Red Nacional de Investigadores en Economía

A first-stage representation for instrumental variables quantile

Javier Alejo (IECON/Universidad de la República)
Antonio Galvao (University of Arizona)
Gabriel Montes-Rojas (Universidad de Buenos Aires)

DOCUMENTO DE TRABAJO 2021-46

Marzo de 2021

Citar como:
Alejo, Javier, Antonio Galvao y Gabriel Montes-Rojas (2021). A first-stage representation for instrumental variables quantile. *Documento de trabajo RedNIE, 2021-46.*

A first-stage representation for instrumental variables quantile regression

Javier Alejo[*] Antonio F. Galvao[†] Gabriel Montes-Rojas[‡]

February 10, 2021

Abstract

This paper develops a first-stage linear regression representation for the instrumental variables (IV) quantile regression (QR) model. The first-stage is analogue to the least squares case, i.e., a conditional mean regression of the endogenous variables on the instruments, with the difference that for 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 endogeneous and exogenous covariates, and the instruments as well, at a given quantile. The first-stage regression is a natural framework to evaluate the validity of instruments. Thus, we are able to use the first-stage result and suggest testing procedures to evaluate the adequacy of instruments in IVQR models by evaluating their statistical significance. In the QR case, the instruments may be relevant at some quantiles but not at others or at the mean. Monte Carlo experiments provide numerical evidence that the proposed tests work as expected in terms of empirical size and power in finite samples. An empirical application illustrates that checking for the statistical significance of the instruments at different quantiles is important.

Keywords: Quantile regression, instrumental variables, first-stage.

JEL: C13, C23.

[*]IECON-Universidad de la República, Gonzalo Ramirez 1926, C.P. 11200, Montevideo, Uruguay. E-mail: javier.alejo@ccee.edu.uy

[†]University of Arizona, Tucson, USA. E-mail: agalvao@email.arizona.edu

[‡]Universidad de Buenos Aires, Ciudad Autónoma de Buenos Aires, Argentina. E-mail: gabriel.montes@fce.uba.ar

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. It is also important, however, that a second condition for a valid instrument, instrument relevance, holds, for if the instruments are only marginally relevant, or "weak," then first-order asymptotics can be a poor guide to the actual sampling distributions of conventional IV regression statistics. Several testing procedures have been proposed to evaluate the presence of weak instruments, as well as alternative robust inference methods. The most popular test to evaluate the weak instruments problem looks at the first-stage (i.e. a linear regression of the endogenous variable on the IV and other exogenous covariates) F-statistics following the rule-of-thumb of Staiger and Stock (1997) and subsequent variants as Sanderson and Windmeijer (2016), Lee et al. (2020) and others. See Stock and Yogo (2005) for an extensive discussion.

Quantile regression (QR) is an important method of modeling heterogeneous effects. 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 (2005, 2006, 2008) (CH hereafter) develop an instrumental variables quantile regression (IVQR) procedure that has been applied in several contexts. It is one of the most prolific approaches in terms of subsequent work, as it provides a general procedure to use IV for endogeneity of regressors (see, e.g., Angrist et al., 2006; Chernozhukov et al., 2009; Galvao, 2011). Other work, based on this idea, develop the GMM counterpart constructed using moment conditions directly, see, for instance, Kaplan and Sun (2017) and de Castro et al. (2019). We refer to Chernozhukov et al. (2020) for an overview of IVQR.

CH comment that their method is a simple solution to a two-stage least-squares (2SLS) analog, which has been formally established in Galvao and Montes-Rojas (2015). However, the first-stage of the IVQR estimator has not been explicitly considered, as it is implemented as an inverse QR estimator. The IVQR estimator contrasts to alternative procedures where the first-stage is implemented in alternative frameworks. 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 first step consists of estimation of the residuals of the reduced-form equation for the endogenous explanatory variable. Ma and Koenker (2006) presents an estimator for a recursive structural equation model.

This paper builds on the IVQR estimator and shows that a first-stage regression model can be explicitly recovered from the CH IVQR estimator. The first-stage IVQR (FS-IVQR) is a conditional mean regression of the endogenous variables on the instruments, 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 endogeneous and exogenous covariates together with the instruments, at a given quantile. The derivation of the result is simple. We write the IVQR estimator as a constrained Lagrangian optimization problem and show that one of the restrictions that must be satisfied is the analogue of the first-stage.

The practical implementation of the FS-IVQR is 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 can be estimated, for instance, using the sparsity method (see, e.g., Koenker (2005)). Second, a standard WLS model is implemented – this is analogue to the first-stage model used in 2SLS. We derive the asymptotic distribution of the two-step estimator.

The first-stage regression is a natural framework to evaluate the validity of instruments since one can test for their statistical significance, that is, how the IV impact on the endogenous variable(s). Furthermore, it can also be used for testing procedures to assess the validity of the IV for given quantiles.[1] A Wald-type test on the coefficients of interest can be used for testing. Provided that weights are consistently estimated, the Wald test is asymptotically Chi-squared with the number of degrees of freedom equal to the number of coefficients tested. We highlight that when testing for one instrument being invalid, the empirical applications are restricted to the case with at least two instruments, with one being valid. This is because the practical implementation of the first-stage relies on a consistent estimation of the weights in the first step. In sum, the proposed method evaluates the relevance condition for the validity of the IV in the IVQR framework. The procedure, thus, allows the empirical researcher to evaluate the individual quality of the IV.

The proposed inference allows for a procedure in empirical work that is parallel to the standard first-stage in two-stage least squares (FS-2SLS), to evaluate the degree of association of the IV to the endogenous variable. Nevertheless, the derivation of the FS-IVQR result illustrates that the rejection of the null hypothesis considered here is a necessary condition for the validity of the IVQR specification to be used in practice for identification. Our procedure should be thus evaluated in a framework where one is concerned with the first-stage relevance of the IV in similar vein as 2SLS for estimating mean effects. It provides a clear link between the IV evaluation in an explicit first-stage in least-squares and QR models.

One important feature of the procedure developed here is that instruments could be statistically insignificant in FS-2SLS, but they could still be related to the endogenous variable in the IVQR set-up. The reason is that the FS-2SLS test only evaluates a mean effect, but

[1] There is a literature on weak identification robust inference for QR models. Without imposing additional conditions, statistical inference for the structural quantile function can be performed using weak-identification robust inference as described in Chernozhukov and Hansen (2008), Jun (2008), or Chernozhukov et al. (2009). In this paper we do not pursue this avenue and leave it for future research.

the FS-IVQR, because of its specific weighting procedure, allows for different first-stage effects 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 Monte Carlo experiments illustrate this issue. 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 tests. The proposed method has correct size in all cases where the structural parameters can be estimated under the null hypothesis. We consider other cases where there is no identification under the null with mixed performances. The developed test has excellent power properties. In particular the experiments highlight the case where the FS-2SLS test suggest the instrument is not valid, but the proposed procedure finds it is for some quantiles.

As an empirical illustration, we apply the FS-IVQR estimator to the Card (1995) data on instrumenting education using college proximity. The analysis reveals heterogeneity in the significance of the IV across quantiles. In fact, while the 2SLS analysis shows that one instrument (proximity to 2-year college) is not statistically significant in the first-stage, it is indeed for high quantiles.

This paper is organized as follows. Section 2 briefly reviews the CH IVQR estimator and rewrites that estimator as a constrained minimization problem and derives the first-stage representation for the IVQR. Section 3 discusses its empirical implementation and derives the estimators' asymptotic distribution. Section 4 presents the first-stage test for validity of instruments. Section 5 provides finite sample Monte Carlo evidence. Section 6 applies the proposed tests to empirical problems. Finally, Section 7 concludes.

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 control 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 generalization of the QR model with endogenous regressors. A linear representation of the model takes the following form

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

where u_d is the nonseparable error or rank. Under some regularity conditions, CH establish the following IV identification function

$$P[y \leq d\alpha_0(\tau) + x\beta_0(\tau)|x, z] = P[u_d \leq \tau|x, z] = \tau. \tag{2}$$

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) can be used to estimate the parameters of interest. For a given quantile τ, the population IVQR estimator for model in (1), is given by

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

where

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

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 GMM estimator where the QR first order conditions are used as moment conditions. In particular, it would involve a Z-estimator solving

$$\mathrm{E}\left[x'\left(\mathbf{1}[y - d\alpha - x\beta < 0] - \tau\right)\right] = \mathbf{0}_k, \tag{3}$$

$$\mathrm{E}\left[z'\left(\mathbf{1}[y - d\alpha - x\beta < 0] - \tau\right)\right] = \mathbf{0}_p,, \tag{4}$$

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 (3)–(4). Kaplan and Sun (2017) 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 IVQR one. This can be seen in the fact that the term $z\gamma$ is not considered altogether from the regression model.

2.2 The IVQR estimator as a constrained minimization problem

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

$$\min_{(\alpha,\beta,\gamma)} \|\gamma\|_A, \tag{5}$$

subject to

$$\mathrm{E}\left[x'\left(\mathbf{1}[y - d\alpha - x\beta - z\gamma < 0] - \tau\right)\right] = \mathbf{0}_k, \tag{6}$$

$$\mathrm{E}\left[z'\left(\mathbf{1}[y - d\alpha - x\beta - z\gamma < 0] - \tau\right)\right] = \mathbf{0}_p. \tag{7}$$

Now we write this constrained optimization as a Lagrangian problem as

$$\mathcal{L}(\alpha,\beta,\gamma,\lambda_x,\lambda_z) = \|\gamma\|_A + \lambda_x \mathrm{E}\left[x'(\mathbf{1}[y - d\alpha - x\beta - x\gamma < 0] - \tau)\right] \\ + \lambda_z \mathrm{E}\left[z'(\mathbf{1}[y - d\alpha - x\beta - x\gamma < 0] - \tau)\right], \tag{8}$$

where λ_x is a $1 \times k$ vector and λ_z is a $1 \times p$ vector. 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 equation (8) are

$$\partial\mathcal{L}/\partial\alpha = -\left\{\lambda_x \mathrm{E}\left[f \cdot x'd\right] + \lambda_z \mathrm{E}\left[f \cdot z'd\right]\right\}' \tag{9}$$

$$\partial\mathcal{L}/\partial\beta = -\left\{\lambda_x \mathrm{E}\left[f \cdot x'x\right] + \lambda_z \mathrm{E}\left[f \cdot z'x\right]\right\}' \tag{10}$$

$$\partial\mathcal{L}/\partial\gamma = \left\{2\gamma' A - \lambda_x \mathrm{E}\left[f \cdot x'z\right] - \lambda_z \mathrm{E}\left[f \cdot z'z\right]\right\}' \tag{11}$$

$$\partial\mathcal{L}/\partial\lambda_x = \mathrm{E}\left[x'(\mathbf{1}[y - d\alpha - x\beta - z\gamma < 0] - \tau)\right]' \tag{12}$$

$$\partial\mathcal{L}/\partial\lambda_z = \mathrm{E}\left[z'(\mathbf{1}[y - d\alpha - x\beta - z\gamma < 0] - \tau)\right]', \tag{13}$$

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 a central role in what follows.

The solution should have all equations above equal to zero when assuming an interior solution as in Assumption 1 below. Thus, from equation (10),

$$\lambda_x' = -\left(\mathrm{E}[f \cdot x'x]\right)^{-1}\left(\mathrm{E}[f \cdot x'z]\right)'\lambda_z'. \tag{14}$$

Then, replacing (14) in (11),

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

such that

$$\lambda_z' = 2\left(\mathrm{E}[f \cdot z'x] - \mathrm{E}[f \cdot z'x](\mathrm{E}[f \cdot x'x])^{-1}\mathrm{E}[f \cdot x'z]\right)^{-1} A\gamma. \tag{15}$$

Finally, replacing (15) in (9),

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

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

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

$$\begin{aligned}
&\left\{\mathrm{E}\left[f \cdot d'z\right] - \mathrm{E}\left[f \cdot d'x\right] (\mathrm{E}\left[f \cdot x'x\right])^{-1}\mathrm{E}\left[f \cdot x'z\right]\right\} \times \\
&\left\{\mathrm{E}\left[f \cdot z'z\right] - \mathrm{E}\left[f \cdot z'x\right] (\mathrm{E}\left[f \cdot x'x\right])^{-1}\mathrm{E}\left[f \cdot x'z\right]\right\}^{-1} A\gamma = \mathbf{0}_r && (16)\\
&\mathrm{E}\left[x \cdot (\mathbf{1}[y - d\alpha - x\beta - z\gamma < 0] - \tau)\right] = \mathbf{0}_k && (17)\\
&\mathrm{E}\left[z \cdot (\mathbf{1}[y - d\alpha - x\beta - z\gamma < 0] - \tau)\right] = \mathbf{0}_p. && (18)
\end{aligned}$$

2.3 First-stage IVQR parameters

Given equations (16)–(18) above, we can see that (16) provides a first-stage representation of the IVQR model. This can be written as

$$\delta' A\gamma = \mathbf{0}_r, \tag{19}$$

where

$$\begin{aligned}
\delta := &\left\{\mathrm{E}\left[f \cdot z'z\right] - \mathrm{E}\left[f \cdot z'x\right] (\mathrm{E}\left[f \cdot x'x\right])^{-1}\mathrm{E}\left[f \cdot x'z\right]\right\}^{-1} \\
&\left\{\mathrm{E}\left[f \cdot z'd\right] - \mathrm{E}\left[f \cdot z'x\right] (\mathrm{E}\left[f \cdot x'x\right])^{-1}\mathrm{E}\left[f \cdot x'd\right]\right\}.
\end{aligned} \tag{20}$$

Here δ is a $p \times r$ vector. Notice that equation (20) is a least-squares projection. In particular, the representation in (20) is a weighted conditional mean regression, 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 endogeneous and exogenous covariates together with the instruments.

Hence, for each endogeneous variable, say d_j for $j = 1, 2, ..., r$, δ_j in equation (20) can be recovered as the solution to the following optimization problem

$$\mu_j := (\psi_j, \delta_j) = \underset{\psi,\delta}{\operatorname{argmin}}\ \mathrm{E}\left[f \cdot (d_j - x\psi - z\delta)^2\right]. \tag{21}$$

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 equation (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 set-up, thus confirming that a first-stage (albeit different) is part of the model.

3 Empirical implementation and asymptotic distribution

In this section we consider the empirical implementation of the first-stage instrumental variables quantile regression (FS-IVQR) and derive the estimators' asymptotic distribution. We propose a two steps estimation procedure, where in the first step we estimate the density using the IVQR model, and in the second step we use a weighted least squares (WLS) regression. For simplicity of exposition, we develop the case of $r = 1$, i.e. one endogenous variable.

3.1 Estimator

The estimator requires a consistent estimator of μ in (21), which will be based on WLS based on the estimator of f, at a given quantile of interest τ. The estimator has two steps as following:

1) In the first step we obtain $\hat{\theta} = (\hat{\alpha}, \hat{\beta}', \hat{\gamma}')'$ from the CH estimator,

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

where

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

Provided that the τth conditional quantile function of $y|s$ is linear, as in (1), then for $h_n \to 0$ we can consistently estimate the parameters of the $\tau \pm h_n$ conditional quantile functions by $\hat{\theta}(\tau \pm h_n)$. And the density $f_i := f_{u_\tau}(0|d = d_i, x = x_i, z = z_i)$ can thus be estimated by the difference quotient

$$\hat{f}_i = \frac{2h_n}{s_i\left(\hat{\theta}(\tau + h_n) - \hat{\theta}(\tau - h_n)\right)}. \tag{22}$$

The estimation in (22) is a natural extension of sparsity estimation methods, suggested by Hendricks and Koenker (1992). The estimator is discussed in further details in Zhou and Portnoy (1996) and Koenker (2005). We introduce the simplifying notation $\hat{f}_i := \hat{f}_{u_\tau}(0|s = s_i)$.[2] The bandwidth for the density estimation can be chosen heuristically as a scaled version

[2] 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.

of Hall and Sheather (1988):

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

2) In the second step the parameters of interest δ can be obtained from a feasible WLS as

$$\hat{\mu} := (\hat{\psi}, \hat{\delta}) = \underset{\psi,\delta}{\operatorname{argmin}} \frac{1}{n}\sum_{i=1}^{n}\left[\hat{f}_i \cdot (d_i - x_i\psi - z_i\delta)^2\right]. \tag{23}$$

Equation (23) 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 (23) above can be written in a simple matrix notation as

$$\hat{\mu} = (W'\hat{V}W)^{-1}W'\hat{V}D. \tag{24}$$

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 *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. Example 1 (location model) Appendix B shows a case where the density function is a constant. A typical example where the weights are not constant across individuals is the location-scale model, see Examples 2 and 3 in Appendix B.

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.501–502), that we reproduce here for convenience. It imposes conditions for θ_0 to be identified and estimated.

Assumption 1. *R1. Sampling.* $\{y_i, x_i, d_i, z_i\}$ *are iid defined on a probability space and take values in a compact set.*
R2. Compactness and convexity. For all $\tau \in (0,1)$, $(\alpha, \beta, \gamma \in \operatorname{int}(\mathcal{A}\times\mathcal{B}\times\mathcal{G})$ *is compact and*

convex.

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

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

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 (23) using the sparsity estimation in (22).

Assumption 2. *Let* $\varepsilon_i := d_i - x_i\psi_0 - z_i\delta_0$*, with* $\mathrm{E}[\varepsilon_i|w_i] = 0$*, and* $\mathrm{E}[\varepsilon_i^2|w_i] = \sigma_i^2$*. Also, let* $f_i := f_{\theta_0}(y - s\theta_0|s = s_i)$ *and assume that* $\mathrm{E}[|f_i^{-2}w_i\varepsilon_i|] < \infty$*. Let* $\Omega_{f\sigma} := \mathrm{E}[f_i^2\sigma_i^2 w_i w_i']$ *and* $\Omega_f := \mathrm{E}[f_i w_i w_i']$*. The limits* $\lim_{n\to\infty} \frac{1}{n}\sum_i^n f_i^2\sigma_i^2 w_i w_i' = \Omega_{f\sigma}$ *and* $\lim_{n\to\infty} \frac{1}{n}\sum_i^n f_i w_i w_i' = \Omega_f$ *exist and are nonsingular (and hence finite).*

Assumption 2 contains conditions for establishing consistency and asymptotic normality of the proposed estimator. The next result presents an intermediate result.

Lemma 1. *Under Assumptions 1–2, as* $n \to \infty$, $h_n \to 0$ *and* $nh_n^2 \to \infty$,

$$\sqrt{n}\,(\hat{\mu} - \mu_0) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}_{k+p}, V(\mu_0)\right), \tag{25}$$

where $\mu_0 := (\psi_0, \delta_0) = \underset{\psi,\delta}{\mathrm{argmin}}\ \mathrm{E}\left[f \cdot (d - x\psi - z\delta)^2\right]$ *and* $V(\mu_0) = \Omega_f^{-1}\Omega_{f\sigma}\Omega_f^{-1}$ *is the asymptotic covariance matrix.*

Proof. In the Appendix A. □

4 Testing

In this section we derive tests for the validity of the IV using the first-stage representation.

4.1 Formulation and null hypothesis

The restriction in equation (19) provides a natural framework to evaluate the relevance of the instruments in QR models.

First, 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 effect of z on the endogenous variables d for the 2SLS.[3] When the instrument is valid, $\delta \neq \mathbf{0}_{p\times r}$.

Second, note that the instrument z does not belong in the structural quantile model (1), hence when z is valid, $\gamma = \mathbf{0}_{p\times r}$ can be used for identification, a key feature of the CH IVQR estimator. Another way to see this is the following. Equation (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 set-up. Therefore, a test for the validity of the instruments can then be based on inference on δ.

Another way of gaining intuition on the test is the following. Assume that $r = 1$ (i.e. only one endogenous variable), then (19) is in fact equal to 0, a scalar. If we further assume that $A = I_p$, then

$$\sum_{q=1}^{p} \delta_q \gamma_q = 0, \tag{26}$$

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.

The formulation of the test proposed in this paper is based on the condition given in equation (16) together with the first-stage IVQR representation in equation (20). A test for validity of the instruments for p instruments can be based on the null hypothesis

$$H_0 : \delta_0 = \mathbf{0}_{p\times r}, \tag{27}$$

against the alternative

$$H_A : \delta_0 \neq \mathbf{0}_{p\times r}. \tag{28}$$

We highlight that, differently from the 2SLS, the first-stage IVQR in (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 variation is captured by the weights f.

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

The expressions of the null and the alternative hypotheses in (27) and (28), respectively, lead to the following testing procedure.

When H_0 is true, under suitable regularity conditions, $\hat{\delta}$ converges in probability to $\mathbf{0}_{p\times r}$

[3] As noted by Galvao and Montes-Rojas (2015) the CH set-up 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.

for a given τ. On the other hand, when H_1 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.

A natural choice to test H_0 against H_1 for the case of $r = 1$ is the Wald statistic as

$$T_n = n\hat{\delta}'\{V_\delta\}^{-1}\hat{\delta}, \tag{29}$$

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.

4.2 Asymptotic distribution

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 eq. (29).

Proposition 1. *Consider Assumptions 1–2, $n \to \infty$, $h_n \to 0$ and $nh_n^2 \to \infty$. Furthermore, assume that $dim(z) = p > p_1 \geq 1$. Then, under $H_0 : \delta_1 = \mathbf{0}_{p_1}$ and local alternatives $H_A : \delta_1 = \boldsymbol{a}_{p_1}/\sqrt{n}$*

$$T_n = n\,(R\hat{\mu})'\,\{R\hat{V}(\hat{\mu})R'\}^{-1}\,(R\hat{\mu}) \xrightarrow{d} \chi^2_{p_1}(\boldsymbol{a}_{p_1}). \tag{30}$$

Proof. In the Appendix A. □

Computation of the test statistic (29) requires a non-parametric 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.

5 Monte Carlo experiments

We analyze 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 model:

$$y_i = d_i + x_i + (1 + cd_i)u_i, \tag{31}$$

$$d_i = az_{1i} + \phi z_{2i} + (1 + bz_{1i})v_i, \tag{32}$$

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. Equations (31)–(32) specify a model where there could be pure location or location-scale specifications in either the first- and/or the second-stages. 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). Next the parameter c determines if the structural second-stage model is a location or location-scale model.

In all cases we consider 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 equation (31) 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).

Tables 1–4 show the empirical size (i.e. $a = b = 0$) of the computed test with 2000 simulations for $n = \{500, 1000\}$ and for the quantiles $\tau = \{0.25, 0.50, 0.75\}$. The simulations show correct empirical size performance in most but not all cases.

Consider first the case where there is a second instrument, $\phi = 1$ in Tables 1 and 2. The tests have approximately correct empirical size. As such they clearly evaluate if the instrument z_1 exerts an effect on the endogenous variable d. Note that the empirical size is improved when we consider a location-scale model $c = 1$.

Now consider the case where there is no available second instrument, $\phi = 0$, in Tables 3 and 4. 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. Note that for the location-only model, Table 3, the test of sparsity estimator is oversized. This is mostly due to the implicitly estimated sparsity function as the test with the true density function has correct size. However, when we use a location-scale model, Table 4, the size is correct for the sparsity estimator. This result suggests that the test can be used if there is a location-scale structure in the second-stage, even when the structural parameters cannot be estimated under the null (because z_1 does not solve the endogeneity problem).

Table 1: Rejection rate of the null hypothesis using $a = b = 0$, model with $c = 0$ and $\phi = 1$

τ	*Size*	$n = 500$			$n = 1000$		
		FS-2SLS	**True f**	**Sparsity f**	**FS-2SLS**	**True f**	**Sparsity f**
0.25	**0.10**	0.100	0.100	0.147	0.104	0.104	0.139
	0.05	0.052	0.048	0.089	0.054	0.051	0.079
	0.01	0.014	0.013	0.038	0.008	0.008	0.024
0.50	**0.10**	0.100	0.095	0.128	0.096	0.097	0.099
	0.05	0.056	0.053	0.072	0.056	0.054	0.059
	0.01	0.011	0.010	0.022	0.011	0.010	0.015
0.75	**0.10**	0.097	0.094	0.137	0.102	0.101	0.133
	0.05	0.047	0.046	0.081	0.054	0.053	0.085
	0.01	0.011	0.010	0.032	0.009	0.009	0.025

Note: Rejection rates of 2000 Monte Carlo experiments.

Table 2: Rejection rate of the null hypothesis using $a = b = 0$, model with $c = 1$ and $\phi = 1$

τ	*Size*	$n = 500$			$n = 1000$		
		FS-2SLS	**True f**	**Sparsity f**	**FS-2SLS**	**True f**	**Sparsity f**
0.25	**0.10**	0.100	0.104	0.101	0.104	0.105	0.108
	0.05	0.052	0.051	0.053	0.054	0.056	0.054
	0.01	0.014	0.015	0.016	0.008	0.009	0.010
0.50	**0.10**	0.100	0.104	0.109	0.096	0.095	0.095
	0.05	0.056	0.055	0.059	0.056	0.056	0.053
	0.01	0.011	0.011	0.010	0.011	0.013	0.012
0.75	**0.10**	0.097	0.103	0.101	0.102	0.104	0.100
	0.05	0.047	0.051	0.049	0.054	0.054	0.058
	0.01	0.011	0.009	0.013	0.009	0.010	0.012

Note: Rejection rates of 2000 Monte Carlo experiments.

An interesting feature in Tables 3 and 4 is that, in some cases, the weighting function may be used to evaluate the first-stage relevance of the IV, even when it is not the correct one. This is not a general result, however, but it illustrates the role of the density function as a weighting factor.

In order to explore this, we consider three examples in Appendix B closely related to the DGP used in the Monte Carlo experiments. When the instrument z is available we should be estimating the correct structural parameters and $f_{u_\tau}(0|d, x, z)$ where $u_\tau = y - Q_\tau(y|d, x, z)$. However, the case where, under the null, z is invalid would be equivalent to the case where there

Table 3: Rejection rate of the null hypothesis using $a = b = 0$, model with $c = 0$ and $\phi = 0$

τ	*Size*	$n = 500$			$n = 1000$		
		FS-2SLS	**True f**	**Sparsity f**	**FS-2SLS**	**True f**	**Sparsity f**
0.25	**0.10**	0.110	0.109	0.228	0.093	0.092	0.265
	0.05	0.057	0.054	0.163	0.054	0.052	0.200
	0.01	0.013	0.013	0.073	0.013	0.013	0.108
0.50	**0.10**	0.107	0.102	0.235	0.097	0.094	0.238
	0.05	0.059	0.055	0.171	0.049	0.049	0.175
	0.01	0.017	0.018	0.101	0.010	0.008	0.094
0.75	**0.10**	0.101	0.100	0.240	0.102	0.101	0.274
	0.05	0.049	0.048	0.171	0.054	0.051	0.198
	0.01	0.010	0.011	0.093	0.008	0.008	0.102

Note: Rejection rates of 2000 Monte Carlo experiments.

Table 4: Rejection rate of the null hypothesis using $a = b = 0$, model with $c = 1$ and $\phi = 0$

τ	*Size*	$n = 500$			$n = 1000$		
		FS-2SLS	**True f**	**Sparsity f**	**FS-2SLS**	**True f**	**Sparsity f**
0.25	**0.10**	0.110	0.114	0.107	0.093	0.097	0.101
	0.05	0.057	0.060	0.062	0.054	0.057	0.054
	0.01	0.013	0.012	0.016	0.013	0.013	0.010
0.50	**0.10**	0.107	0.107	0.108	0.097	0.097	0.094
	0.05	0.059	0.059	0.059	0.049	0.050	0.051
	0.01	0.017	0.018	0.020	0.010	0.009	0.009
0.75	**0.10**	0.101	0.100	0.108	0.102	0.109	0.105
	0.05	0.049	0.050	0.053	0.054	0.051	0.052
	0.01	0.010	0.010	0.009	0.008	0.009	0.011

Note: Rejection rates of 2000 Monte Carlo experiments.

are no instruments available. That is, we would not be able to solve the endogeneity in the second-stage. Note that, for this case, the density function that will be implicitly used is that of $u^*_\tau = y - Q_\tau(y|d,x)$. The examples in Appendix B compare $f_{u^*_\tau}(0|d,x)$ with $f_{u_\tau}(0|d,x,z)$.[4] The partial results suggest that if $f_{u_\tau}(0|d,x,z)$ and $f_{u^*_\tau}(0|d,x)$ are proportional to each other when they vary with d, we could implement the first-stage test under the null of all IV being invalid.

To analyze the empirical power of the test, we performed 2000 simulations only for the case with $n = 1000$ and we calculated 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. The results appear in Figures 1-4. For each figure we have two blocks, (i) and (ii), 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 $b = \{0, 0.10, ..., 0.90, 1\}$ such that z_1 has only a scale effect on d.

We first consider the case where the relation $y|(d,x)$ is a pure location model, that is, $c = 0$, and 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), scale-only first-stage, however, FS-2SLS and FS-IVQR (true density) have no power in detecting the effect of z_1 on d. The test with the estimated sparsity function rejects the null as b increases.

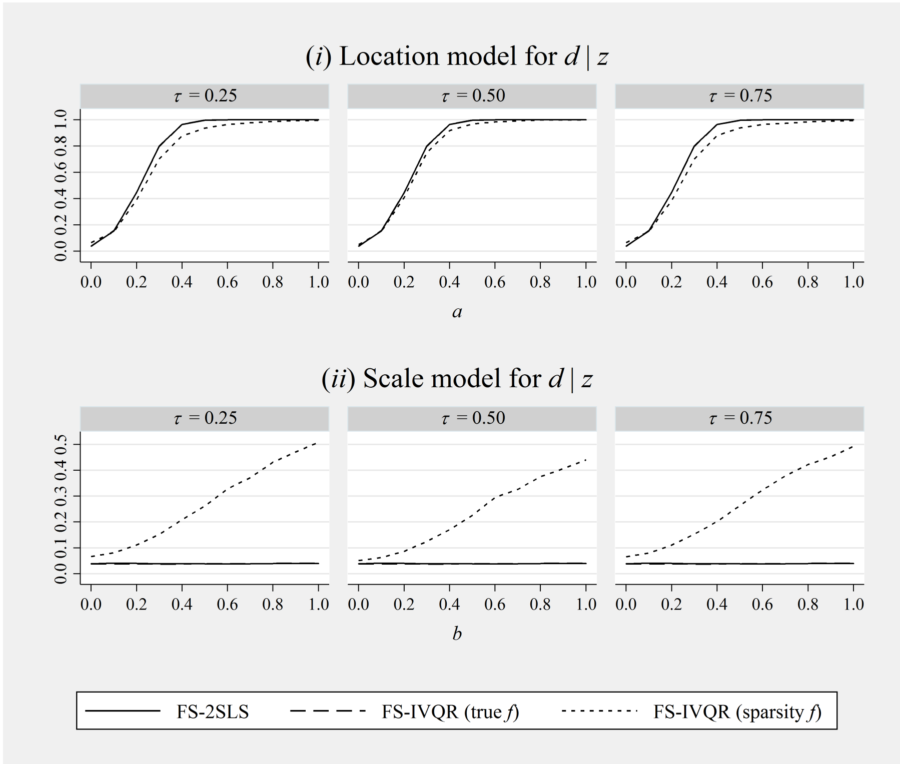
We now analyze the case of location-scale in $y|(d,x)$, i.e. $c = 1$, also with $\phi = 1$. The results of Figure 2 show that the FS-2SLS test is similar to the QRIV-based tests under pure location model for $d|z_1$ (block (i) of Figure 2). In particular, both using the true and the estimated sparsity function correctly rejects when a increases. However, the results of the FS-IVQR differ when we are in the presence of a pure-scale model for $d|z1$ (block (ii) of Figure 2). 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 theoretical and estimated) are able to correctly detect it.

Consider now the case where we have $c = 0$ and $\phi = 0$. The problems noted in the size tables are exacerbated here since it is not possible to identify the parameters of $Q_\tau(y|(d,x))$ via the IVQR, as shown in Figure 3. Interestingly, the density estimation using the sparsity function introduces some misspecification that allow us to evaluate the effect of z_1. Note however, that as noted in the size evaluation, this test is oversized, and therefore it cannot be used for valid inference.

Finally, consider the last case when $c = 1$ and $\phi = 0$ (Figure 4). The FS-IVQR tests work

[4] Let $\tilde{\alpha}$ and $\tilde{\beta}$ be the parameters that result from the estimation of the biased structural model without instruments, $Q_\tau(y|d,x) = d\tilde{\alpha} + x\tilde{\beta}$. Note that $u_\tau = y - Q_\tau(y|d,x,z) = y - d\alpha_0 - x\beta_0$ can be written as $y - d\tilde{\alpha} - x\tilde{\beta} - bias(d,x)$, where $bias(d,x) = d(\alpha_0 - \tilde{\alpha}) + x(\beta_0 - \tilde{\beta})$ such that $u_\tau = u^*_\tau - bias(d,x)$.

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



in this case. In both (i) and (ii) the tests detects 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 4 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 $c = 1$ and $\phi = 1$)

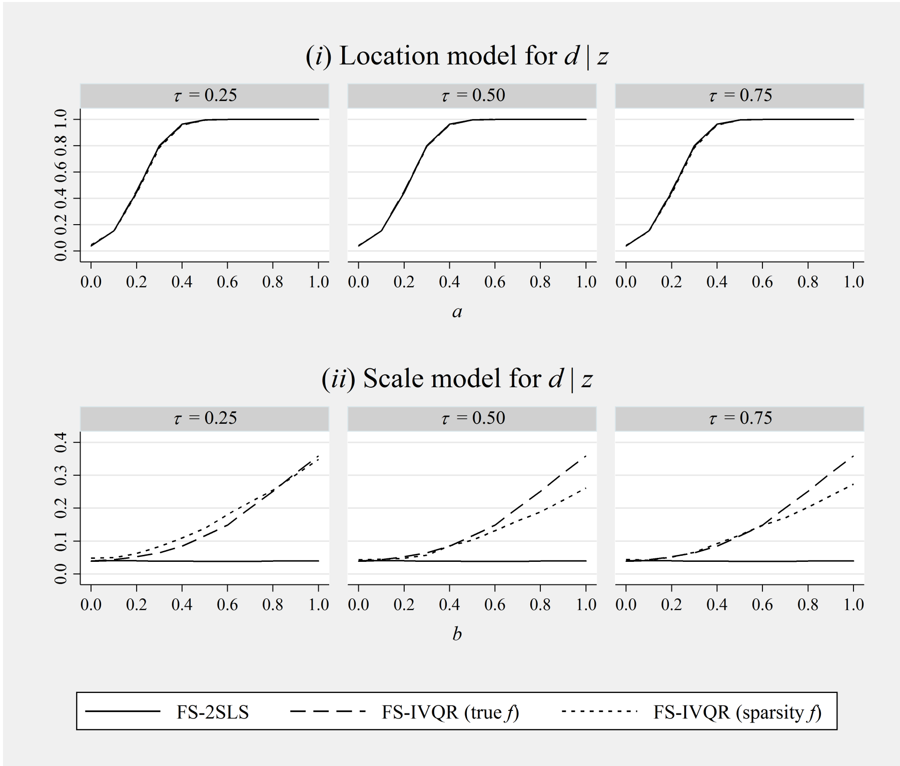


Figure 3: Power for $H_0 : \delta_1 = 0$ (model with $c = 0$ and $\phi = 0$)

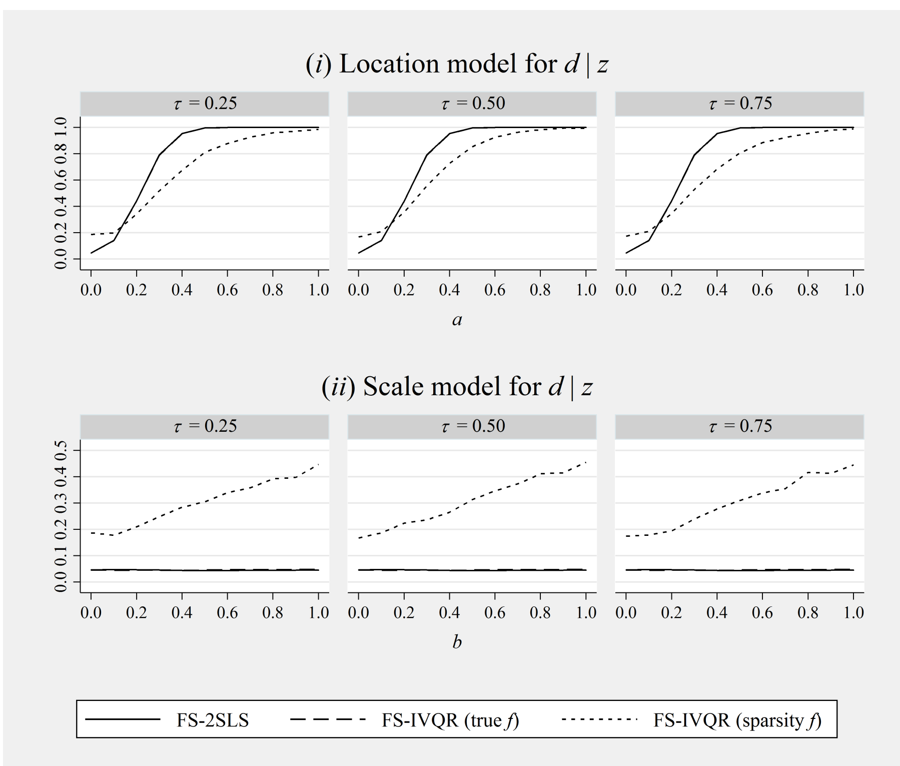
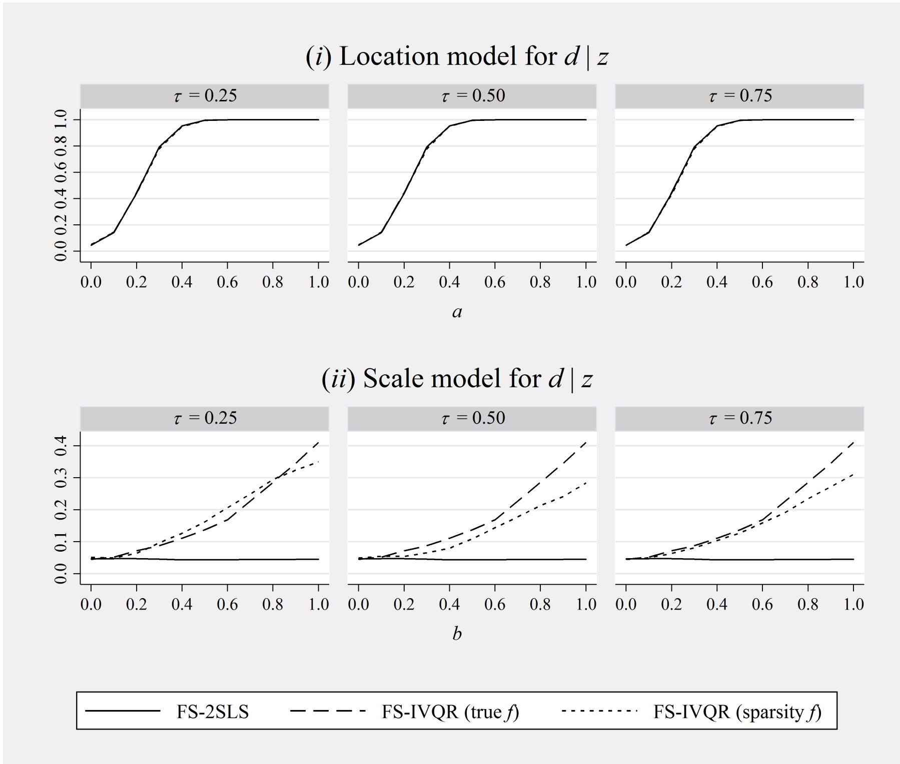


Figure 4: Power for for $H_0 : \delta_1 = 0$ (model with $c = 1$ and $\phi = 0$)



6 Empirical application: Card (1995) college proximity as an instrument for education

In this section we show an application of the proposed test in the estimation of a Mincer equation to estimate returns to schooling. The data used are from the paper of Card (1995) and correspond to 3010 individuals of the US National Longitudinal Survey of Young Men.[5] 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 analyzed by Angrist et al. (2006). Card (1995) proposes to implement an IV strategy using two measures of proximity to the university as external variables to the wage equation.

Table 5 shows the results of the first-stage to check if the IV are valid, together with the estimated second-stage results. The first column corresponds to the conditional mean model and the next ones are the regressions proposed for IVQR for $\tau \in \{0.25, 0.50, 0.75\}$. The results shows that the first instrument (lived near 2-year college in 1966) is not relevant for the low

[5]Downloaded from `http://davidcard.berkeley.edu/data_sets/proximity.zip`

Table 5: Returns to schooling (Card, 1995)

	2SLS	IV Quantile Regression		
		$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$
		First-stage estimates		
Lived Near 2-year College in 1966	0.123	0.0644	0.471***	0.154**
	(0.0774)	(0.129)	(0.0704)	(0.0709)
Lived Near 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	-0.000681	0.00457***	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)
		Second-stage estimates		
Education	0.157***	0.176***	0.268***	0.104
	(0.0524)	(0.0521)	(0.0271)	(0.0662)
Experience	0.119***	0.120***	0.180***	0.0932***
	(0.0227)	(0.0248)	(0.0140)	(0.0341)
Experience-Squared	-0.00236***	-0.00201***	-0.00337***	-0.00221***
	(0.000347)	(0.000347)	(0.000352)	(0.000438)
Black indicator	-0.123**	-0.110**	-0.00925	-0.148***
	(0.0520)	(0.0519)	(0.0342)	(0.0469)
Constant	3.237***	2.698***	1.400***	4.360***
	(0.883)	(0.870)	(0.466)	(1.119)
Observations	3,010	3,010	3,010	3,010

Source: Card (1995). 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.

quantiles and the mean but it is significant for middle and high quantiles. Also, note that although the second instrument (lived near 4-year college in 1966) rejects the null hypothesis for the conditional mean, this variable has different degree of significance across quantiles. In particular, this is for $\tau = 0.75$ where the instrument is relevant only at 10% significance. These results are very important since although the proximity to the university seems to be a strong instrument to identify the causal effect of education on the conditional mean, our test also indicates a certain limitation when the object of study is to evaluate the impact on the lower part of conditional distribution of wages. Therefore, this alerts for the quality of the asymptotic properties of the IVQR estimates in the presence of invalid instruments.

7 Conclusions

This paper proposes a first-stage model and inference procedures 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. Monte Carlo experiments clearly illustrate that one may encounter cases where the IV are not valid for the mean, but are still valid for some quantiles. The same issue appears in the empirical application.

The analysis may be extended in the following two directions. First, this approach can be used to identify local 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.

Appendix A: Proofs

Proof of Lemma 1. First, consider an estimator of the parameter μ using the true weighting matrix V as

$$V = \begin{bmatrix} f_1 & & \\ & \ddots & \\ & & f_n \end{bmatrix}, \tag{33}$$

that is given by the following

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

where $W = [X, Z]$. Replacing D by $(W\mu_0 + \varepsilon)$ in the definition of $\tilde{\mu}$ we have that

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

By the Slutsky's Theorem, the proof of the lemma requires showing that

$$\frac{W'VW}{n} \xrightarrow{p} \Omega_f, \tag{34}$$

and

$$\frac{W'V\varepsilon}{\sqrt{n}} \xrightarrow{d} N(0, \Omega_{f\sigma}). \tag{35}$$

To show (34), its left side has the (j,k) element given by

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

by the Law of Large Numbers and Assumption 2. To show (35), first note that

$$\mathrm{E}\left[W'V\varepsilon\right] = \mathrm{E}\left[W'V\mathrm{E}[\varepsilon|W]\right] = 0,$$

by Assumption 2. Furthermore, $W'V\varepsilon$ is a sum of i.i.d. random vectors $f_{\theta_0}(s_i) \cdot w_i \cdot \varepsilon_i$ with common covariance matrix having the (j,k) element

$$\begin{aligned} Cov\left(f_i w_{ij}\varepsilon_i, f_i w_{ik}\varepsilon_i\right) &= \mathrm{E}\left[f_i^2 w_{ij} w_{ik} \varepsilon_i^2\right] = \mathrm{E}\left[f_i^2 w_{ij} w_{ik} \mathrm{E}[\varepsilon_i^2|w_i]\right] \\ &= \mathrm{E}\left[f_i^2 w_{ij} w_{ik} \sigma_i^2\right]. \end{aligned}$$

Thus, each vector $f_i \cdot w_i \cdot \varepsilon_i$ has covariance matrix $\Omega_{f\sigma}$. Therefore, by the Multivariate Central Limit Theorem, (35) holds.

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

To establish that consider the estimator with the estimated weights as following

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

such that

$$\sqrt{n}(\hat{\mu} - \mu) = \left(\frac{W'\hat{V}W}{n}\right)\frac{W'\hat{V}\varepsilon}{\sqrt{n}}. \tag{36}$$

First, we show that

$$\frac{W'\hat{V}\varepsilon}{\sqrt{n}} - \frac{W'V\varepsilon}{\sqrt{n}} \xrightarrow{p} 0. \tag{37}$$

Note that

$$\frac{W'(\hat{V} - V)\varepsilon}{\sqrt{n}} = n^{-1/2}\sum_{i=1}^{n} w_i\varepsilon_i \left(\hat{f}_i - f_i\right). \tag{38}$$

We want to show that the right hand side of (38) is $o_p(1)$. Using the sparsity function estimator in (22) along with some calculations, we have that

$$\hat{f}_i = f_i + \frac{2h_n}{f_i^2}s_i(\hat{\theta} - \theta) + o_p((nh^2)^{-2/3}).$$

We refer the reader to Ota, Kato, and Hara (2019) for details on the remainder term.

Hence, using the previous equation, the jth component of the right hand side of equation (38) can be written as

$$\sqrt{n}(\hat{\theta}_j - \theta_{0,j})2h_n\frac{1}{n}\sum_{i=1}^{n}\frac{1}{f_i^2}w_{ij}\varepsilon_i.$$

The first factor $\sqrt{n}(\hat{\theta}_j - \theta_{0,j}) = o_p(1)$ by Assumption 1 and CH. Moreover, note that the average of the i.i.d. variables $f_i^{-2}w_i\varepsilon_i$ obeys the Law of Large Numbers by the moment restrictions in Assumption 2, and the result follows.

Next, we show that

$$\frac{W'\hat{V}W}{n} - \frac{W'VW}{n} \xrightarrow{p} 0, \tag{39}$$

which follows from the same argument as above.

The convergences (37) and (39) are enough to show that the right-hand side of (36) satisfies

$$\left(\frac{W'\hat{V}W}{n}\right)\frac{W'\hat{V}\varepsilon}{\sqrt{n}} - \left(\frac{W'VW}{n}\right)\frac{W'V\varepsilon}{\sqrt{n}} \xrightarrow{p} 0$$

just by making simple use of the equality

$$\hat{a}\hat{b} - ab = \hat{a}(\hat{b} - b) + (\hat{a} - a)b.$$

Finally, Slutsky's theorem yields the result. □

Proof of Proposition 1. The proof of this result is simple. It follows from observing that by Lemma 1,

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

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,

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

□

Appendix B: Examples of weighting factors

1. Location model

Consider a pure location model, using two equations

$$y = d + u,$$

$$d = az + v,$$

with $(u, v) \sim N(0, 0, 1, 1, \rho)$ a bivariate normal with zero mean, unit variance and correlation parameter ρ and $z \sim N(0, 1)$. Then, it follows that $d \sim N(0, 1 + a^2)$ and $y \sim N(0, 2 + a^2 + 2\rho)$.

Consider now the model where we condition on both (d, z). For this case, $u|d, z \sim N(\rho v, (1-\rho^2))$ by the marginal of the bivariate normal density. Then,

$$Q_\tau(u|d, z) = \rho v + \sqrt{1 - \rho^2}\Phi^{-1}(\tau).$$

Then, $u_\tau = y - Q_\tau(y|d, z) = u - Q_\tau(u|d, z)$. Note that $E(u_\tau|d, z) = E(u_\tau|d, z) - Q_\tau(u|d, z) = -\sqrt{1 - \rho^2}\Phi^{-1}(\tau)$. Thus, the density is

$$f_{u_\tau}(U|d, z) = \frac{1}{\sqrt{1 - \rho^2}}\phi\left(\frac{U + \sqrt{1 - \rho^2}\Phi^{-1}(\tau)}{\sqrt{1 - \rho^2}}\right),$$

where $\phi()$ is the density function of a standard normal. If we evaluate it at 0,

$$f_{u_\tau}(0|d, z) = \frac{1}{\sqrt{1 - \rho^2}}\phi\left(\Phi^{-1}(\tau)\right).$$

Now consider the joint density of $(u, d) \sim N(0, 0, 1, 1 + a^2, \kappa)$, where $\kappa = \frac{\rho}{\sqrt{1+a^2}}$. Then, it follows that $u|d \sim N(E(u|d), Var(u|d))$, where $E(u|d) = \kappa d$ and $Var(u|d) = (1 - \kappa^2)$.

As such, we can obtain the quantiles of interest,

$$Q_\tau(y|d) = d + \kappa d + \Phi^{-1}(\tau)(1 - \kappa^2)^{1/2}.$$

Note that without endogeneity, i.e. $\rho = 0$, then $\kappa = 0$, and the correct τ-quantile model should be

$$Q_\tau(y|d, \rho = 0) = d + \Phi^{-1}(\tau).$$

Now, $u_\tau^* = y - Q_\tau(y|d) = d + u - (d + \kappa d + \Phi^{-1}(\tau)(1 - \kappa^2)^{1/2}) = u - \kappa d - \Phi^{-1}(\tau)(1 - \kappa^2)^{1/2}$. Then, $E(u_\tau^*|d) = -\Phi^{-1}(\tau)(1 - \kappa^2)^{1/2}$, and $Var(u_\tau^*|d) = Var(u|d) = (1 - \kappa^2)$.

Then,

$$f_{u_\tau^*}(U|d) = \frac{1}{\sqrt{(1 - \kappa^2)}}\phi\left(\frac{U - E(u_\tau^*|d)}{\sqrt{Var(u_\tau^*|d)}}\right),$$

such that,

$$f_{u_\tau^*}(0|d) = \frac{1}{\sqrt{(1 - \kappa^2)}}\phi\left(\Phi^{-1}(\tau)\right).$$

In all cases, $f_{u_\tau^*}(0|d)$ and $f_{u_\tau}(0|d, z)$ are constant that do not change with d or z. It is interesting to evaluate when $a = 0$, such that $(1 - \kappa^2) = (1 - \rho^2)$. Note that in this case, $f_{u_\tau^*}(0|d) = f_{u_\tau}(0|d, z)$.

2. Location-scale model 1

Now consider a location-scale model of the form

$$y = d + (1 + cd)u,$$

$$d = az + v,$$

where a and c are parameters. As in the previous case $(u, v) \sim N(0, 0, 1, 1, \rho)$. Then, $u|(d, z) \sim u|v \sim N(\rho v, 1 - \rho^2)$. Thus, $Q_\tau(u|d, z) = \rho v + \sqrt{1 - \rho^2}\Phi^{-1}(\tau)$. Note that it does not depend on z.

In this case, $Q_\tau(y|d, z) = d + (1 + cd)Q_\tau(u|d, z)$, and then, $u_\tau = y - Q_\tau(y|d, z) = (1 + cd)(u - Q_\tau(u|d, z))$.

As such, we can obtain,

$$f_{u_\tau}(U|d, z) = \frac{1}{|1 + cd|\sqrt{1 - \rho^2}}\phi\left(\frac{U + (1 + cd)\sqrt{1 - \rho^2}\Phi^{-1}(\tau)}{(1 + cd)\sqrt{1 - \rho^2}}\right).$$

If we evaluate it at 0,

$$f_{u_\tau}(0|d,z) = \frac{1}{|1+cd|\sqrt{1-\rho^2}}\phi\left(\Phi^{-1}(\tau)\right).$$

Note that this depends d, and then, the weights are not uniform.

Now, consider the of $u|d$. Consider first the joint distribution of $(u,d) \sim N(0,0,1,1+a^2,\kappa)$ where $\kappa = \rho/\sqrt{1+a^2}$. Now, $u|d \sim N(\kappa d, (1-\kappa^2))$, then $E(u|d) = \kappa d$ and $Var(u|d) = (1-\kappa^2)$.

For this case let $u_\tau^* = y - Q_\tau(y|d) = d + (1+cd)u - d - (1+cd)Q_\tau(u|d) = (1+cd)(u - Q_\tau(u|d))$. Since $u|d$ is Gaussian then $(1+cd)(u - \kappa d - \Phi^{-1}(\tau)(1-\kappa^2)^{1/2})$. Then, $E(u_\tau^*|d) = (1+cd)(-\Phi^{-1}(\tau)(1-\kappa^2)^{1/2})$ and $Var(u_\tau^*|d) = (1+cd)^2(1-\kappa^2)$. As such, we can obtain,

$$f_{u_\tau^*}(U|d) = \frac{1}{|1+cd|\sqrt{1-\kappa^2}}\phi\left(\frac{U + (1+cd)(1-\kappa^2)^{1/2}\Phi^{-1}(\tau)}{(1+cd)(1-\kappa^2)^{1/2}}\right).$$

If we evaluate it at 0,

$$f_{u_\tau^*}(0|d) = \frac{1}{|1+cd|\sqrt{1-\kappa^2}}\phi\left(\Phi^{-1}(\tau)\right).$$

Note that both $f_{u_\tau}(0|d,z)$ and $f_{u_\tau^*}(0|d)$ share the same relationship with d. In fact, the weighting procedure will be equivalent, as they are proportional to each other.

3. Location-scale model 2

Now consider a location-scale model where both the first and second stage are affected in the variance component,

$$y = d + (1+cd)u,$$

$$d = az + (1+bz)v,$$

where a, b, and c are parameters. As in the previous case $(u,v) \sim N(0,0,1,1,\rho)$. Define $w = (1+bz)v$ and note that $(u,w|z) \sim N(0,0,1,(1+bz)^2,\rho)$. Then, $u|d,z \sim u|w,z \sim N(\rho v, 1-\rho^2)$. Thus, $Q_\tau(u|d,z) = \rho v + \sqrt{1-\rho^2}\Phi^{-1}(\tau)$. Note that it does not depend on b.

In this case, $Q_\tau(y|d,z) = d + (1+cd)Q_\tau(u|d,z)$, and then, $u_\tau = y - Q_\tau(y|d,z) = (1+cd)(u - Q_\tau(u|d,z))$.

As such, we can obtain,

$$f_{u_\tau}(U|d,z) = \frac{1}{|1+cd|\sqrt{1-\rho^2}}\phi\left(\frac{U + (1+cd)\sqrt{1-\rho^2}\Phi^{-1}(\tau)}{(1+cd)\sqrt{1-\rho^2}}\right).$$

If we evaluate it at 0,

$$f_{u_\tau}(0|d,z) = \frac{1}{|1+cd|\sqrt{1-\rho^2}}\phi\left(\Phi^{-1}(\tau)\right).$$

Note that this depends d, and then, the weights are not uniform.

Now, it is not standard to obtain the distribution of $u|d$. To exemplify this, suppose $z = \{0, 1\}$ is a simple binary variable with $p = Pr(z = 1)$ and independent of (u, v). Then, the joint density is $f(u, v, z) = \phi_\rho(u, v)p^z(1-p)^{1-z}$ and using the Jacobian transformation we obtain:

$$f(u, d, z) = \frac{1}{|1+bz|}\phi_\rho\left(u, \frac{d-az}{1+bz}\right)p^z(1-p)^{1-z}$$

Therefore,

$$f(u, d) = \phi_\rho(u, d)(1-p) + \frac{1}{|1+b|}\phi_\rho\left(u, \frac{d-a}{1+b}\right)p$$

and

$$f(d) = \phi(d)(1-p) + \frac{1}{|1+b|}\phi\left(\frac{d-a}{1+b}\right)p$$

Putting all that together, the conditional density is

$$f(u|d) = \frac{\phi_\rho(u, d)(1-p) + \frac{1}{|1+b|}\phi_\rho\left(u, \frac{d-a}{1+b}\right)p}{\phi(d)(1-p) + \frac{1}{|1+b|}\phi\left(\frac{d-a}{1+b}\right)p}.$$

If we assume that $p = \frac{|1+b|}{1+|1+b|}$ this expression simplifies to

$$f(u|d) = \frac{\phi_\rho(u, d) + \phi_\rho\left(u, \frac{d-a}{1+b}\right)}{\phi(d) + \phi\left(\frac{d-a}{1+b}\right)}.$$

We can rewrite this as a function of standard normal densities noting that $\phi_\rho(u, d) = \phi_\rho(u|d)\phi(d)$ with $\phi_\rho(u|d) = \frac{1}{\sqrt{1-\rho^2}}\phi\left(\frac{u-\rho d}{\sqrt{1-\rho^2}}\right)$, then

$$f(u|d) = \frac{1}{\sqrt{1-\rho^2}}\phi\left(\frac{u-\rho d}{\sqrt{1-\rho^2}}\right)\omega(d) + \frac{1}{\sqrt{1-\rho^2}}\phi\left(\frac{u-\rho\frac{d-a}{1+b}}{\sqrt{1-\rho^2}}\right)(1-\omega(d)),$$

where $\omega(d) = \frac{\phi(d)}{\phi(d)+\phi\left(\frac{d-a}{1+b}\right)}$. Therefore, conditional on d this density is a Gaussian mixture of two distributions with different means. Two particular cases are: (i) $\rho = 0$ (exogeneity) where $f(u|d) = \phi(u)$; (ii) $a = b = 0$ (d and z unrelated) which reduces to $f(u|d) = \phi_\rho(u|d)$. Obviously, in the rest of the cases $Q_\tau(u|d)$ does not have an explicit analytical solution and therefore neither $u^*_\tau = y - Q_\tau(y|d) = (1+cd)(u - Q_\tau(u|d))$.

The interesting feature to notice is that in all cases, the distribution of u^*_τ depends basically on d, and $(1+cd)$ should be used to standardize its density function in a similar way to u_τ.

References

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

Angrist, J., Chernozhukov, V. and Fernández-Val, I. (2006). Quantile regression under misspecification, with an application to the U.S. wage structure. *Econometrica, 74*, 539-563.

Card, D. (1995). Using Geographic Variation in College Proximity to Estimate the Return to Schooling. in Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp, ed, by Louis N. Christofides, E. Kenneth Grant, and Robert Swidinsky. Toronto: University of Toronto Press, 201-222.

Chen, L-A. and Portnoy, S. Two-stage regression quantiles and two-stage trimmed least squares estimators for structural equation models. *Communication in Statistics, Theory Methods, 25*, 1005–1032.

Chernozhukov, V. and Hansen, C. (2005). An IV model of quantile treatment effects. *Econometrica, 73*, 245–261.

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

Chernozhukov, V. and Hansen, C. (2008). Instrumental variable quantile regression: A robust inference approach. *Journal of Econometrics, 142*, 379–398.

Chernozhukov, V., Hansen, C. and Jansson, M. (2009). Finite sample inference for quantile regression models *Journal of Econometrics, 152*, 93–103.

Chernozhukov, V., Hansen, C. and Wuthrich, C. (2020). Instrumental variable quantile regression *Handbook of Quantile Regression,* https://arxiv.org/abs/2009.00436

Chesher, A. (2003). Identification in nonseparable models. *Econometrica, 71*, 1405–1441.

de Castro, L., Galvao, A.F., Kaplan, D.M. and Liu, X. (2019) Smoothed GMM for quantile models. *Journal of Econometrics, 213*, 121–144.

Galvao, A. (2011). Quantile regression for dynamic panel data with fixed effects. *Journal of Econometrics, 164*, 142–157.

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

Hall, P. and Sheather, S.J. (1988). On the distribution of a studentized quantile. *Journal of the Royal Statistical Society: Series B, 50 (3)*, 381–391.

Hendricks, W. and Koenker, R. (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*, 118–138.

Kaplan, D.M. and Sun, Y. (2017). Smoothed estimating equations for instrumental variables quantile regression. *Econometric Theory, 33*, 105–157.

Kim, T-H and Muller, C. (2004). Two-stage quantile regression when the first stage is based on quantile regression. *Econometrics Journal, 7*, 218–231.

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

Lee, S. (2007). Endogeneity in quantile regression models: A control function approach. *Journal of Econometrics, 101*, 1131–1158.

Lee, D. S., McCrary, J., Moreira, M. J. and Porter, J. (2020). Valid t-ratio inference for IV. Draft version. https://arxiv.org/abs/2010.05058

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

Ota, H., Kato, K. and Hara, S. Quantile regression approach to conditional mode estimation. *Electronic Journal of Statistics, 13*, 3120–3160.

Powell, J., (1983). The asymptotic normality of two-stage least absolute deviations estimators. *Econometrica, 51*, 1569–1576.

Sanderson, E. and Windmeijer, F. (2016). A weak instrument F-test in linear IV models with multiple endogenous variables. *Journal of Econometrics, 190*, 212-221.

Staiger, D. and Stock, J.H. (1997). Instrumental variables regression with weak instruments. *Econometrica, 65*, 557-566.

Stock, J. and Yogo, M. (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.

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