The Econometrics Journal

Econometrics Journal (2023), volume **26**, pp. 350–377.
https://doi.org/10.1093/ectj/utad010

A first-stage representation for instrumental variables quantile regression

JAVIER ALEJO[†], ANTONIO F. GALVAO[‡] AND GABRIEL MONTES-ROJAS[§]

[†]*IECON-Universidad de la República, Montevideo, Uruguay*

[‡]*Michigan State University, East Lansing, USA.*
E-mail: agalvao@msu.edu

[§]*CONICET and IIEP-BAIRES, Universidad de Buenos Aires, CABA, Argentina*

First version received: 30 March 2023; final version accepted: 31 March 2023.

Summary: This paper develops a first-stage linear regression representation for an instrumental variables (IV) quantile regression (QR) model. The quantile first stage is analogous to the least-squares case, i.e., a linear projection of the endogenous variables on the instruments and other exogenous covariates, with the difference that the QR case is a weighted projection. The weights are given by the conditional density function of the innovation term in the QR structural model, at a given quantile. We also show that the required Jacobian identification conditions for IVQR models are embedded in the quantile first stage. We then suggest procedures to evaluate the validity of instruments by evaluating their statistical significance using the first-stage representation. Monte Carlo experiments provide numerical evidence that the proposed tests work as expected in terms of empirical size and power. An empirical application illustrates the methods.

Keywords: *first stage*, *instrumental variables*, *quantile regression*.

JEL codes: *C14*, *C26*.

1. INTRODUCTION

Instrumental variables (IV) methods are one of the main workhorses to estimate causal relationships in empirical analysis. Standard IV regression methods stress that for instruments to be valid they must be exogenous and must be related to the endogenous variables. The latter condition is usually evaluated by using a first-stage auxiliary regression, where a linear model is used to make inference on the degree of association of the IV and the endogenous variables. While this is usually accepted as a valid procedure, its representation is in fact specific to the two-stage least squares (2SLS) model for average models. This paper derives a first-stage representation for quantile regression (QR) models.

Several IV methods have been proposed in QR to solve endogeneity when the covariates are correlated with the error term in a regression model. Chernozhukov and Hansen (2004, 2005, 2006, 2008) (CH hereafter) develop an instrumental variables quantile regression (IVQR) procedure that has been applied in several contexts. This is also usually termed the inverse QR estimator, although we will use CH-IVQR to refer to this specific estimator. This method is one of the most prolific approaches in terms of subsequent work, as it provides a general IV procedure for solving endogeneity of regressors (see, e.g., Chernozhukov et al., 2007, 2009; Galvao, 2011; Chetverikov et al., 2016). We refer to Chernozhukov et al. (2020) for an overview of IVQR.[1]

CH highlight that the IVQR method is a simple solution for endogeneity and a 2SLS analogue.[2] However, the first stage of the IVQR has not been considered as it is implemented as an inverse QR estimator. This estimator contrasts with alternative IV procedures where the first stage is explicitly implemented. For instance, Amemiya (1982), Powell (1983), Chen and Portnoy (1996), and Kim and Muller (2004) use an explicit first stage that fits the endogenous variable(s) as a function of exogenous covariates and IV, and this is then plugged in a second stage. Lee (2007) also adopts a two-step control-function approach where in the first step consists of the estimation of the residuals of the reduced-form equation for the endogenous explanatory variable. Ma and Koenker (2006) present an estimator for a recursive structural equation model.

This paper shows that a first-stage regression model can be explicitly recovered from the CH-IVQR estimator. The first-stage IVQR (FS-IVQR) is a linear projection of the endogenous variables on the instruments and other exogenous variables, with the difference that the QR case is a weighted regression, that is, it has the representation of a weighted least-squares (WLS) regression of the endogenous variable(s) on the IV and the exogenous regressors. The weights are given by the conditional density function of the innovation term in the QR structural model, conditional on the endogenous and exogenous covariates together with the instruments, at a given quantile. This result provides a clear analogy between the first stage in 2SLS and IVQR. The derivation of the result is simple, but conceptually important. We write the IVQR model as a constrained Lagrangian optimisation problem and show that one of the restrictions that must be satisfied is the analogue of the first stage.

The CH-IVQR method requires an identification condition that is based on the full rank of the Jacobian for the exogeneity of the instruments. The lack of identification when the Jacobian is not full rank implies that estimating the parameters can be extremely difficult and the first-order asymptotics can be a poor guide of the actual sampling distributions (see, e.g., Dufour, 1997). In this paper, we show that a necessary condition for the Jacobian identification condition to be satisfied in IVQR models is embedded in the first-stage quantile representation. Hence, the FS-IVQR representation is directly related to the Jacobian requirement of CH-IVQR.

The practical implementation of the FS-IVQR estimator is straightforward and as follows. First, from the IVQR one estimates the conditional density function at a selected quantile which produces an estimate of the weights. The weighting factor is estimated from the IVQR errors using, for instance, sparsity or kernel methods (see, e.g., Koenker, 2005). Second, a standard WLS regression is implemented by regressing the endogenous variable on the instruments and exogenous variables with weights from the first step—this is parallel to the first-stage model used in 2SLS, but using weights. We derive the limiting distribution of the FS-IVQR estimator and show that, under some standard regularity conditions, it is asymptotically normal.

The first-stage regression for conditional average models has been used as a natural framework to evaluate the validity of instruments since one can test for their statistical significance, that is, how IVs impact the endogenous variable(s). A complete parallel testing procedure to evaluate the validity of IV in the QR case only works when one is able to estimate the structural parameters—and consequently the weights—consistently, which in turn requires at least one valid instrument. We thus suggest three practical procedures for testing the first-stage validity of the instruments. First, we consider a simple test for the validity of instruments at a given value of the coefficient of

[1] There is also a more recent literature on GMM QR, see, e.g., Firpo et al. (2022) and references therein.

[2] This has been formally established in Galvao and Montes-Rojas (2015).

the endogenous variable. This can be implemented simply by performing the WLS together with a Wald statistic. This serves conceptually to illustrate the WLS implementation and as a building block for a robust testing procedure discussed below. Second, we consider a test for a subset of IV, assuming that other valid IVs are available. Finally, we propose a robust testing procedure where we evaluate the IV validity for a grid of values of coefficients of the endogenous variable. This procedure does not require to estimate the structural parameters consistently as they are implemented for all potential values of them.

Based on these procedures, we suggest two potential options to the empirical researcher. On the one hand, if there is evidence of a large FS-IVQR statistic for the IV—when evaluated at the CH-IVQR structural parameter estimate of the endogenous variable—then the Chernozhukov and Hansen (2006) inference procedures based on strong IV for the structural parameters could be used.[3] On the other hand, if there is evidence of low values of the FS-IVQR statistic for the IV—at the estimated CH-IVQR estimator—then we suggest using a practical robust inference procedure based on Anderson–Rubin weak identification, as in Chernozhukov and Hansen (2008), Jun (2008), Chernozhukov et al. (2009), and Andrews and Mikusheva (2016).

One important feature of the procedure developed in this paper is that instruments could be statistically insignificant in the first stage of a mean-based 2SLS model, but they could still be related to the endogenous variable in the IVQR setup. The reason is that the 2SLS test only evaluates a mean effect, but the FS-IVQR, because of its specific weighting procedure, allows for different first-stage coefficients across quantiles. As a result, the IV could be relevant at some quantiles, but not for the mean (and vice versa), an issue that has been discussed in Chesher (2003) and subsequent literature. The test developed here thus allows inference on the validity of the IV for the exogeneity condition across quantiles, rather than only a mean effect.

We use a Monte Carlo exercise to evaluate the finite sample performance of the proposed tests. The tests have correct size in all cases studied, where the structural parameters can be consistently estimated under the null hypothesis. We consider alternative cases where there is no identification under the null. The tests have excellent power properties. In particular, these experiments highlight the case where the first-stage 2SLS test for the mean-based model suggests the instrument is not valid, but the proposed FS-IVQR procedure finds it is for some quantiles.

As an empirical illustration, we apply the FS-IVQR estimator to the Card (1995a) data on instrumenting education using college proximity. The analysis shows how to evaluate the FS-IVQR statistical significance of the IV, and reveals that one of the proposed instruments (proximity to 2-year college) should be discarded, but the other (proximity to 4-year college) is valid. However, the latter can only be used for $\tau = 0.25, 0.50$, but it is not valid for $\tau = 0.75$.

The paper is organised as follows. Section 2 briefly reviews the CH-IVQR estimator, rewrites that estimator as a constrained minimisation problem and derives the first-stage representation for the IVQR. Then it shows that the FS-IVQR estimator is related to the identification condition in CH. Section 3 discusses its empirical implementation and derives the estimators' asymptotic distribution. Section 4 presents the first-stage tests for the validity of instruments. Section 5 provides finite sample Monte Carlo evidence. Section 6 applies the proposed tests to an empirical problem. Finally, Section 7 concludes. All proofs are collected in the Appendix.

[3] It is difficult to derive an analogous F-statistic type rule-of-thumb for categorising weak instruments as in, among others, Staiger and Stock (1997), Sanderson and Windmeijer (2016), and Lee et al. (2022); for ordinary least-squares (OLS) models (see Stock and Yogo, 2005, for an extensive discussion).

2. A FIRST-STAGE REPRESENTATION FOR IVQR

2.1. The IVQR estimator and its variants

Let (y, d, x, z) be random variables, where y is a scalar outcome of interest, d is a $1 \times r$ vector of endogenous variables, x is a $1 \times k$ vector of exogenous control variables, and z is a $1 \times p$ vector of exogenous instrumental variables, with $p \geq r$. Define $w = (x, z)$ and $s = (d, x, z)$.

Chernozhukov and Hansen (2006) developed estimation and inference for a generalisation of the QR model with endogenous regressors. A linear representation of the model takes the following form

$$y = d\alpha_0(u_d) + x\beta_0(u_d), \ \ u_d|x, z \sim \text{Uniform}(0, 1), \tag{2.1}$$

where u_d is the nonseparable error or rank and the subscript indicates the endogenous covariates of the model, and α_0 ($r \times 1$ vector) and β_0 ($k \times 1$ vector) are the parameters of interest. Under some regularity conditions, CH establish the following IV identification function

$$\text{P}[y \leq d\alpha_0(\tau) + x\beta_0(\tau)|x, z] = \text{P}[u_d \leq \tau|x, z] = \tau. \tag{2.2}$$

Although each parameter and estimator is indexed by the quantile $\tau \in (0, 1)$, throughout the paper we will suppress the dependence on τ.

The restriction in (2.2) can be used to estimate the parameters of interest. For a given quantile τ, the population IVQR estimator for model in (2.1), is given by

$$\min_{\alpha} \|\gamma(\alpha)\|_A,$$

where

$$(\beta(\alpha), \gamma(\alpha)) = \underset{\beta,\gamma}{\text{argmin}} \ \text{E}\,[\rho_\tau(y - d\alpha - x\beta - z\gamma)],$$

and $\rho_\tau(u) = u(\tau - \mathbf{1}(u < 0))$ is the check function, and $\| \cdot \|_A = \cdot' A \cdot$ is the Euclidean distance for any positively definite matrix A of dimension $p \times p$.

As noted by Chernozhukov and Hansen (2006, p. 501), the IVQR estimator is asymptotically equivalent to a particular generalised method of moments (GMM) estimator where the QR first-order conditions are used as moment conditions. In particular, it would involve a Z-estimator solving

$$\text{E}\left[x'\left(\mathbf{1}[y - d\alpha - x\beta < 0] - \tau\right)\right] = \mathbf{0}_k, \tag{2.3}$$

$$\text{E}\left[z'\left(\mathbf{1}[y - d\alpha - x\beta < 0] - \tau\right)\right] = \mathbf{0}_p, \tag{2.4}$$

where $\mathbf{1}(\cdot)$ is the indicator function. Here $\mathbf{0}_k$ and $\mathbf{0}_p$ are null vectors with dimensions $k \times 1$ and $p \times 1$, respectively.

Different estimators have been proposed in the GMM framework based on identifying the structural parameters from equations (2.3)–(2.4). Kaplan and Sun (2017), Chen and Lee (2018), and De Castro et al. (2019) provide general estimation procedures based on smoothing techniques of the non-differentiable indicator function. However, the constructed estimator differs from the CH-IVQR one. This can be seen in that the term $z\gamma$ is not considered in the moment condition. Our procedure follows the CH estimator and their specific notation.

2.2. The IVQR estimator as a constrained minimisation problem

The IVQR estimator proposed by Chernozhukov and Hansen (2006), for a given quantile τ, can be written as a constrained minimisation problem of (2.5), where the constraints are the moment conditions (2.6) and (2.7), that is,

$$\min_{(\alpha,\beta,\gamma)} \|\gamma\|_A, \tag{2.5}$$

subject to

$$\mathrm{E}\left[x'\left(\mathbf{1}[y - d\alpha - x\beta - z\gamma < 0] - \tau\right)\right] = \mathbf{0}_k, \tag{2.6}$$

$$\mathrm{E}\left[z'\left(\mathbf{1}[y - d\alpha - x\beta - z\gamma < 0] - \tau\right)\right] = \mathbf{0}_p. \tag{2.7}$$

Now we write this constrained optimisation as a Lagrangian problem as

$$\begin{aligned}\mathcal{L}(\alpha, \beta, \gamma, \lambda_x, \lambda_z) = \|\gamma\|_A &+ \lambda_x \mathrm{E}\left[x'(\mathbf{1}[y - d\alpha - x\beta - z\gamma < 0] - \tau)\right] \\ &+ \lambda_z \mathrm{E}\left[z'(\mathbf{1}[y - d\alpha - x\beta - z\gamma < 0] - \tau)\right],\end{aligned} \tag{2.8}$$

where λ_x is a $1 \times k$ vector and λ_z is a $1 \times p$ vector.[4] Therefore, the IVQR estimator is given by the empirical counterpart of

$$\underset{(\theta,\lambda_x,\lambda_z)}{\operatorname{argmin}} \mathcal{L}(\theta, \lambda_x, \lambda_z),$$

where $\theta = (\alpha', \beta', \gamma')'$.

The first derivatives of the Lagrangian in (2.8) are

$$\partial\mathcal{L}/\partial\alpha = -\left\{\lambda_x \mathrm{E}\left[f \cdot x'd\right] + \lambda_z \mathrm{E}\left[f \cdot z'd\right]\right\}' \tag{2.9}$$

$$\partial\mathcal{L}/\partial\beta = -\left\{\lambda_x \mathrm{E}\left[f \cdot x'x\right] + \lambda_z \mathrm{E}\left[f \cdot z'x\right]\right\}' \tag{2.10}$$

$$\partial\mathcal{L}/\partial\gamma = \left\{2\gamma' A - \lambda_x \mathrm{E}\left[f \cdot x'z\right] - \lambda_z \mathrm{E}\left[f \cdot z'z\right]\right\}' \tag{2.11}$$

$$\partial\mathcal{L}/\partial\lambda_x = \mathrm{E}\left[x'(\mathbf{1}[y - d\alpha - x\beta - z\gamma < 0] - \tau)\right]' \tag{2.12}$$

$$\partial\mathcal{L}/\partial\lambda_z = \mathrm{E}\left[z'(\mathbf{1}[y - d\alpha - x\beta - z\gamma < 0] - \tau)\right]', \tag{2.13}$$

where $f := f_{u_\tau}(0|d, x, z)$ denotes the density function of $u_\tau := y - d\alpha_0(\tau) - x\beta_0(\tau)$ conditional on $s = (d, x, z)$, evaluated at the τ-th conditional quantile, which is zero. Note that f is specific for each quantile τ. This density function plays an important role in equations (2.9)–(2.13) and in what follows.

The solution should have all equations above equal to zero when assuming an interior solution as in Assumption 3.1. Thus, from (10),

$$\lambda_x' = -\left(\mathrm{E}[f \cdot x'x]\right)^{-1}\left(\mathrm{E}[f \cdot x'z]\right)\lambda_z'. \tag{2.14}$$

Then, replacing (2.14) in (2.11),

$$\left(\mathrm{E}[f \cdot z'z] - \mathrm{E}[f \cdot z'x](\mathrm{E}[f \cdot x'x])^{-1}\mathrm{E}[f \cdot x'z]\right)\lambda_z' = 2A\gamma,$$

[4] See Pouliot (2019) and Kaido and Wüthrich (2021) for recent contributions that tackle the problem of practical implementation of the IVQR methods.

such that

$$\lambda_z' = 2\left(\mathrm{E}[f \cdot z'z] - \mathrm{E}[f \cdot z'x](\mathrm{E}[f \cdot x'x])^{-1}\mathrm{E}[f \cdot x'z]\right)^{-1} A\gamma. \tag{2.15}$$

Finally, replacing (2.15) in (2.9),

$$\begin{aligned}
&\mathrm{E}\left[f \cdot d'x\right]\lambda_x' + \mathrm{E}\left[f \cdot d'z\right]\lambda_z' \\
&\quad = 2\left\{\mathrm{E}\left[f \cdot d'z\right] - \mathrm{E}\left[f \cdot d'x\right](\mathrm{E}[f \cdot x'x])^{-1}\mathrm{E}[f \cdot x'z]\right\} \\
&\qquad \times \left\{\mathrm{E}[f \cdot z'z] - \mathrm{E}[f \cdot z'x](\mathrm{E}[f \cdot x'x])^{-1}\mathrm{E}[f \cdot x'z]\right\}^{-1} A\gamma = \mathbf{0}_r,
\end{aligned}$$

where $\mathbf{0}_r$ is an $r \times 1$ vector of zeros.

Therefore, we can restate the IVQR problem for $(\alpha', \beta', \gamma')'$ as a system of three equations given by

$$\begin{aligned}
&\left\{\mathrm{E}\left[f \cdot d'z\right] - \mathrm{E}\left[f \cdot d'x\right]\left(\mathrm{E}\left[f \cdot x'x\right]\right)^{-1}\mathrm{E}\left[f \cdot x'z\right]\right\} \\
&\times \left\{\mathrm{E}\left[f \cdot z'z\right] - \mathrm{E}\left[f \cdot z'x\right]\left(\mathrm{E}\left[f \cdot x'x\right]\right)^{-1}\mathrm{E}\left[f \cdot x'z\right]\right\}^{-1} A\gamma = \mathbf{0}_r
\end{aligned} \tag{2.16}$$

$$\mathrm{E}\left[x \cdot (\mathbf{1}[y - d\alpha - x\beta - z\gamma < 0] - \tau)\right] = \mathbf{0}_k \tag{2.17}$$

$$\mathrm{E}\left[z \cdot (\mathbf{1}[y - d\alpha - x\beta - z\gamma < 0] - \tau)\right] = \mathbf{0}_p. \tag{2.18}$$

2.3. *First-stage IVQR parameters*

Given (2.16)–(2.18), we can see that (2.16) provides a first-stage representation of the IVQR model. This can be written as

$$\delta' A\gamma = \mathbf{0}_r, \tag{2.19}$$

where

$$\begin{aligned}
\delta := &\left\{\mathrm{E}\left[f \cdot z'z\right] - \mathrm{E}\left[f \cdot z'x\right]\left(\mathrm{E}\left[f \cdot x'x\right]\right)^{-1}\mathrm{E}\left[f \cdot x'z\right]\right\}^{-1} \\
&\times \left\{\mathrm{E}\left[f \cdot z'd\right] - \mathrm{E}\left[f \cdot z'x\right]\left(\mathrm{E}\left[f \cdot x'x\right]\right)^{-1}\mathrm{E}\left[f \cdot x'd\right]\right\}.
\end{aligned} \tag{2.20}$$

Here δ is a $p \times r$ matrix. Notice that (2.20) is a least-squares projection coefficient. In particular, the representation in (2.20) is a weighted projection, where the endogenous variable(s), d, is(are) regressed on the IV, z, and the exogenous variables, x. This is the analogue to the first stage in the 2SLS case, with the difference that the QR case is a weighted regression. The weights are given by the conditional density function of the innovation term in the QR structural model, conditional on the endogenous and exogenous covariates together with the instruments.

Hence, for each endogenous variable, say d_j for $j = 1, 2, ..., r$, δ_j in (2.20), can be recovered as the solution to the following optimisation problem

$$\mu_j := (\psi_j, \delta_j) = \underset{\psi,\delta}{\operatorname{argmin}}\ \mathrm{E}\left[f \cdot (d_j - x\psi - z\delta)^2\right]. \tag{2.21}$$

Note that the parameter δ also depends on $\theta = (\alpha', \beta', \gamma')'$, through the conditional density function f at quantile τ. Thus, this first-stage representation depends on the structural (second-stage) parameters and, as such, it is different from the 2SLS case in mean regression models.

We notice that the first stage in (2.21) is different from those in the existing literature using two-stage regressions for conditional quantile models. Amemiya (1982), Powell (1983), Chen and Portnoy (1996), and Kim and Muller (2004) propose different two-step procedures in which the first step fits the endogenous variable(s) as a function of exogenous covariates and IV, and this is then plugged in a second-stage. Nevertheless, these papers use least squares without weighting or standard quantile regression in the first stage. Our procedure derives the first stage from the IVQR setup, thus confirming that a first-stage (albeit different) is part of the model.

2.4. *Further intuition on the FS-IVQR*

Now we further discuss the first stage derived in the previous section, in particular (2.20). For simplification, we consider a model without additional exogenous covariates x and one endogenous variable d, i.e., $r = 1$. In this case (2.20) simplifies substantially and can be written as

$$\delta = \mathrm{E}\left[f \cdot z'z\right]^{-1} \mathrm{E}\left[f \cdot z'd\right],$$

which is the solution of a simplified version of the optimisation problem in (2.21) as

$$\delta = \underset{\delta}{\operatorname{argmin}}\, \mathrm{E}\left[f \cdot (d - z\delta)^2\right].$$

First, we relate the proposed estimator to the rank identification conditions for the IVQR estimator of CH. As discussed in CH, the IVQR optimisation problem is asymptotically equivalent to solving the following moment condition

$$\Pi((\alpha, \gamma), \tau) = \mathrm{E}\left[z'(\mathbf{1}[y - d\alpha - z\gamma < 0] - \tau)\right] = \mathbf{0}_p.$$

To establish the asymptotic properties of the IVQR estimator, it is required that the Jacobian matrices, $\frac{\partial}{\partial \alpha}\Pi((\alpha, \gamma), \tau)$ and $\frac{\partial}{\partial \gamma}\Pi((\alpha, \gamma), \tau)$, are continuous and full column rank (see below the conditions for the derivation of the asymptotic properties of the estimator, in particular, Assumption 1, item R3). We show here that these conditions are embedded in the FS-IVQR representation.

The rank Jacobian conditions are

$$\operatorname{rank}\left(\frac{\partial \Pi((\alpha, \gamma), \tau)}{\partial \alpha}\right) = \operatorname{rank}\left(\mathrm{E}[f \cdot z'd]\right) \geq r,$$

$$\operatorname{rank}\left(\frac{\partial \Pi((\alpha, \gamma), \tau)}{\partial \gamma}\right) = \operatorname{rank}\left(\mathrm{E}[f \cdot z'z]\right) = p.$$

The first equation implies that for the case of one endogenous variable, $r = 1$, $\mathrm{E}[f \cdot z'd]$ has at least one nonzero column, and the second equation requires p noncollinear valid instruments. Now notice that from the FS-IVQR representation given by (2.20), in the case without exogenous regressors, simplifies to:

$$\mathrm{E}[f \cdot z'd] = \mathrm{E}[f \cdot z'z]\delta.$$

Therefore, the matrices involved in the rank conditions directly appear in representation (2.20). Also, if the FS-IVQR parameter $\delta = 0$, then the rank conditions cannot be satisfied. Note that this is a necessary condition, but not a sufficient one. Furthermore, by checking how close δ is to zero one is in fact evaluating the strength of the identification condition.

Second, the restriction in (2.19) provides a natural framework to evaluate the relevance of the instruments in IVQR models. The first-stage regression representation in (2.21) is a weighted linear projection, where the weights are the conditional density function of the innovation term in the QR structural model, conditional on the endogenous and exogenous covariates together with the instruments. This exposes a caveat of the QR IV model. In order to estimate the parameters in (2.21) consistently, one needs a consistent estimate of the density f, and hence at least one valid instrument must be available to the researcher. This is in contrast with the standard conditional average models where the first stage is a simple OLS regression without weights. The required weights in the QR case will be further discussed below when we suggest a test for the validity of the IV.

Third, notice that the parameter δ captures the strength of the instrument in the sense it measures the correlation between the instrument z and the endogenous variable d weighted by the density function f. This is the QR counterpart of the first-stage partial correlation of z on the endogenous variables d for the 2SLS. As noted by Galvao and Montes-Rojas (2015) the CH setup is equivalent to the 2SLS in least-squares models. In fact the CH estimator is the QR counterpart of a 2SLS estimator. The expression above also shows that there is an implicit first stage, similar to that in 2SLS problems. As such, this provides an analytical expression to evaluate the relevance of the IV. When the instrument is valid, $\delta \neq \mathbf{0}_{p\times r}$.

Fourth, note that the instrument z does not belong in the structural quantile model (2.1), hence $\gamma = \mathbf{0}_{p\times r}$ can be used for identification, a key feature of the CH-IVQR estimator. Equation (2.19) also shows that when $\delta = \mathbf{0}_{p\times r}$, the value of γ is irrelevant and, therefore, it cannot be used in the IVQR procedure to solve endogeneity. As such, $\delta \neq \mathbf{0}_{p\times r}$ is a necessary condition for the IV to have a purpose in the CH setup. Therefore, a test for the validity of the instruments can be based on a test for statistical significance of δ.

Finally, another way of gaining intuition on the test is the following. Note that for the case of one endogenous variable, $r = 1$, (2.19) is in fact equal to 0, a scalar. If we further assume that $A = I_p$, then

$$\sum_{q=1}^{p} \delta_q \gamma_q = 0, \tag{2.22}$$

where $\delta = [\delta_1, \ldots, \delta_p]'$ is the column vector that has the first-stage effect of all IV on d. Note again that if $\delta = \mathbf{0}_{p\times 1}$, then the vector γ could have any value and its implied restrictions would be irrelevant.

3. EMPIRICAL IMPLEMENTATION AND ASYMPTOTIC DISTRIBUTION

In this section we propose a two-step estimator for the first-stage instrumental variables quantile regression (FS-IVQR), consider its empirical implementation, and derive the estimators' asymptotic distribution. The two steps estimation procedure consists of estimating the conditional density using the IVQR model in the first step, and in the second step employing a weighted least squares (WLS) regression. For simplicity of exposition, we present the case of $r = 1$, i.e., one endogenous variable, but as discussed above the case of $r > 1$ can be implemented using separate regressions.

3.1. FS-IVQR estimator

The FS-IVQR estimator requires a consistent estimator of μ in (2.21), which will be based on WLS based on the estimator of f, at a given quantile of interest τ. The CH estimator $\hat{\theta} = (\hat{\alpha}, \hat{\beta}', \hat{\gamma}')'$ can be used to compute such $\hat{f}$.

The estimator has two steps as following:

(a) In the first step we obtain the structural parameters' estimates $\hat{\theta} = (\hat{\alpha}, \hat{\beta}', \hat{\gamma}')'$ using the Chernozhukov and Hansen (2006) IVQR estimator, that is[5]

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \|\hat{\gamma}(\alpha)\|_A,$$

where

$$(\hat{\beta}(\alpha), \hat{\gamma}(\alpha)) = \underset{\beta,\gamma}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} [\rho_\tau(y_i - d_i\alpha - x_i\beta - z_i\gamma)].$$

(b) In the second step the parameters of interest δ can be obtained from a feasible WLS.[6] The QR literature provides different alternatives for the specific estimating f. One could use a kernel estimator for the conditional density, as in Powell (1991). This procedure would use the error term $\hat{u}_\tau := y - d\hat{\alpha}(\tau) - x\hat{\beta}(\tau) - z\hat{\gamma}(\tau)$ from the CH-IVQR estimator. Another alternative is to use sparsity estimation methods, as suggested by Hendricks and Koenker (1992). This estimator is discussed in further details in Zhou and Portnoy (1996), Koenker (2005), and Ota et al. (2019). We discuss practical implementation of the density estimation in Section 5. Then, the WLS is

$$\hat{\mu} := (\hat{\psi}, \hat{\delta}) = \underset{\psi,\delta}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \left[\hat{f}_i \cdot (d_i - x_i\psi - z_i\delta)^2\right]. \tag{3.1}$$

Equation (3.1) produces $\hat{\delta}$, which is the main object of interest.

Define Y, X, D, and Z as the matrices formed from a random sample of $\{y_i, d_i, x_i, z_i\}_{i=1}^n$. Similarly, define $W = [X, Z]$. Define the weighting diagonal matrix

$$\hat{V} = \begin{bmatrix} \hat{f}_1 & & \\ & \ddots & \\ & & \hat{f}_n \end{bmatrix}.$$

Then, the estimator in (3.1) can be written in a simple matrix notation as

$$\hat{\mu} = (W'\hat{V}W)^{-1}W'\hat{V}D. \tag{3.2}$$

Notice that if f_i is a constant for all i, then the proposed FS-IVQR method should deliver same estimates as FS-2SLS for the mean. This would happen, for example, in the case of independent and identically distributed (i.i.d.) innovations in the second-stage structural model. Thus, there will be differences between the two estimators only when f_i varies across i, that is, when the weighting factor is not a constant.

[5] We refer the reader to Chernozhukov and Hansen (2006) for the discussion of this estimator.

[6] We are assuming that there is only one endogenous variable, $r = 1$. Otherwise the analysis below should be repeated separately for each endogenous variable as there will be a different first stage for each one.

3.2. Asymptotic distribution

In this subsection, we derive the asymptotic distribution of the proposed estimator. The asymptotic properties of the IVQR estimator can be found in Chernozhukov and Hansen (2006), and the assumptions therein are those required for inference. We consider Assumption 2 in Chernozhukov and Hansen (2006, pp. 501–2), which we reproduce here for convenience. It imposes conditions for θ_0 to be identified and estimated.

ASSUMPTION 3.1. *R1. Sampling.* $\{y_i, x_i, d_i, z_i\}$ *are i.i.d. defined on a probability space and take values in a compact set.*

R2. Compactness and convexity. For all $\tau \in (0, 1)$, $(\alpha, \beta, \gamma) \in \text{int}(\mathcal{A} \times \mathcal{B} \times \mathcal{G})$ *where* $(\mathcal{A} \times \mathcal{B} \times \mathcal{G})$ *is compact and convex.*

R3. Full rank and continuity. y *has bounded conditional density (conditional on* $s = (d, w, z)$*), and for* $\theta = (\alpha, \beta, \gamma)$,

$$\Pi(\theta, \tau) := \mathrm{E}\,[(\tau - \mathbf{1}(y < d\alpha + x\beta + z\gamma)) \cdot [x, z]],$$

Jacobian matrices $\frac{\partial}{\partial(\alpha', \beta')}\Pi(\theta, \tau)$ *and* $\frac{\partial}{\partial(\beta', \gamma')}\Pi(\theta, \tau)$ *are continuous and have full rank, uniformly over* $\mathcal{A} \times \mathcal{B} \times \mathcal{G}$ *and the image of* $\mathcal{A} \times \mathcal{B} \times \mathcal{G}$ *under the mapping* $(\alpha, \beta) \mapsto \Pi(\theta, \tau)$ *is simply connected. Assume that* $\theta_0 = (\alpha_0, \beta_0', \gamma_0')'$ *is the unique solution to the CH problem.*

We impose additional conditions for deriving the limiting properties of the feasible first-stage estimator in (3.1).

ASSUMPTION 3.2. *(i) Assume that* $\mathrm{E}[d_i w_i w_i' d_i] < \infty$. *Let* $\Omega_{f\sigma} := Var(f_i d_i w_i) < \infty$, *where* $w_i = [x_i, z_i]$ *and* $\Omega_f := \mathrm{E}[f_i w_i w_i'] < \infty$ *be nonsingular. (ii) Assume that the density estimator* $\hat{f}_i$ *satisfies the following expansion* $\hat{f}_i - f_i = m_n^{-1} \sum_{j \neq i}^{n} \Psi_i(u_{j\tau}) + o_p(m_n^{-1})$, *where* $u_{j\tau} := y_j - d_j\alpha_0(\tau) - x_j\beta_0(\tau)$, $\{\Psi_i\}$ *is bounded, the contribution of* $\{\Psi_i(u_i)\}$ *is negligible, and* $m_n^{-1} = o(n^{-1/2})$ *and* $nm_n^{-2} = o(n^{-1/2})$.

Assumption 3.2 contains conditions for establishing consistency and asymptotic normality of the proposed estimator. Part (i) simply imposes standard assumptions on moment conditions. Part (ii) provides a condition on the density estimation together with rates of convergence, that are satisfied for conditional density estimators commonly used in QR methods. This condition is the same as in Koenker (2005, pp. 161–3), where it is used for estimated weights in the context of QR models, where for a parametric model $m_n = n$, and for a nonparametric kernel estimator $m_n = n^{4/5}$. The next lemma presents the result. Estimation of weights, f_i, has appeared in the QR literature, and it is a difficult question to resolve in its full generality. Koenker and Zhao (1994) treat the linear location-scale model, providing a uniform Bahadur representation for the empirically weighted QR process. Some related results, in the context of autoregressive conditional heteroscedasticity-type models, are provided in Koenker and Zhao (1996). Zhao (2001) considers a model in which the weights are estimated by nearest-neighbour nonparametric methods.

LEMMA 3.1. *Under Assumptions 3.1–3.2, as* $n \to \infty$,

$$\sqrt{n}\,(\hat{\mu} - \mu_0) \xrightarrow{d} N\left(\mathbf{0}_{k+p}, V(\mu_0)\right),$$

where $\mu_0 := \mathrm{E}[f_i w_i w_i']^{-1}\mathrm{E}[f_i w_i d_i]$, *and* $V(\mu_0) := \Omega_f^{-1}\Omega_{f\sigma}\Omega_f^{-1}$ *is the asymptotic covariance matrix.*

4. TESTS FOR VALIDITY OF THE IV

In this section we suggest tests for the validity of the IV using the first-stage representation. The formulation of the test proposed in this paper is based on the condition given in (2.16) together with the first stage IVQR representation in (2.20). A test for validity of the instruments for p instruments can be based on the null hypothesis

$$H_0 : \delta_0 = \mathbf{0}_{p\times r}, \tag{4.1}$$

against the alternative

$$H_A : \delta_0 \neq \mathbf{0}_{p\times r}. \tag{4.2}$$

We highlight that, differently from the 2SLS, the first-stage IVQR in (2.21) is for a given quantile τ. Thus, for the same variables d and instruments z, the strength of the instruments may vary across different quantiles. This difference is captured by the weights f, which are absent in the standard 2SLS model.

Note that the procedure works for $r \geq 1$, that is for one or more than one endogenous variables. In the $r > 1$ case, separate tests could be applied as in 2SLS analysis where there may be a different first stage for each endogenous variable. To simplify the procedures below we assume that $r = 1$, that is, there is only one endogenous variable.

A natural choice to test H_0 against H_1, in (4.1) and (4.2), respectively, together with the result in Lemma 3.1, for the case of $r = 1$ is the Wald statistic as

$$T_n = n\hat{\delta}'\{V_\delta\}^{-1}\hat{\delta}, \tag{4.3}$$

where V_δ is the asymptotic covariance matrix of $\sqrt{n}\hat{\delta}$ under H_0. In practice, V_δ is replaced by a suitable consistent estimate. We will discuss the practical implementation as well the limiting distribution in the next section. When H_0 is true, under suitable regularity conditions, $\hat{\delta}$ converges in probability to $\mathbf{0}_{p\times r}$ for a given τ. However, when H_A is true, $\hat{\delta}$ converges in probability to $\delta_0 \neq \mathbf{0}_{p\times r}$. Therefore, it is reasonable to reject H_0 if the magnitude of $\hat{\delta}$ is suitably large.

The main issue at stake is that the above test implementation requires to use the correct weights, f, or a consistent estimate. This may be unfeasible if one is not able to estimate the parameters in the structural equation, i.e., the coefficient α. Thus, we propose three different methods to empirically evaluate the validity of the instruments.

First, we consider a simple test for the strength of instruments at a given value of the coefficient α. This can be implemented simply by performing the WLS together with the Wald statistic described above. While this may not be of interest for an applied researcher (as α is unknown in practice) it serves conceptually to illustrate the WLS implementation, and as a building block for a robust testing procedure discussed below.

Second, we consider individual tests for a subset of IV, assuming that others are available and they are valid.

Finally, we propose a robust inference procedure where we evaluate the IV validity for a grid of α values. This procedure does not require to estimate the structural parameters consistently, as they are implemented for all potential values of them.

4.1. Testing for validity of the instruments

We first consider a simple test for the validity of instruments as in (4.1) at a given value of the coefficient α, such that this coefficient is known. In this case, one is able to use a standard QR to estimate (β, γ) and construct consistent estimates of the weights f.

This test is very simple to implement in practice. Recall that $\mu = (\psi', \delta')'$ are the parameters of the FS-IVQR, and $\theta = (\alpha, \beta', \gamma')$ are the parameters of the structural QR. As mentioned above, when α is known, one can use a QR to estimate $(\beta', \gamma')'$ and consistently recover $\hat{f}$. Then, the WLS can be performed to compute $\hat{\mu}$. Let $\hat{V}(\hat{\mu})$ be a consistent estimator of $V(\mu_0)$, which can be obtained from the WLS procedure. The next result provides the limiting distribution of the Wald test statistic in (4.3) for a subset of instruments.

PROPOSITION 4.1. *Consider Assumptions 3.1–3.2, $n \to \infty$. Then, under $H_0 : \delta = \mathbf{0}_p$ and local alternatives $H_A : \delta = \boldsymbol{a}_p/\sqrt{n}$,*

$$T_n = n\hat{\mu}' \left(\hat{V}(\hat{\mu})\right)^{-1} \hat{\mu} \xrightarrow{d} \chi^2_p(\boldsymbol{a}_p).$$

4.2. Testing for a subset of the instruments

The previous section considered the case of a given value of α. However, in most applications the researcher might not have such value available. In this section, we relax this constraint. Nevertheless, given that the procedure requires a consistent estimate of the weights, we impose a restriction on the number of instruments available.

Consider a subset of the instruments, $p_1 < p$, and consider a partition of $\delta = [\delta_1', \delta_2']'$ of the corresponding first-stage parameters of interest, with dimensions p_1 and p_2 (with $p = p_1 + p_2$), respectively. Consider a $p_1 \times (k + p)$ matrix $R = [\mathbf{0}_{p_1 \times k}, \boldsymbol{I}_{p_1}, \mathbf{0}_{p_1 \times p_2}]$ where $\boldsymbol{I}_{p_1}$ is an identity matrix of dimension $p_1 \times p_1$. Thus, $R\mu = \delta_1$ is the subvector of interest. Let $\hat{V}(\hat{\mu})$ be a consistent estimator of $V(\mu_0)$, which can be obtained from the WLS procedure. The next result derives the limiting distribution of the test statistic in (4.3) for a subset of instruments.

PROPOSITION 4.2. *Consider Assumptions 3.1–3.2. Furthermore, assume that $dim(z) = p > p_1 \geq 1$. Then, under $\delta_2 \neq \mathbf{0}_{p_2}$ and $H_0 : \delta_1 = \mathbf{0}_{p_1}$ and local alternatives $H_A : \delta_1 = \boldsymbol{a}_{p_1}/\sqrt{n}$,*

$$T_n = n\,(R\hat{\mu})'\,\{R\hat{V}(\hat{\mu})R'\}^{-1}\,(R\hat{\mu}) \xrightarrow{d} \chi^2_{p_1}(\boldsymbol{a}_{p_1}).$$

Computation of the test statistic in (4.3) requires a nonparametric estimator of f, the conditional density of $u_\tau|d, x, z$ evaluated at the specific quantile of interest τ. Given that the weights need to be estimated, the proposed FS-IVQR has specific properties when testing under the null hypothesis of an invalid instrument. The condition on the number of IV being larger than the number of parameters tested in the null hypothesis is required for consistent estimation of θ under the null, which in turn, is used for the consistent estimation of f.

4.3. Robust inference for the first stage

The previous subsections presented frameworks of tests for the statistical significance of the FS-IVQR parameters. One important limitation is that the estimator depends crucially on a given value or estimation of α, the coefficient of the endogenous variable in the structural equation, as this is a requirement for computing the weights f. In turn, these may not be available nor valid if one works under the null of all IV being invalid, i.e., $\delta = 0$ for the p instruments. Here we suggest another alternative.

Following Chernozhukov and Hansen (2008; see also Chernozhukov et al. 2020, sect. 3.3) we propose here a robust inference analysis that does not have this restriction. Those authors present an inference model based on the Anderson–Rubin framework, valid for weak instruments.[7] We apply this idea to our case. Define $T_n(\alpha)$ as the implied Wald statistic arising from (4.3), but which is obtained for a given fixed α. The corresponding $T_n(\alpha)$ statistic checks for the statistical significance of the IV (as in a standard 2SLS framework) for all possible values α (in practice for a grid in a compact set). It is thus a robust-type analysis to check if instruments are strong (in a 2SLS sense).

$T_n(\alpha)$ is computed in two steps. First, by running a QR regression of $y - d\alpha$ on (x, z) and computing the weights as in (5.3). Second, by running WLS as described in (3.2). Note that using the assumptions on the bandwidth as per Lemma 3.1, the estimation of α has no effect on the WLS distribution used in the FS-IVQR. Thus, we can use the same inference procedures described above.

This framework thus proposes a way to inform the econometrician about the first-stage validity of the instruments, and it does not suffer from the lack of identification of the structural parameters that may arise if the IVs are not valid.

As an initial exploratory analysis, consider two potential cases.

(i) Consider first the case of $T_n(\alpha)$ being a *large* value for all α. The definition of large can be based on the consensus on the 2SLS literature on weak IV, and we left it for the reader to decide. A common procedure would be to use F-test values for the first stage to decide. For this case, the IV are strong for whatever value α were the true one. This suggests that 'strong instrument analysis' can be carried out, e.g., IVQR Chernozhukov and Hansen (2006) inference procedures.

(ii) Consider now the case where the $T_n(\alpha)$ were *small* for a large set of values α. This suggests that the IVs are weak, and it suggests to implement weak IV inference procedures such as robust inference in Chernozhukov and Hansen (2008).

In turn, the decision about whether 'strong' standard IVQR inference or 'weak' IV robust inference should be used, can be based on the particular estimator, $\hat{\alpha}$, i.e., the CH-IVQR estimator, that is, on $T_n(\hat{\alpha}) \equiv T_n$ as in (4.3). Note that this analysis may not provide a univocal prescription for empirical practitioners in some cases. Although beyond the scope of this paper, inference procedures should be adjusted for first-stage model selection, as suggested, for instance, in Pötscher (1991) and Leeb and Pötscher (2005, 2006).

5. MONTE CARLO EXPERIMENTS

We analyse in this section the performance of the proposed test with finite samples through a series of Monte Carlo simulation exercises. The data generating process (DGP) has the following location-scale model:

$$y_i = d_i + x_i + (1 + d_i)u_i, \tag{5.1}$$

$$d_i = c_1 + az_{1i} + \phi z_{2i} + (1 + bz_{1i})v_i, \tag{5.2}$$

[7] The main intuition in their paper is that they construct confidence intervals for α based on a Wald test for the condition $\gamma = 0$, which is the vector of coefficients of the IV in the CH-IVQR estimator in (2.5). The key element there is that they do not require to estimate α, as the confidence interval is implicitly calculated.

where x_i, z_{1i} and z_{2i} are three independent variables with distribution $U(0, 1)$; u_i and v_i have standard bivariate normal distribution with correlation 0.50. For all cases we set $c_1 = 10$. Equations (5.1)–(5.2) specify a model where there could be pure location or location-scale specifications in the first stage, thus allowing the instruments to have different effects on the endogenous variable. Note that the parameters a and b determine the type of effect that the instrument z_1 has on the endogenous covariate d. For example, if $a \neq 0$ and $b = 0$ the instrument z_1 has a pure location effect on d (pure location-shift model), while if $a = 0$ and $b \neq 0$ the effect is only on the variance of the endogenous covariate (pure scale-shift model).

We implement the density estimator using the sparsity function model with the difference quotient

$$\hat{f}_i = \frac{2h_n}{s_i\left(\hat{\theta}(\tau + h_n) - \hat{\theta}(\tau - h_n)\right)}. \tag{5.3}$$

We use this estimation in lieu of a kernel estimator for simplicity.[8] The estimator in (5.3) is a natural extension of sparsity estimation methods suggested by Hendricks and Koenker (1992). The bandwidth for the density estimation is chosen as a scaled version of Hall and Sheather (1988):

$$h_n = 2n^{-1/3}\Phi^{-1}(0.975)^{2/3}\left[\frac{3}{2}\cdot\frac{\phi\left\{\Phi^{-1}(\tau)\right\}^4}{2\Phi^{-1}(\tau)^2 + 1}\right]^{1/3}.$$

5.1. First-stage parameter

We consider here tests for $H_0 : \delta_1 = 0$ where this is the first-stage parameter associated with the z_1 instrument defined in the previous sections. We consider two different cases to investigate the numerical properties of the tests. In the first case, $\phi = 1$, there is a second instrument, z_2, such that the model correctly identifies the parameters in the structural (5.1) for all possible values of a and b, even under the case that $a = b = 0$. In the second case, we set $\phi = 0$ and, therefore, under the null hypothesis the consistent estimation of the weights f is problematic. Also, in this case, when $a = b = 0$, there is no valid available instrument.

We will consider three different test statistics from different estimators. First, for comparison purposes, we present a Wald test for the coefficient in z_1 using a simple regression model of d on (x, z_1, z_2) in a standard 2SLS framework, denoted FS-2SLS. Second, we test for $H_0 : \delta_1 = 0$ using the true density function, f, as weights, that is, using the true θ_0, denoted FS-IVQR (true density). We note that this is not observed in practice, and we include these results for comparison purposes. Our proposed test studied in the previous section is the third one, denoted FS-IVQR (sparsity), where we use the sparsity function estimation described above. Note that the three tests differ only in the weighting procedure used in the regression of d on (x, z_1, z_2) or (x, z_1).

Tables 1–2 show the empirical size (i.e., $a = b = 0$) of the computed test with 1,000 simulations for $n = \{500, 1{,}000\}$ and for the quantiles $\tau = \{0.25, 0.50, 0.75\}$.

Consider first the case where there is a second instrument, $\phi = 1$ in Table 1. The tests have approximately correct empirical size in all cases. As such, they clearly evaluate if the instrument z_1 exerts an effect on the endogenous variable d. In all cases they have a similar performance to the FS-2SLS case.

[8] We also implemented the methods using a kernel estimator for the error, $\hat{u}_\tau = y - d\hat{\alpha}(\tau) - x\hat{\beta}(\tau) - z\hat{\gamma}(\tau)$. Since the results are qualitatively similar we omit them for brevity.

Table 1. Rejection rate of the null hypothesis using $a = b = 0$ and $\phi = 1$.

		$n = 500$			$n = 1{,}000$		
τ	*Size*	FS-2SLS	True f	Sparsity f	FS-2SLS	True f	Sparsity f
0.25	**0.10**	0.104	0.109	0.100	0.073	0.078	0.078
	0.05	0.054	0.050	0.054	0.031	0.034	0.042
	0.01	0.015	0.017	0.017	0.006	0.007	0.006
0.50	**0.10**	0.104	0.109	0.101	0.073	0.078	0.076
	0.05	0.054	0.050	0.061	0.031	0.034	0.037
	0.01	0.015	0.017	0.013	0.006	0.007	0.007
0.75	**0.10**	0.104	0.109	0.107	0.073	0.078	0.083
	0.05	0.054	0.050	0.056	0.031	0.034	0.004
	0.01	0.015	0.017	0.016	0.006	0.007	0.008

Note: Empirical rejection rates of 1,000 Monte Carlo experiments.

Table 2. Rejection rate of the null hypothesis using $a = b = 0$ and $\phi = 0$.

		$n = 500$			$n = 1{,}000$		
τ	*Size*	FS-2SLS	True f	Sparsity f	FS-2SLS	True f	Sparsity f
0.25	**0.10**	0.117	0.125	0.113	0.0114	0.106	0.117
	0.05	0.063	0.066	0.070	0.054	0.053	0.063
	0.01	0.011	0.009	0.019	0.010	0.010	0.014
0.50	**0.10**	0.117	0.125	0.117	0.114	0.106	0.112
	0.05	0.063	0.066	0.061	0.054	0.053	0.057
	0.01	0.011	0.009	0.016	0.010	0.010	0.013
0.75	**0.10**	0.117	0.125	0.113	0.114	0.106	0.114
	0.05	0.063	0.066	0.063	0.054	0.053	0.053
	0.01	0.011	0.009	0.012	0.010	0.010	0.010

Note: Empirical rejection rates of 1,000 Monte Carlo experiments.

Now consider the case where there is no available second instrument, $\phi = 0$ in Table 2. The idea of this experiment is to evaluate the test performance when there is lack of identification under the null. In this case, the weights in the structural model cannot be estimated consistently under the null. Since the proposed test evaluates the relationship between z_1 and d, the main issue is whether this relationship can be evaluated in other than the OLS model. The simulations show that the size is correct for the sparsity estimator. This result suggests that the test can be used even when the structural parameters cannot be estimated under the null (because z_1 does not solve the endogeneity problem).

To analyse the empirical power of the tests, we study the cases where in (i) we evaluate a pure location first-stage model of z_1 on d using $a = \{0, 0.10, ..., 0.90, 1\}$ and $b = 0$, and in (ii) we set $a = 0$ and we vary $b = \{0, 0.10, ..., 0.90, 1\}$, and we perform 100 simulations for each case. The only sample size considered is $n = 1{,}000$, and we calculate the rejection rates of the proposed procedure for the quantiles $\tau = \{0.25, 0.50, 0.75\}$. As benchmark we also use the test rejection rates obtained in the FS-2SLS method, i.e., the Wald test of an OLS regression of d on z_1. The results appear in Figures 1 and 2. For each figure we have two blocks, (i) and (ii), that correspond to either varying a or b, respectively.

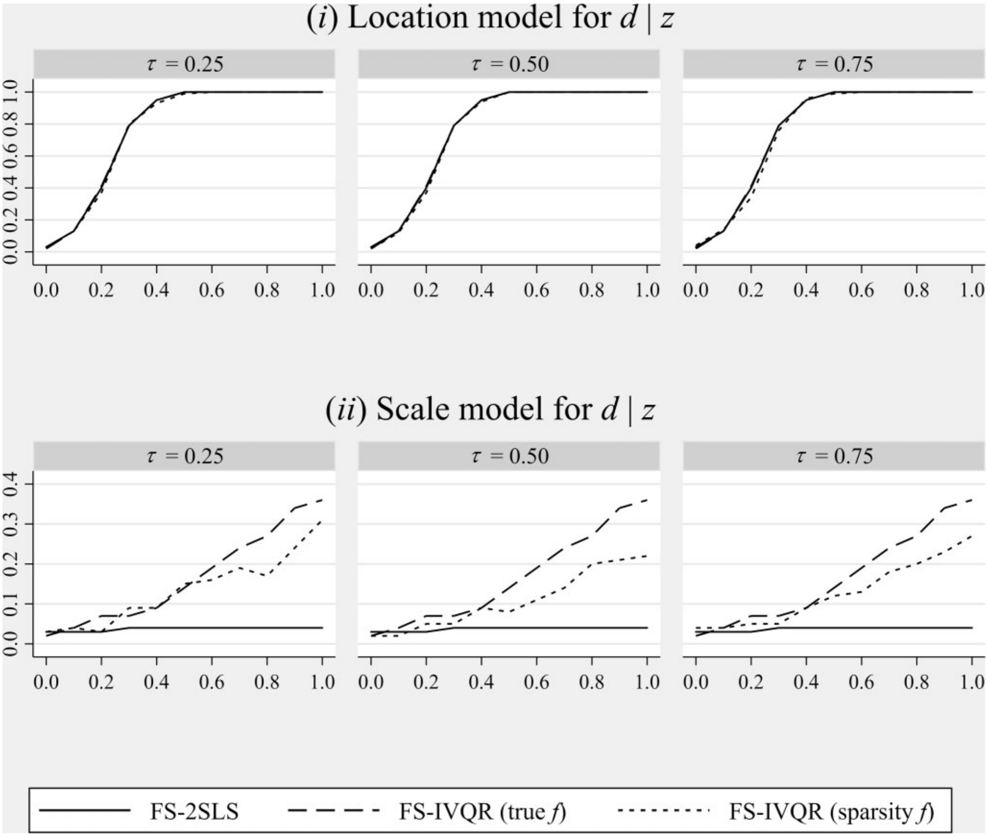

Figure 1. Power for $H_0 : \delta_1 = 0$ (model with $\phi = 1$).

We first consider the case where there is a second valid instrument $\phi = 1$. Figure 1, block (i) pure location first stage, shows that the FS-IVQR power computed with true and estimated densities behaves similarly to FS-2SLS. That is, they correctly reject as a increases. The estimated density model has slightly less power than the one with the true density. For block (ii), the results of the FS-IVQR differ when we are in the presence of a pure-scale model for $d|z_1$. Note that in this case there is no relationship between d and z_1 at the mean (FS-2SLS), but it does affect the other points of the conditional distribution. Therefore, the first stage of 2SLS does not find any relationship between the endogenous variable and the instrument, while the FS-IVQR estimators (both true and estimated weights) are able to correctly detect it.

Finally, consider the last case when $\phi = 0$ in Figure 2. The FS-IVQR tests also work in this case. In both (i) and (ii) cases, the tests detect an association between the instrument and the endogenous variable. In case (ii) the FS-IVQR rejects as b increases while FS-2SLS does not. As noted in Table 2 the test works even for the case where $a = b = 0$ and the endogeneity problem in the structural estimators cannot be solved.

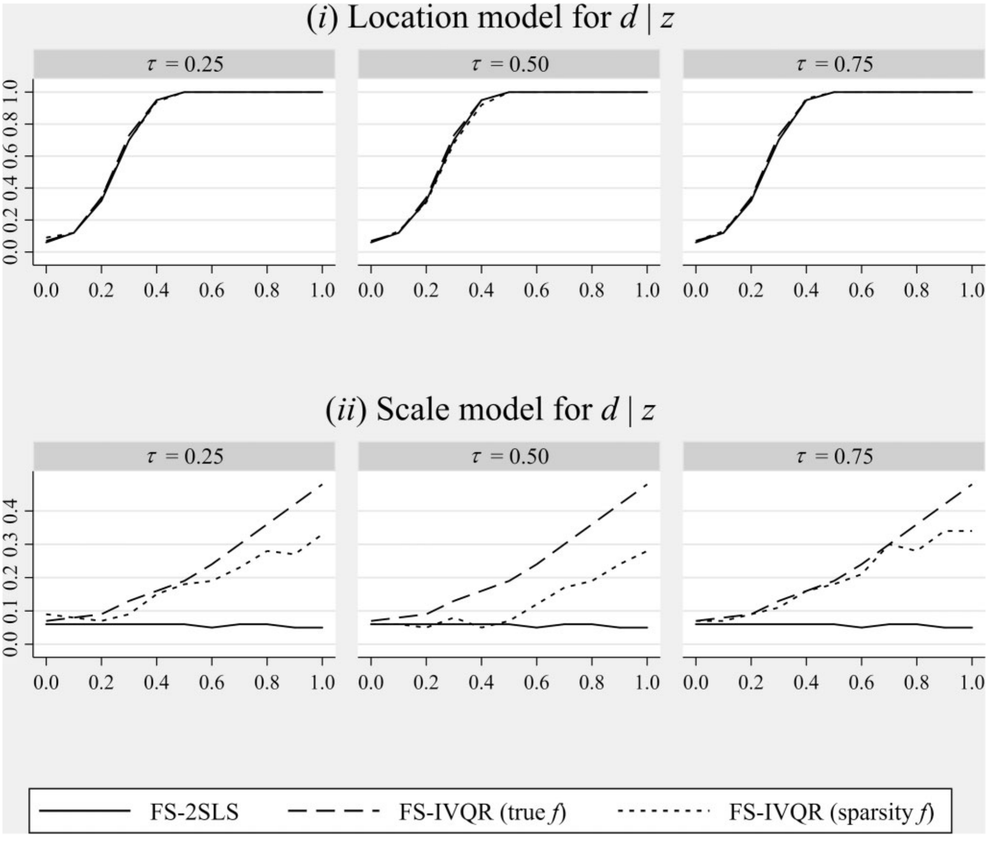

Figure 2. Power for $H_0 : \delta_1 = 0$ (model with $\phi = 0$).

5.2. Robust inference

We consider here the robust inference procedure discussed above. This is the case where we test for the first-stage statistical significance of the IV, but the structural parameters cannot be identified under the null. As discussed above one valid strategy is to implement the test for $H_0 : \delta(\alpha) = 0$ based on $T_n(\alpha)$ for a grid of values of α.

We consider the DGP in 5.2 using $\phi = 0$, i.e., there is only one IV (z_1), and as such the coefficient of the endogenous variable α requires a valid IV. We fix $n = 1{,}000$ and 500 simulations for each DGP. We evaluate $\tau \in \{0.25, 0.50, 0.75\}$ quantiles as above. We report here the simulations for the location-shift model of $d|z$, i.e., $b = 0$, but we consider different values of a. In the first case we use $a = 0$ such that the instrument is not valid. In a second case we use $a = 1$ and then the instrument is valid.

First, we consider the case where $a = 0$ and thus z_1 is not related to d. Note that for this case we cannot correctly identify α. Figure 3 presents this simulations. For each value of α we compute the distribution of the $T(\alpha)$. Note that should the true alpha be used for the computation, $T(\alpha) \sim \chi^2_1$. The simulations show that this is the case, as the percentiles computed are similar to

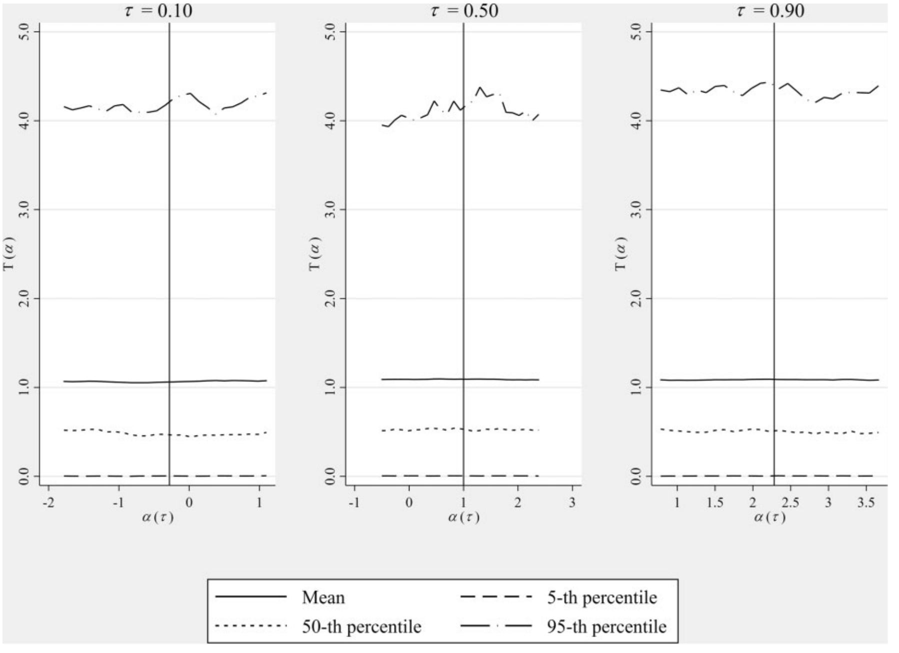

Figure 3. Percentiles of the simulated distribution of $T(\alpha)$, $a = 0$.

those of a chi-squared. More importantly, the simulations show that the distribution is not affected by the α values used to evaluate the test statistic. For instance, the 95th percentile is close to that of a chi-squared with one degree of freedom (3.8415, about 4).

Second, we consider now the case where $a = 1$, such that z_1 has a mean effect on d, Figure 4. For both cases the figures indicate that the instrument is statistically significant for the entire range of α values.

Both cases illustrate that the test statistic does not vary across the values of α. Although not reported, different specifications based on variations of the Monte Carlo DGP results in the same conclusion: the $T_n(\alpha)$ distribution is not related to the value of α.

6. EMPIRICAL APPLICATION: COLLEGE PROXIMITY AS AN INSTRUMENT FOR EDUCATION

In this section we show an empirical application of the proposed test to a Mincer equation to estimate returns to schooling. We use the data in Card (1995a) (taken from Card, 1995b) and correspond to 3,010 individuals of the US National Longitudinal Survey of Young Men. Following the same specification of that paper, the model describes wages as a function of the years of education and other exogenous controls such as work experience, race, and a set of geographic and regional variables. A classic problem with this model is that ability is unobservable and, therefore, its omission induces a potential bias due to endogeneity of the OLS estimator. Specification errors have analogous consequences on QR estimators, as analysed by Angrist et al. (2006). Angrist et al. proposes to implement an IV strategy using two measures of proximity to the university as

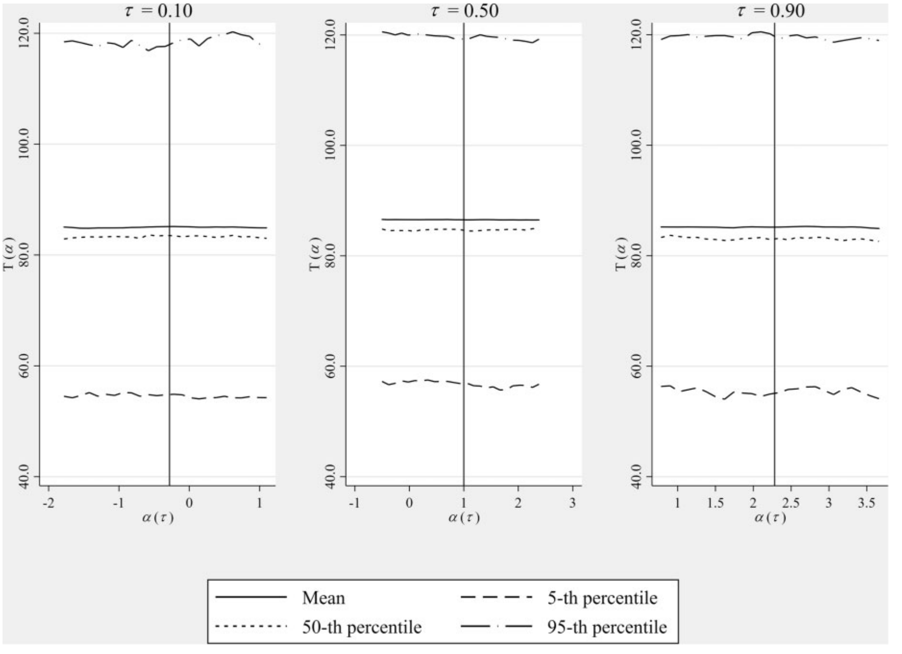

Figure 4. Percentiles of the simulated distribution of $T(\alpha)$, $a = 1$.

external variables to the wage equation: *nearc*2 (lived near 2-year college in 1966) and *nearc*4 (lived near 4-year college in 1966).

We implement first the robust testing procedure, where we evaluate the statistical first-stage significance of the IV for different α values. Here the endogenous variable is education and, therefore, we implement the FS-IVQR tests for different values of the coefficient of the returns to education. We use a grid of values on $\alpha \in [0, 1]$, which will be the same used in the CH-IVQR estimation. We consider three model specifications: (i) a model where only *nearc*2 is used as IV; (ii) a model where only *nearc*4 is used as IV; (iii) a model where both *nearc*2 and *nearc*4 are used as IV. Then for (i) we use a test for the FS-IVQR significance of *nearc*2, for (ii) we use a test for the FS-IVQR significance of *nearc*4, and for (iii) we use a test for the FS-IVQR joint significance of *nearc*2 and *nearc*4. We take a sceptical approach regarding the critical value that should be used to define that an instrument is weak or strong, as the econometric literature is still debating on this issue. Nevertheless, we rely on the (so far) accepted applied consensus that for a model with one IV, an F (i.e., chi-squared in our case) value of 10 is a strong IV.

The results appear in Figures 5 and 6. The former figure reports the corresponding $T_n(\alpha)$ for a grid of αs. Note that the estimated statistics have large variation across small changes in α, as the algorithm finds corner solutions very often. The latter figure presents a standard smoother of these results (we use a Nadaraya–Watson estimator with a bandwidth of 0.10 and Epanechnikov kernel). The results illustrate two different cases regarding the IV validity. The test statistic for *nearc*2 (based on a model that only uses this variable as IV) is below any critical value to be used for weak IV almost everywhere. For whatever critical value to be used for first-stage significance it is clear that this variable is a weak instrument. The test statistic for *nearc*4 (based on a model

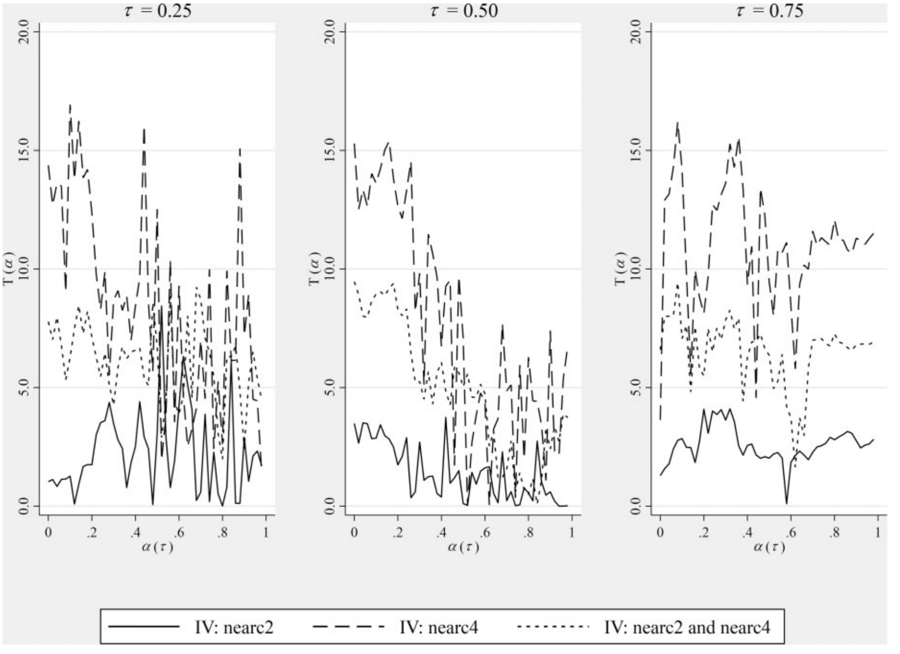

Figure 5. Robust IV validity, non-smooth.

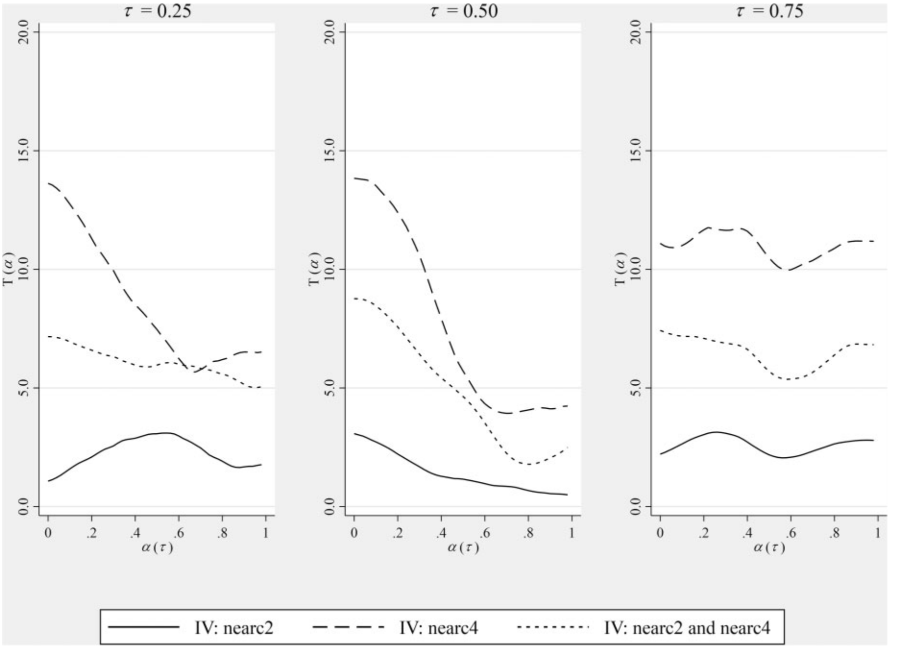

Figure 6. Robust IV validity, smooth.

Table 3. Returns to schooling (Card, 1995a), *nearc*4 as IV.

	2SLS	IV quantile regression 25	50	75
		First-stage estimates		
Lived wear 4-year college in 1966	0.320***	0.589***	0.402**	−0.103
	(0.0879)	(0.197)	(0.168)	(0.231)
Experience	– 0.413***	– 0.624***	– 0.00172	– 0.956***
	(0.0337)	(0.0600)	(0.0484)	(0.121)
Experience-squared	0.000869	0.00833***	– 0.0193***	0.0240***
	(0.00165)	(0.00298)	(0.00242)	(0.00532)
Black indicator	– 0.936***	– 0.532**	– 1.072***	– 0.395*
	(0.0937)	(0.212)	(0.173)	(0.214)
Constant	16.64***	16.79***	16.07***	17.08***
	(0.241)	(0.285)	(0.324)	(1.014)
		Structural equation estimates		
Education	0.132**	0.152***	0.132***	0.0880
	(0.0548)	(0.0315)	(0.0437)	(0.166)
Experience	0.108***	0.112***	0.105***	0.0840
	(0.0236)	(0.0215)	(0.00733)	(0.159)
Experience-squared	– 0.00233***	– 0.00208***	– 0.00223**	– 0.00203
	(0.000333)	(0.000423)	(0.000932)	(0.00400)
Black indicator	– 0.147***	– 0.152***	– 0.145***	– 0.158**
	(0.0538)	(0.0249)	(0.0503)	(0.0695)
Constant	3.666***	3.105***	3.734***	4.652
	(0.922)	(0.546)	(0.711)	(2.830)
Observations	3,010	3,010	3,010	3,010

Notes: Standard errors in parentheses. SE robust for OLS estimates. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Regional and geographic dummies are used, but omitted. Source: Card (1995b).

that only uses this variable as IV) is a strong IV for most of the α interval to be considered. For $\tau = 0.25$ and $\tau = 0.50$, $T_n(\alpha)$ is above 10 for $\alpha \in [0, 0.36]$, and it is above 3.8415 (5% critical value for χ_1^2) for all $\alpha \in [0, 1]$. For $\tau = 0.75$ the smooth version of $T_n(\alpha)$ is above 10 for $\alpha \in [0, 1]$. However, as discussed below, when we evaluate the particular estimate at the CH-IVQR value, the instrument is not significant. As a result, we cannot apply the CH-IVQR to $\tau = 0.75$, and this model should not be considered. The test statistic for *nearc*2&*nearc*4 (based on a model that uses both IV) lies in between, although the critical values to be considered here should correspond to a 2 IV case, and thus the first-stage validity provides mixed results. In most cases it lies above the critical 5% value for a χ_2^2.

The previous analysis suggests that only *nearc*4 is a valid instrument (for $\tau = 0.25, 0.50$), and *nearc*2 should be discarded. We thus estimate first a model using only *nearc*4 as the only instrument (our preferred specification), and second, another one using *nearc*2&*nearc*4 as instruments (for comparison purposes).

Tables 3 and 4 show the results of the first-stage and structural equations' results for models with *nearc*4 as the only instrument, and *nearc*2&*nearc*4 as the two-instruments case, respectively. For the latter we set A, the weighting matrix in the CH-IVQR estimator, equal to the inverse of the asymptotic covariance matrix of $\hat{\gamma}$, as suggested by CH. For the one instrument model,

Table 4. Returns to schooling (Card, 1995a), *nearc*2&*nearc*4 as IV.

	2SLS	IV quantile regression		
		25	50	75
		First-stage estimates		
Lived wear 2-year college in 1966	0.123	0.0644	0.471***	0.154**
	(0.0774)	(0.129)	(0.0703)	(0.0709)
Lived wear 4-year college in 1966	0.321***	0.380***	0.298***	0.140*
	(0.0878)	(0.146)	(0.101)	(0.0737)
Experience	– 0.412***	– 0.450***	– 0.489***	– 0.494***
	(0.0337)	(0.0871)	(0.0247)	(0.0344)
Experience-squared	0.000848	−0.000682	0.00456***	0.00449**
	(0.00165)	(0.00496)	(0.00122)	(0.00192)
Black indicator	– 0.945***	– 0.926***	– 0.886***	– 0.753***
	(0.0939)	(0.162)	(0.113)	(0.0701)
Constant	16.60***	16.42***	17.00***	16.68***
	(0.242)	(0.393)	(0.173)	(0.211)
		Structural equation estimates		
Education	0.157***	0.176***	0.268***	0.104
	(0.0524)	(0.0521)	(0.0270)	(0.0661)
Experience	0.119***	0.120***	0.180***	0.0932***
	(0.0227)	(0.0248)	(0.0139)	(0.0341)
Experience-squared	– 0.00236***	– 0.00201***	– 0.00337***	– 0.00221***
	(0.000347)	(0.000347)	(0.000352)	(0.000438)
Black indicator	– 0.123**	– 0.109**	– 0.00925	– 0.148***
	(0.0520)	(0.0519)	(0.0342)	(0.0467)
Constant	3.237***	2.698***	1.392***	4.360***
	(0.883)	(0.870)	(0.465)	(1.116)
Observations	3,010	3,010	3,010	3,010

Notes: Standard errors in parentheses. SE robust for OLS estimates. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Regional and geographic dummies are used, but omitted. Source: Card (1995b).

the estimated CH-IVQR returns to education are 0.152 (S.E. 0.0315) for $\tau = 0.25$, 0.132 (S.E. 0.0437) for $\tau = 0.50$ and 0.088 (0.166) for $\tau = 0.75$, not statistically significant in this case. Note that for the latter we have the statistical significance of the FS-IVQR evaluated at the CH-IVQR estimate signals that *near*4 is not valid. The difference between the testing procedure in the figures and the point estimate in the table should be highlighted. For the former we use a fixed value of α to compute the sparsity density estimator in each case. For the latter, we have a different CH-IVQR estimate for different τs to compute the sparsity density estimator. Thus they may not coincide with each other in finite samples. As such, we cannot use inference for the CH-IVQR estimator for this particular $\tau = 0.75$, and other inference procedures (such as Chernozhukov and Hansen, 2008) should be used instead. For the two-instrument case, the estimated CH-IVQR returns to education are 0.176 (S.E. 0.0521) for $\tau = 0.25$, 0.268 (0.0271) for $\tau = 0.50$, and 0.104 (0.0662) for $\tau = 0.75$, not statistically significant in the latter.

The FS-IVQR should be analysed with caution and it cannot be taken as in the 2SLS case. That is, the FS-IVQR coefficients in these tables are for the individual statistical significance of each instrument. The reported inference for each IV assumes that we can estimate the structural coefficients consistently.[9] Our approach requires that this analysis should be complemented with the robust IV analysis in Figures 5 and 6. When we evaluate the $T_n(\alpha)$ statistics at the estimated α coefficients, we note that $nearc4$ has values that are consistent with a strong IV for $\tau = 0.25, 0.50$, but we cannot evaluate the model for $\tau = 0.75$.

The proposed method allows for an evaluation of the first stage in a QR framework. It does not, however, provides results for the selection of the 'best' subset of IV to be used in empirical settings. Although it is beyond the scope of this paper, if that route were followed for a practitioner, it should follow post-model selection adjustment for valid inference as discussed in Pötscher (1991) and Leeb and Pötscher (2005, 2006).

7. CONCLUSIONS

This paper proposes a first-stage model and a testing procedure to evaluate the degree of association between the IV and the endogenous regressor(s) in the IVQR estimator. The procedure developed here allows to evaluate instruments in a similar vein to that in 2SLS models for the conditional average, that is, by looking at the statistical significance of the instruments in the first-stage regression. In turn, this will allow to investigate IV validity for specific quantiles.

Nevertheless, due to the requirement of consistent estimation of the weights in the first stage, it is important to notice that the testing procedure is not equivalent to the first stage in 2SLS. Tests for each individual instrument require the availability of at least another instrument. However, we propose a robust analysis that does not rely on this requirement.

The analysis may be extended in the following directions. First, this approach can be used to identify quantile-specific treatment effects, where an IV estimate being significant at some quantiles corresponds to a particular effect of a treatment. Second, the procedure outlined here could be combined with the second-stage inference to produce statistics similar to the Staiger and Stock (1997) F-statistics rule-of-thumb. In particular, to study weak instruments issues in QR models. Third, the analysis in this paper could be further extended for the cases of multiple endogenous variables as well as many instruments.

REFERENCES

Amemiya, T. (1982). Two stage least absolute deviations estimators. *Econometrica 50*, 689–711.

Andrews, I. and A. Mikusheva (2016). Conditional inference with a functional nuisance parameter. *Econometrica 84*, 1571–612.

Angrist, J., V. Chernozhukov and I. Fernández-Val (2006). Quantile regression under misspecification, with an application to the U.S. wage structure. *Econometrica 74*, 539–63.

Card, D. (1995a). Using geographic variation in college proximity to estimate the return to schooling. In L. N. Christofides, E. K. Grantand and R. Swidinsky (Eds.), *Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp*, 201–22. Toronto: University of Toronto Press.

Card, D. (1995b). Database for 'Using geographic variation in college proximity to estimate the return to schooling'. Data available in https://www.stata.com/data/jwooldridge/eacsap/card.dta.

Chen, L.-Y. and S. Lee (2018). Exact computation of GMM estimators for instrumental variable quantile regression models. *Journal of Applied Econometrics 33*(4), 553–67.

[9] We thank an anonymous referee for this observation.

Chen, L.-A. and S. Portnoy (1996). Two-stage regression quantiles and two-stage trimmed least squares estimators for structural equation models. *Communication in Statistics, Theory Methods 25*, 1005–32.

Chernozhukov, V. and C. Hansen (2004). The effects of 401(k) participation on the wealth distribution: An instrumental quantile regression analysis. *Review of Economics and Statistics 86*, 735–51.

Chernozhukov, V. and C. Hansen (2005). An IV model of quantile treatment effects. *Econometrica 73*, 245–61.

Chernozhukov, V. and C. Hansen (2006). Instrumental quantile regression inference for structural and treatment effects models. *Journal of Econometrics 132*, 491–525.

Chernozhukov, V. and C. Hansen (2008). Instrumental variable quantile regression: A robust inference approach. *Journal of Econometrics 142*, 379–98.

Chernozhukov, V., C. Hansen and M. Jansson (2007). Inference approaches for instrumental variable quantile regression. *Economics Letters 95*, 272–7.

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 (2020). Instrumental variable quantile regression. In R. Koenker, V. Chernozhukov, X. Heand and L. Peng (Eds.), *Handbook of Quantile Regression*, 119–44. Boca Raton, FL: CRC Press.

Chesher, A. (2003). Identification in nonseparable models. *Econometrica 71*, 1405–41.

Chetverikov, D., B. Larsen and C. Palmer (2016). IV quantile regression for group-level treatments, with an application to the distributional effects of trade. *Econometrica 84*, 809–33.

De Castro, L., A. F. Galvao, D. M. Kaplan and X. Liu (2019). Smoothed GMM for quantile models. *Journal of Econometrics 213*, 121–44.

Dufour, J.-M. (1997). Some impossibility theories in econometrics with applications to structural and dynamic models. *Econometrica 65*, 1365–88.

Firpo, S., A. F. Galvao, C. Pinto, A. Poirier and G. Sanroman (2022). GMM quantile regression. *Journal of Econometrics 230*(2), 432–52.

Galvao, A. F. (2011). Quantile regression for dynamic panel data with fixed effects. *Journal of Econometrics 164*, 142–57.

Galvao, A. F. and G. Montes-Rojas (2015). On the equivalence of instrumental variables estimators for linear models. *Economics Letters 134*, 13–15.

Hall, P. and S. J. Sheather (1988). On the distribution of a studentized quantile. *Journal of the Royal Statistical Society: Series B 50*(3), 381–91.

Hendricks, W. and R. Koenker (1992). Hierarchical spline models for conditional quantiles and the demand for electricity. *Journal of the American Statistical Association 87*, 58–68.

Jun, S. J. (2008). Weak identification robust tests in an instrumental quantile model. *Journal of Econometrics 144*, 118–38.

Kaido, H. and K. Wüthrich (2021). Decentralization estimators for instrumental variable quantile regression models. *Quantitative Economics 12*, 443–75.

Kaplan, D. M. and Y. Sun (2017). Smoothed estimating equations for instrumental variables quantile regression. *Econometric Theory 33*, 105–57.

Kim, T.-H. and C. Muller (2004). Two-stage quantile regression when the first stage is based on quantile regression. *Econometrics Journal 7*, 218–31.

Koenker, R. (2005). *Quantile Regression*. New York: Cambridge University Press.

Koenker, R. and Q. Zhao (1994). L-estimation for linear heteroscedastic models. *Journal of Nonparametric Statistics 3*, 223–35.

Koenker, R. and Q. Zhao (1996). Conditional quantile estimation and inference for ARCH models. *Econometric Theory 12*, 793–813.

Lee, D. S., J. McCrary, M. J. Moreira and J. Porter (2022). Valid t-ratio inference for IV. *American Economic Review 112*, 3260–90.

Lee, S. (2007). Endogeneity in quantile regression models: A control function approach. *Journal of Econometrics 141*, 1131–58.

Leeb, H. and B. M. Pötscher (2005). Model selection and inference: Facts and fiction. *Econometric Theory 21*, 21–59.

Leeb, H. and B. M. Pötscher (2006). Can one estimate the conditional distribution of post-model-selection estimators? *Annals of Statistics 34*, 2554–91.

Ma, L. and R. Koenker (2006). Quantile regression methods for recursive structural equation models. *Journal of Econometrics 134*, 471–506.

Ota, H., K. Kato and S. Hara (2019). Quantile regression approach to conditional mode estimation. *Electronic Journal of Statistics 13*, 3120–60.

Pötscher, B. M. (1991). Effects of model selection. *Econometric Theory 7*, 163–85.

Pouliot, G. A. (2019). Instrumental variables quantile regression with multivariate endogenous variable. Working paper, University of Chicago. https://bfi.uchicago.edu/wp-content/uploads/MILP_for_IVQR_V4_4.pdf.

Powell, J. L. (1983). The asymptotic normality of two-stage least absolute deviations estimators. *Econometrica 51*, 1569–76.

Powell, J. L. (1991). Estimation of monotonic regression models under quantile restrictions. In W. Barnett, J. Powell and G. Tauchen (Eds.), *Nonparametric and Semiparametric Models in Econometrics*, 357–84. Cambridge: Cambridge University Press.

Sanderson, E. and F. Windmeijer (2016). A weak instrument F-test in linear IV models with multiple endogenous variables. *Journal of Econometrics 190*, 212–21.

Staiger, D. and J. H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica 65*, 557–66.

Stock, J. H. and M. Yogo (2005). Testing for weak instruments in linear IV regression. In D. W. K. Andrews (Ed.), *Identification and Inference for Econometric Models*, 80–108. New York: Cambridge University Press.

Zhao, Q. (2001). Asymptotically efficient median regression in the presence of heteroscedasticity of unknown form. *Econometric Theory 17*, 765–84.

Zhou, K. Q. and S. L. Portnoy (1996). Direct use of regression quantiles to construct confidence sets in linear models. *Annals of Statistics 24*, 287–306.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Replication Package

Co-editor Victor Chernozhukov handled this manuscript.

APPENDIX: PROOFS OF RESULTS

Proof of Lemma 3.1: First, consider the estimator of the parameter $\mu_0 := (\mathrm{E}[f_i.w_i w_i'])^{-1}\mathrm{E}[f_i.w_i d_i'])$ using the true weighting matrix V as

$$V = \begin{bmatrix} f_1 & & \\ & \ddots & \\ & & f_n \end{bmatrix}, \tag{A.1}$$

that is given by

$$\tilde{\mu} = (W'VW)^{-1}W'VD,$$

where $W = [X, Z]$.

The proof of the lemma requires first showing that, as $n \to \infty$,

$$\frac{W'VW}{n} \xrightarrow{p} \mathrm{E}[f_i.w_i w_i'] := \Omega_f, \tag{A.2}$$

$$\frac{W'VD}{n} \xrightarrow{p} \mathrm{E}[f_i.w_i d_i], \tag{A.3}$$

and

$$\sqrt{n}(\tilde{\mu} - \mu_0) \xrightarrow{d} N(0, \Omega_f^{-1}\Omega_{f\sigma}\Omega_f^{-1}). \tag{A.4}$$

To show (A.2), its left side has a (j, k) element given by $\frac{1}{n}\sum_{i=1}^n f_i w_{ij} w_{ik}$. By the Law of Large Numbers and Assumptions 3.1 and 3.2, we have that

$$\frac{1}{n}\sum_{i=1}^{n} f_i w_{ij} w_{ik} \xrightarrow{p} \mathrm{E}\left[f_i w_{ij} w_{ik}\right].$$

Similar arguments can be used to show (A.3).

To show (A.4), note that

$$\tilde{\mu} = \left(\frac{1}{n}W'VW\right)^{-1}\frac{1}{n}W'VD.$$

Consider $W'VD$, which is a sum of i.i.d. random vectors $f_i \cdot w_i \cdot d_i$ with common covariance matrix, $\Omega_{f\sigma}$. Note that, by definition, $\mathrm{E}[f_i d_i w_i] = \mathrm{E}[f_i w_i w_i']\mu_0$. Therefore, by Assumptions 3.1 and 3.2 and the central limit theorem, we have that

$$\sqrt{n}\left(\frac{1}{n}W'VD - \Omega_f\mu_0\right) \xrightarrow{d} N\left(0, \Omega_{f\sigma}\right).$$

Thus,

$$\sqrt{n}(\tilde{\mu} - \mu_0) \xrightarrow{d} N\left(0, \Omega_f^{-1}\Omega_{f\sigma}\Omega_f^{-1}\right),$$

where this display holds because of the convergence in distribution above, $\frac{1}{n}W'VW \xrightarrow{p} \Omega_f$ as $n \to \infty$, the continuous mapping theorem, and Slutsky's theorem.

Finally, we have to show that using estimated weights does not affect the liming distribution when considering

$$\hat{\mu} = (W'\hat{V}W)^{-1}W'\hat{V}D,$$

where the matrix $\hat{V}$ contains $\{\hat{f}_i\}$ the density estimator based on a (possibly) nonparametric estimator (evaluated at the CH-IVQR estimator) of the density $\{f_i\}$.

We apply the results in Koenker (2005, pp. 162–3) for the asymptotic distribution of weighted estimators with estimated weights. For the particular application in the book, it is for a weighted QR estimator; for our case, we use it for a weighted least-squares (WLS) estimator.

The parameter to be estimated is

$$\mu_0 = \underset{\mu}{\operatorname{argmin}}\ E[f_i \cdot (d_i - w_i \mu)^2].$$

Note that in our case we have the following infeasible estimator

$$\tilde{\mu} = \underset{\mu}{\operatorname{argmin}} \sum_i^n f_i \cdot (d_i - w_i \mu)^2,$$

for the estimator with the true weights, f_i. The feasible estimator is

$$\hat{\mu} = \underset{\mu}{\operatorname{argmin}} \sum_i^n \hat{f}_i \cdot (d_i - w_i \mu)^2,$$

for the estimated weights, $\hat{f}_i$. Koenker's result is about the asymptotic distribution of $n^{1/2}(\hat{\mu} - \mu_0)$ and how it compares it with $n^{1/2}(\tilde{\mu} - \mu_0)$. In particular, it shows that $n^{1/2}(\tilde{\mu} - \mu_0)$ and $n^{1/2}(\hat{\mu} - \mu_0)$ are asymptotically normal with the same variance.

In both cases, it uses a Bahadur-type representation of an $\sqrt{n}$-consistent estimator such that for our particular case (i.e., least squares) it is

$$n^{1/2}(\tilde{\mu} - \mu_0) = n^{-1/2} \sum_{i=1}^n f_i \cdot (y_i - w_i \mu_0) w_i' + o_p(1),$$

$$n^{1/2}(\hat{\mu} - \mu_0) = n^{-1/2} \sum_{i=1}^n \hat{f}_i \cdot (y_i - w_i \mu_0) w_i' + o_p(1).$$

Then by the Assumption 3.2 (ii), together with the bounded conditions in Assumption 3.1 R1, the weights satisfy the same conditions in Koenker's derivation, and the main result is that

$$n^{-1/2} \sum_{i=1}^n \hat{f}_i \cdot (y_i - w_i \mu_0) w_i' = n^{-1/2} \sum_{i=1}^n f_i \cdot (y_i - w_i \mu_0) w_i' + o_p(1).$$

Thus, we conclude that both estimators, $\tilde{\mu}$ and $\hat{\mu}$ have the same asymptotic distribution. □

Proof of Proposition 4.1: The proof of this result is simple. If α is known and correct, $\hat{f} \xrightarrow{p} f$. Then, the result follows directly from Lemma 3.1, the null hypothesis, a consistent estimator of $V(\mu_0)$, and the Slutsky's theorem. □

Proof of Proposition 4.2: From the conditions of the proposition we have that $\hat{f} \xrightarrow{p} f$. Hence, conditions of Lemma 3.1 are satisfied, and it follows that,

$$\sqrt{n}(\hat{\mu} - \mu_0) \xrightarrow{d} N(\mathbf{0}, V(\mu_0)).$$

Notice that $R\mu = \delta_1$, hence under the null hypothesis,

$$\sqrt{n}(R\hat{\mu} - \mathbf{0}) \xrightarrow{d} N\left(\mathbf{0}, RV(\mu_0)R'\right).$$

Let $\hat{V}(\hat{\mu})$ be a consistent estimator of $V(\mu_0)$, and $V_{\delta_1} := RV(\mu_0)R'$, then by the Slutsky's theorem,

$$T_n = n\hat{\delta}_1' \left(V_{\delta_1}\right)^{-1} \hat{\delta}_1 \xrightarrow{d} \chi^2_{p_1}(\boldsymbol{a}_{p_1}).$$

□