J. H. (2005): *Asset Pricing*, Princeton, NJ: Princeton University Press.

de Castro, L. and A. F. Galvao (2017): "Dynamic Quantile Models of Rational Behavior," University of Iowa, mimeo.

de Jong, R. M. (1998): "Weak Laws of Large Numbers for Dependent Random Variables," *Annales d'Économie et de Statistique*, 209–225.

Fernandes, M., E. Guerre, and E. Horta (2017): "Smoothing Quantile Regressions," Mimeo, available at http://bibliotecadigital.fgv.br/dspace/handle/10438/18390.

Galvao, A. F. and K. Kato (2016): "Smoothed Quantile Regression for Panel Data," *Journal of Econometrics*, 193, 92–112.

Giovannetti, B. C. (2013): "Asset Pricing under Quantile Utility Maximization," *Review* *of Financial Economics*, 22, 169–179.

Hall, R. E. (1988): "Intertemporal Substitution in Consumption," *Journal of Political Economy*, 96, 339–357.

Hansen, B. E. (2017): "A Stein-Like 2SLS Estimator," *Econometric Reviews*, XX, XXX.

Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

Hansen, L. P. and K. J. Singleton (1983): "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns," *Journal of Political Economy*, 92, 249–265.

Horowitz, J. L. (1992): "A Smoothed Maximum Score Estimator for the Binary Response Model," *Econometrica*, 60, 505–531.

——— (1998): "Bootstrap Methods for Median Regression Models," *Econometrica*, 66, 1327–1351.

Jun, S. J. (2008): "Weak Identification Robust Tests in an Instrumental Quantile Model," *Journal of Econometrics*, 144, 118–138.

Kaplan, D. M. and Y. Sun (2017): "Smoothed Estimating Equations for Instrumental Variables Quantile Regression," *Econometric Theory*, 33, 105–157.

Kato, K. (2012): "Asymptotic Normality of Powell's Kernel Estimator," *Annals of the Institute of Statistical Mathematics*, 64, 255–273.

Kinal, T. W. (1980): "The Existence of Moments of k-Class Estimators," *Econometrica*, 48, 241–249.

Koenker, R. and G. Bassett, Jr. (1978): "Regression Quantiles," *Econometrica*, 46, 33–50.

Lancaster, T. and S. J. Jun (2010): "Bayesian quantile regression methods," *Journal of Applied Econometrics*, 25, 287–307.

Ljungqvist, L. and T. J. Sargent (2012): *Recursive Macroeconomic Theory*, Cambridge, Massachusetts: MIT Press, 3rd ed.

Lucas, R. E. (1978): "Asset Prices in an Exchange Economy," *Econometrica*, 46, 1429–1446.

MaCurdy, T. (2007): "A Practitioner's Approach to Estimating Intertemporal Relationships Using Longitudinal Data: Lessons from Applications in Wage Dynamics," in *Handbook of Econometrics*, ed. by J. J. Heckman and E. E. Leamer, Elsevier, vol. 6A, chap. 62, 4057–4167.

MaCurdy, T. and H. Hong (1999): "Smoothed Quantile Regression in Generalized Method of Moments," Mimeo.

MaCurdy, T. and C. Timmins (2001): "Bounding the Influence of Attrition on Intertemporal Wage Variation in the NLSY," Mimeo, available at `http://public.econ.duke.edu/~timmins/bounds.pdf`.

Manski, C. F. (1988): "Ordinal Utility Models of Decision Making under Uncertainty," *Theory and Decision*, 25, 79–104.

Newey, W. K. (2004): "Efficient Semiparametric Estimation via Moment Restrictions," *Econometrica*, 72, 1877–1897.

Newey, W. K. and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, ed. by R. F. Engle and D. L. McFadden, Elsevier, vol. 4, chap. 36, 2111–2245.

Newey, W. K. and J. L. Powell (1990): "Efficient Estimation of Linear and Type I Censored Regression Models Under Conditional Quantile Restrictions," *Econometric Theory*, 6, 295–317.

Newey, W. K. and K. D. West (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Oberhofer, W. and H. Haupt (2016): "Asymptotic Theory for Nonlinear Quantile Regression under Weak Dependence," *Econometric Theory*, 32, 686–713.

Ogaki, M. and C. M. Reinhart (1998): "Measuring Intertemporal Substitution: The Role of Durable Goods," *Journal of Political Economy*, 106, 1078–1098.

Otsu, T. (2008): "Conditional Empirical Likelihood Estimation and Inference for Quantile Regression Models," *Journal of Econometrics*, 142, 508–538.

Politis, D. N. and J. P. Romano (1994): "The Stationary Bootstrap," *Journal of the American Statistical Association*, 89, 1303–1313.

Pötscher, B. M. and I. R. Prucha (1994): "Generic Uniform Convergence and Equicontinuity Concepts for Random Functions: An Exploration of the Basic Structure," *Journal of Econometrics*, 60, 23–63.

Powell, J. L. (1984): "Least Absolute Deviations Estimation for the Censored Regression Model," *Journal of Econometrics*, 25, 303–325.

——— (1991): "Estimation of Monotonic Regression Models under Quantile Restrictions," in *Nonparametric and semiparametric methods in econometrics and statistics: proceedings of the Fifth International Symposium in Economic Theory and Econometrics*, ed. by W. A. Barnett, J. Powell, and G. Tauchen, Cambridge University Press, 357–384.

——— (1994): "Estimation of Semiparametric Models," in *Handbook of Econometrics*, ed. by R. F. Engle and D. L. McFadden, Elsevier, vol. 4, chap. 41, 2443–2521.

R Core Team (2013): *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Rostek, M. (2010): "Quantile Maximization in Decision Theory," *Review of Economic Studies*, 77, 339–371.

Toda, A. A. and K. Walsh (2015): "The Double Power Law in Consumption and Implications for Testing Euler Equations," *Journal of Political Economy*, 123, 1177–1200.

Toda, A. A. and K. J. Walsh (2017): "Fat Tails and Spurious Estimation of Consumption-Based Asset Pricing Models," *Journal of Applied Econometrics*, forthcoming.

van der Vaart, A. W. (1998): *Asymptotic Statistics*, Cambridge: Cambridge University Press.

Whang, Y.-J. (2006): "Smoothed Empirical Likelihood Methods for Quantile Regression Models," *Econometric Theory*, 22, 173–205.

Wooldridge, J. M. (1986): "Asymptotic Properties of Econometric Estimators," Ph.D. thesis, Department of Economics, University of California, San Diego.

Yogo, M. (2004): "Estimating the Elasticity of Intertemporal Substitution When Instruments are Weak," *Review of Economics and Statistics*, 86, 797–810.