`ivqregress`: Stata Command for Instrumental Variables Quantile Regression

Di Liu

StataCorp

March 31, 2022

Abstract

We describe the design of `ivqregress`, a Stata command to estimate the linear quantile regression model with endogenous variables using the instrumental variables approach. In particular, `ivqregress` implements the inverse quantile regression estimator in Chernozhukov and Hansen (2006) and the smoothed estimating equation estimator in Kaplan and Sun (2017). We also give details about the design of a suite of post-estimation tools to visualize, make inference, and diagnose the instrumental variable quantile regression model.

Contents

1 Introduction

In empirical applications, we are usually interested in the causal effects of some covariate on the outcome variable. The traditional linear regression model is an excellent way to model how the covariate affects the outcome's conditional mean. However, sometimes we would like to study features of the outcome distribution different than the mean to have a full picture of the causal effects of covariates. For example, a policy maker may be more interested in how a summer job training program affects the income's lower quantile instead of just its mean.

Quantile regression in Koenker and Bassett (1978) can help us grasp a better picture than the regular linear regression by estimating the causal effects on different quantiles of the outcome's conditional distribution. For a discussion on quantile regression, see `qreg`.

In practice, some covariates of interests are often endogenous due to various reasons such as self-selection, omission of some relevant variable, and measurement error. For example, suppose we are interested in how the participation of the 401(k) program affects the net wealth. However, participation in the 401(k) program is endogenous because the people who do and do not participate may have different saving preferences, which will affect the net wealth growth.

Endogenous covariates make quantile regression estimates inconsistent, as is the case for linear regression model. Analogous to the instrumental variable least square estimator, there are instrumental variable quantile regression estimator to compute the different quantiles of casual effects consistently. For a discussion of instrumental variables estimation, see `ivregress`.

`ivqregress` fits a quantile regression model with endogeneity using two estimators: the inverse quantile regression estimator proposed in Chernozhukov and Hansen (2006) and the smoothed estimating equation estimator outlined in Kaplan and Sun (2017). Intuitively, `ivqregress` can be thought of as the `ivregress` version of the `qreg`, although the underlying estimators are not as straightforward as the two-stage least-squares estimator.

This paper describes the design of the `ivqregress` and a suite of post-estimation tools to estimate, visualize, make the inference, and diagnose the instrumental variable quantile regression model. In particular, the Stata commands in the IVQR toolbox can be grouped into the following categories.

Estimation

- `ivqregress iqr` estimates the IVQR model by the inverse quantile regression (IQR) estimator proposed in Chernozhukov and Hansen (2006).
- `ivqregress smooth` estimates the IVQR model by the smoothed estimation equation (SEE) estimator proposed in Kaplan and Sun (2017).

Visualization

- `estat coefplot` visualizes how the treatment effects vary at different quantiles of the outcome.
- `marginsplot` plots the potential outcome's conditional quantile function.

Inference

- `estat endogeffects` makes inference of particular interest in the context of IVQR model. In particular, `estat endogeffects` test four hypotheses: 1. the quantile treatment effects are zero; 2. the quantile treatment effects are constant across different quantiles; 3. the quantile treatment effects are unambiguously beneficial; 4. the treatment is exogenous.
- `estat dualci` provides confidence interval robust to weak instruments for the endogenous treatment effects. It is allowed only after `ivqregress iqr`.
- `test` and `testnl` infer classical linear and nonlinear hypotheses after estimation.
- `margins` helps to compute the potential outcome's conditional quantile function.

Diagnosis

- `estat waldplot` helps diagnosis the convergence of the inverse quantile regression estimator (`ivqregress iqr`). In particular, `estat waldplot` visualizes the optimization process during the computation in `ivqregress iqr` and shows if the searching domain contains the true value of the parameter with a predefined probability level.

We organize this paper as follows. Section 2 describes the instrumental variable quantile regression model and the primary moment condition for estimation. Section 3 reviews the existing estimators, discusses their strength and weakness and identify two estimators implemented in `ivqregress`. In particular, `ivqregress` implements the inverse quantile regression estimator and the smoothed estimating equation estimator, discussed in Sections 4 and 5, respectively. Section 6 outlines the general inference approach of particular interest in the IV quantile regression model. Section 7 describes the syntax of `ivqregress`. Section 8 describes the post-estimation of `ivqregress`. Finally, section 9 illustrates the use of `ivqregress` and its post-estimation through some examples.

2 The instrumental variable quantile regression model

2.1 The model

The general instrumental variables quantile regression model was first proposed by Chernozhukov and Hansen (2005). `ivqregress` is based on the linear IVQR model described in Chernozhukov and Hansen (2006) and Chernozhukov and Hansen (2008).

To simplify notation, we use the capital letter to denote the random variable and a lowercase letter to denote the random variable's actual value. We use the bold letter to represent a vector. All the vectors are assumed to be column vectors.

The linear IVQR model can be written in the form of "random coefficients" model

$$Y = \mathbf{D}'\boldsymbol{\alpha}(U_{\mathbf{D}}) + \mathbf{X}'\boldsymbol{\beta}(U_{\mathbf{D}}) \quad \text{where } U_{\mathbf{D}}|\mathbf{X},\mathbf{Z} \sim Uniform(0,1) \tag{1}$$

$$\mathbf{D} = \delta(\mathbf{X},\mathbf{Z},V) \quad \text{where } V \text{ statistically depends on } U_{\mathbf{D}} \tag{2}$$

$$\tau \rightarrow \mathbf{D}'\boldsymbol{\alpha}(\tau) + \mathbf{X}'\boldsymbol{\beta}(\tau) \quad \text{is strictly increasing in } \tau \tag{3}$$

where

- Y is a scalar outcome variable,
- $U_{\mathbf{D}}$ is a scalar random variable that charaterizes the heterogeneity of the outcome and captures all the unobservables in the outcome Equation 1,
- $\mathbf{D}$ is a vector of endogenous variables that statistically depend on $U_{\mathbf{D}}$,
- $\mathbf{X}$ is a vector of exogenous variables that are independent of $U_{\mathbf{D}}$,
- $\boldsymbol{\alpha}()$ and $\boldsymbol{\beta}()$ are random coefficient vectors that depend on $U_{\mathbf{D}}$,
- the endogenous variables $\mathbf{D}$ are dermined via Equation 2,
- $\mathbf{Z}$ is a vector of instrumental variables that are independent of $U_{\mathbf{D}}$ but correlated with $\mathbf{D}$,
- V is a scalar unobserved random variable that impacts $\mathbf{D}$ and also correlated with $U_{\mathbf{D}}$, and
- the observable variables are $\{Y_i, \mathbf{X}_i, \mathbf{D}_i, \mathbf{Z}_i\}_{i=1}^n$ with a sample of size n.

There are two objectives of the analysis.

1. Estimate the conditional quantile function of the latent potential outcome $Y_{\mathbf{d}}$ when fixing $\mathbf{D} = \mathbf{d}$ and conditional on $\mathbf{X}$. More precisely, the potential outcome $Y_{\mathbf{d}} = \mathbf{d}'\alpha(U_{\mathbf{d}}) + \mathbf{X}'\boldsymbol{\beta}(U_{\mathbf{d}})$ with $U_{\mathbf{d}}$ as a scalar random variable from uniform distribution conditional on $\mathbf{X}$ (note that $\mathbf{d}$ are treated as constant here). The conditional quantile function of $Y_{\mathbf{d}}$ can be written as

$$S_{Y_{\mathbf{d}}}(Y_{\mathbf{d}}|\tau, \mathbf{X}, \mathbf{d}) = \mathbf{d}'\boldsymbol{\alpha}(\tau) + \mathbf{X}'\boldsymbol{\beta}(\tau) \tag{4}$$

   Notice that $S_{Y_{\mathbf{d}}}()$ is generally different from the conditional quantile function for the observed outcome Y because $\mathbf{D}$ are endogenous. $S_{Y_{\mathbf{d}}}()$ is also referred to as structural quantile equation (SQE).

2. Estimate the quantile treatment effects of $\mathbf{D}$. Suppose $\mathbf{D}$ is a binary variable that can only take 0 and 1, The quantile treatment effects (QTE) are defined as

$$S_{Y_1}(Y_1|\tau, \mathbf{X}, 1) - S_{Y_0}(Y_0|\tau, \mathbf{X}, 0) = \boldsymbol{\alpha}(\tau) \tag{5}$$

   Or, if $\mathbf{D}$ is continuous, QTE is defined as

$$\frac{\partial S_{Y_{\mathbf{d}}}()}{\partial \mathbf{d}} = \boldsymbol{\alpha}(\tau) \tag{6}$$

   The linear-in-parameter assumption greatly simplifies the estimation of QTE, which is $\boldsymbol{\alpha}(\tau)$ in both discrete and continuous case.

Suppose we estimate the function $S_{Y_{\mathbf{d}}}()$ at different values of τ spanned between 0 and 1. In that case, we can have a fuller picture of the conditional distribution of $Y_{\mathbf{d}}$ than just estimating the mean of the distribution. In addition, the estimates for $\boldsymbol{\alpha}(\tau)$ at different values of τ reveal how the treatment effects vary at different conditional quantile indexes.

2.2 Main assumptions and moment condition

The primary condition in the IVQR model (Chernozhukov and Hansen (2005)) is

ASSUMPTION 1. *Consider a probability space* (Ω, F, P) *and the set of potential outcome variables* $(Y_{\mathbf{d}}, d \in \mathbb{D})$*, endogenous variables* $\mathbf{D}$*, exogenous covariates* $\mathbf{X}$*, and instruments* $\mathbf{Z}$*. The following conditions hold:*

A1 *(Potential outcomes) Conditional on* $\mathbf{X} = \mathbf{x}$*, for each* $\mathbf{d}$*,* $Y_{\mathbf{d}} = q(\mathbf{d}, \mathbf{x}, U_{\mathbf{d}})$*, where* $q(\mathbf{d}, \mathbf{x}, \tau)$ *is increasing in* τ *and* $U_{\mathbf{d}} \sim U(0, 1)$*.*

A2 *(Independence) Conditional on* $\mathbf{X} = \mathbf{x}$*,* $U_{\mathbf{d}}$ *are independent of* $\mathbf{Z}$*.*

A3 *(Selection)* $\mathbf{D} = \delta(\mathbf{Z}, \mathbf{X}, V)$ *for some unknown function* $\delta(\cdot)$ *and random vector* V*.*

A4 *(Rank similarity) Conditional on* $(\mathbf{X}, \mathbf{Z}, V)$*,* $\{U_{\mathbf{d}}\}$ *are identically distributed.*

A5 *(Observables) Observed variables consist of* $Y = q(\mathbf{D}, \mathbf{X}, U_{\mathbf{D}})$*,* $\mathbf{D}$*,* $\mathbf{X}$*, and* $\mathbf{Z}$*.*

In the linear IVQR model, $q(\mathbf{D}, \mathbf{X}, U_{\mathbf{D}}) = \mathbf{D}'\boldsymbol{\alpha}(U_{\mathbf{D}}) + \mathbf{X}'\boldsymbol{\beta}(U_{\mathbf{D}})$ as specified in Equation 1. The following is the main implication of Assumption A1-A5.

THEOREM 1 (Main implications of IVQR model)**.** *Suppose conditions A1-A5 hold. Then for all* $\tau \in (0,1)$*, a.s.*

$$P(Y \leq q(\mathbf{D}, \mathbf{X}, U_{\mathbf{D}}) | \mathbf{X}, \mathbf{Z}) = \tau \tag{7}$$

and $U_{\mathbf{D}} \sim U(0,1)$ *conditional on* $\mathbf{X}$ *and* $\mathbf{Z}$*.*

For the proof of Theorem 1 and a discussion on the rank similarity, see Section A.1 in Appendix. Theorem 1 also implies the following unconditional moment condition that can be used to estimate the IVQR model. Namely, Equation 7 together with the linear-in-parameter assumptions imply

$$\mathbf{E}\left[\left(\tau - \mathbb{1}(Y - \mathbf{D}'\boldsymbol{\alpha} - \mathbf{X}'\boldsymbol{\beta} \leq 0)\right) \boldsymbol{\Psi}(\mathbf{X}, \mathbf{Z})\right] = 0 \tag{8}$$

where $\mathbb{1}()$ is the indicator function, $\boldsymbol{\Psi}(\mathbf{X}, \mathbf{Z})$ is some transformation of $\mathbf{X}$ and $\mathbf{Z}$. In practice, $\boldsymbol{\Psi}(\mathbf{X}, \mathbf{Z}) = (\boldsymbol{\Phi}(\mathbf{X}, \mathbf{Z})', \mathbf{x}')'$, and $\boldsymbol{\Phi}(\mathbf{X}, \mathbf{Z})$ is the projection of $\mathbf{D}$ into the linear space spanned by $\mathbf{X}$ and $\mathbf{Z}$.

The empirical version of Equation 8 is

$$\frac{1}{n}\sum_{i=1}^{n}\left[\left(\tau - \mathbb{1}(Y_i - \mathbf{D}_i'\boldsymbol{\alpha} - \mathbf{X}_i'\boldsymbol{\beta} \leq 0)\right) \boldsymbol{\Psi}(\mathbf{X}_i, \mathbf{Z}_i)\right] = 0 \tag{9}$$

Equation 8 and 9 are the foundation for different estimation strategies, that will be discussed in Sections 3, 4, 5, and A.3.

3 Reviews of estimation approaches

3.1 Nonconvex and nonsmooth GMM objective function

To estimate the linear IV quantile regression model, we need to solve the moment condition specified in Equation 8. One necessary condition for identification is that $dim(\Psi(\mathbf{X}, \mathbf{Z})) > k_{\mathbf{X}} + k_{\mathbf{D}}$ with $dim(\mathbf{X}) = k_x$ and $dim(\mathbf{D}) = k_d$.

For $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta})$ and $\{\mathbf{W}_i\}_{i=1}^n = \{(Y_i, \mathbf{D}_i', \mathbf{X}_i', \mathbf{Z}_i')'\}_{i=1}^n$, let

$$\mathbf{g}_\tau(\mathbf{W}_i, \boldsymbol{\theta}) = ((\tau - \mathbb{1}(Y_i - \mathbf{D}_i'\boldsymbol{\alpha} - \mathbf{X}_i'\boldsymbol{\beta} \leq 0))\, \boldsymbol{\Psi}(\mathbf{X}_i, \mathbf{Z}_i) \tag{10}$$

The empirical moment condition specified in Equation 9 can be written as

$$\hat{\mathbf{g}}_N(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^n \mathbf{g}_\tau(\mathbf{W}_i, \boldsymbol{\theta}) \tag{11}$$

We can estimate $\boldsymbol{\theta}_0$ by GMM as

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\Theta} n\hat{\mathbf{g}}_N'\boldsymbol{\Omega}_N\hat{\mathbf{g}}_N \tag{12}$$

where $\boldsymbol{\Omega}_N$ is the GMM weighting matrix and it is usually set to be

$$\boldsymbol{\Omega}_N = \left(\tau(1-\tau)\frac{1}{n}\sum_{i=1}^n \boldsymbol{\Psi}_i\boldsymbol{\Psi}_i'\right)^{-1} \tag{13}$$

Due to the indicator function in Equation 10, the GMM objective function is nonsmooth and nonconvex. See Figure 3.1 for an illustration.

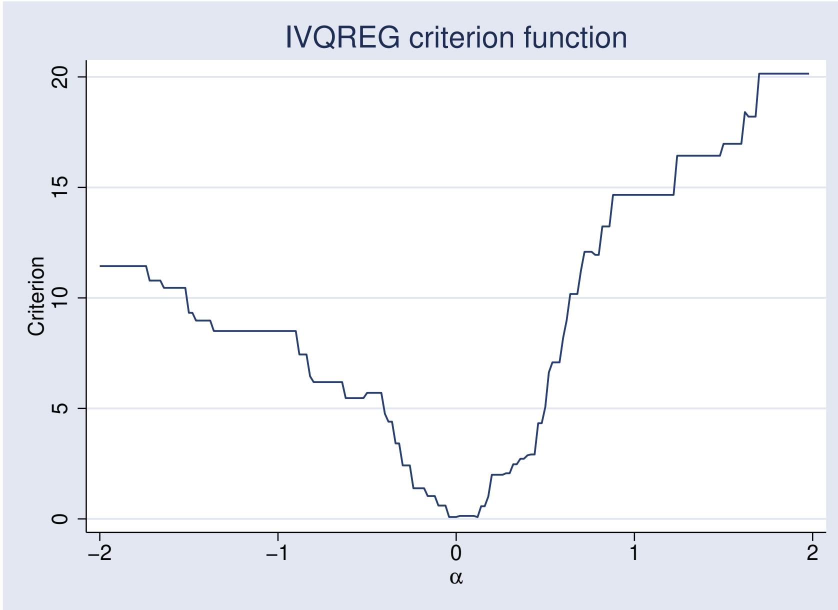


Figure 1: IVQREG GMM criterion function

For simplicity, the outcome variable only depends on one endogenous variable D, and the true value of α equals to zero. We plot the GMM criterion function at different values of α. While the criterion function is globally minimized at zero, the plot is flat in certain regions and exhibits several local minimums. Thus, it is computationally difficult to solve the optimization problem using the moment condition in Equation 9.

3.2 Different approaches

In literature, there are several estimation approaches to alleviate the difficulties posed by non-smoothness and non-convexity. We list these approaches in the chromatic order.

1. Chernozhukov and Hong (2003): an MCMC approach.
2. Chernozhukov and Hansen (2006): inverse quantile regression
3. Kaplan and Sun (2017) and de Castro et al. (2019): smoothed estimating equation
4. Chen and Lee (2018): exact GMM via mixed-integer quadratic programming
5. Kaido and Wüthrich (2021): decentralization

To avoid directly optimizing the original GMM optimization, Chernozhukov and Hong (2003) propose using the Markov Chain Monte Carlo method to simulate the distribution of the parameter $\boldsymbol{\theta}$. We can implement this approach using the `bayesmh`. However, as usually seen in the Bayesian approach, the estimator's performance may depend on other inputs such as prior specification. This method is worth implementation, or at least internally. It is interesting to compare the performance of this MCMC method with other frequentist approaches.

Chernozhukov and Hansen (2006) propose estimating the IVQR model by conducting a series of regular quantile regressions, and they name this method as inverse quantile regression. This method is easy to implement but suffers from the curse of dimensionality. The inverse quantile regression needs to do a grid search of dimension R^{k_d}. However, this method can efficiently produce estimates at different quantiles when there are a few endogenous variables. This method should be implemented as a benchmark because it is already widely used.

Kaplan and Sun (2017) and de Castro et al. (2019) suggest to smooth the indicator function in Equation 9 and then to solve the smoothed GMM objective function. This method makes the objective function smooth, but the non-convexity is still there. Thus, to implement this method, we need an efficient optimization algorithm such as simulated annealing for the nonconvex problem in Stata. We can use the Mata function `solvenl()` to solve the estimating equation for the just-identified case. Compared with the inverse quantile regression, the main advantage of this method is that it does not suffer from the curse of dimensionality. By the way, the over-identified case can be transformed into a just-identified case by linear projection.

The recent advances of mixed-integer quadratic programming make it possible to directly solve the nonconvex and nonsmooth optimization problems like Equation 12. Chen and Lee (2018) show how to solve the original GMM objective function in this path. This approach depends on the third-party package for the mixed-integer quadratic programming algorithm. Furthermore, this method is slow even when the sample size is moderate. Given the difficulty of implementing Stata's own mixed-integer quadratic programming algorithm, we should not implement this method.

Finally, Kaido and Wüthrich (2021) propose recasting the original IVQR GMM optimization problem into an iterative series of regular quantile regression. The idea is to divide the problem into a simple and easy-to-compute subproblem, denoting decentralization. The standard errors are obtained via standard bootstrap. This method is promising because it is easy to implement as the inverse quantile regression, and it does not suffer from the curse of dimensionality. Furthermore, this method does not need extra tuning parameters such as priors in the MCMC approach or the bandwidth choice in the smoothed GMM approach. However, this method requires adding the `noconstant` option to `qreg`. Also, some preliminary simulation results show that this method is sensitive to the starting values, and it may even be inconsistent when the model is over-identified.

Considering the different properties of the above approaches, we propose the following priority order of implementation.

1. Inverse quantile regression (benchmark)

2. Smoothed EE

3. Decentralization (postpone due to `noconstant qreg` and over-identification issues)

4. MCMC approach (postpone)

Section 4 describes the inverse quantile regression estimator. Section 5 introduces the smoothed EE estimator. Finally, section A.3 documents the decentralized estimator.

4 Inverse quantile regression

4.1 Main results of IQR

Rather than directly optimizing the original GMM objective function in Equation 12, Chernozhukov and Hansen (2006) propose to do a grid search over $\boldsymbol{\alpha}$, and each step is a regular quantile regression. They label this process as "inverse quantile regression" (IQR). The intuition is from the moment condition in the linear IV quantile regression model in Equation 8. It implies that the τ-th quantile of $Y - \mathbf{D}'\boldsymbol{\alpha}$ conditional on $\mathbf{X}$ and $\mathbf{Z}$ is $\mathbf{X}'\boldsymbol{\beta}$:

$$Q_{Y-\mathbf{D}'\boldsymbol{\alpha}_0}(\tau|X,Z) = \mathbf{X}'\boldsymbol{\beta}_0 + \widehat{\boldsymbol{\Phi}}(\tau)'\boldsymbol{\gamma}_0, \qquad \text{with } \boldsymbol{\gamma}_0 = 0 \tag{14}$$

where $\widehat{\boldsymbol{\Phi}}(\tau)$ is the transformation of $\mathbf{Z}$ and $\mathbf{X}$. In practice, $\widehat{\boldsymbol{\Phi}}(\tau)$ is the linear projection of $\mathbf{D}$ on $\mathbf{X}$ and $\mathbf{Z}$.

Equation 14 means that, based on the true value $\boldsymbol{\alpha}_0$, the conditional quantile regression of $Y - \mathbf{D}'\boldsymbol{\alpha}_0$ on $\mathbf{X}$ and $\mathbf{Z}$ will yield the coefficient on $\widehat{\boldsymbol{\Phi}}(\tau)$ to zero. Given a sequence of $\boldsymbol{\alpha}$, denoted as $A = \{\boldsymbol{\alpha}_j\}_{j=1}^{J}$, we just need to run J regular quantile regression of $y - \mathbf{D}'\boldsymbol{\alpha}_j$ on $\mathbf{X}$ and $\widehat{\boldsymbol{\Phi}}(\tau)$. Denote the estimate for $\boldsymbol{\gamma}_j$ as $\hat{\boldsymbol{\gamma}}(\boldsymbol{\alpha}_j)$. Then, the estimator $\hat{\boldsymbol{\alpha}}_{IQR}$ for $\boldsymbol{\alpha}$ is $\boldsymbol{\alpha}_j$ such that $\hat{\boldsymbol{\gamma}}(\boldsymbol{\alpha}_j)$ is the smallest. Algorithm 1 describes the IQR procedure.

Algorithm 1: Inverse quantile regression

1. Define a sequence of $\boldsymbol{\alpha}$'s, denoted by $A = \{\boldsymbol{\alpha}_j\}_{j=1}^{J}$.

2. Define the quantile level τ.

3. For each variable in $\mathbf{D}$, compute its linear projection on the space spanned by $\mathbf{X}$ and $\mathbf{Z}$. Denote the predicted $\mathbf{D}$ as $\widehat{\boldsymbol{\Phi}}$.

4. For j from 1 to J, make the following loop.

   (a) Run regular τ-th quantile regression of $Y - \mathbf{D}'\boldsymbol{\alpha}_j$ on $\mathbf{X}$ and $\widehat{\boldsymbol{\Phi}}$. Denote the estimate for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ as $\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}_j)$ and $\hat{\boldsymbol{\gamma}}(\boldsymbol{\alpha}_j)$, respectively. Also denote the $\hat{\boldsymbol{\Omega}}(\boldsymbol{\alpha}_j)$ as the estimated variance matrix for $\sqrt{N}(\hat{\boldsymbol{\gamma}}(\boldsymbol{\alpha}_j) - \boldsymbol{\gamma}(\boldsymbol{\alpha}_j))$.

   (b) Compute the norm of $\hat{\boldsymbol{\gamma}}(\boldsymbol{\alpha}_j)$ as $W(\boldsymbol{\alpha}_j) = N\hat{\boldsymbol{\gamma}}(\boldsymbol{\alpha}_j)'\hat{\boldsymbol{\Omega}}(\boldsymbol{\alpha}_j)^{-1}\hat{\boldsymbol{\gamma}}(\boldsymbol{\alpha}_j)$

5. The IQR estimate for $\boldsymbol{\alpha}$ is $\hat{\boldsymbol{\alpha}}_{IQR}$ such that $W(\hat{\boldsymbol{\alpha}}_{IQR})$ is the smallest among $\{W(\boldsymbol{\alpha}_j)\}_{j=1}^{J}$. Formally,

$$\hat{\boldsymbol{\alpha}}_{IQR} = \arg\min_{\boldsymbol{\alpha}\in A} W(\boldsymbol{\alpha})$$

6. Given $\hat{\boldsymbol{\alpha}}_{IQR}$, the estimate for $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}_{IQR})$.

To obtain the standard error of the IQR estimator, we have the following theorem.

THEOREM 2. *(Theorem 3 in Chernozhukov and Hansen (2006)) Given some regularity assumptions, for* $\epsilon_i(\tau) = Y_i - \mathbf{D}_i'\boldsymbol{\alpha}(\tau) - \mathbf{X}_i'\boldsymbol{\beta}(\tau)$ *and* $l_i(\tau, \boldsymbol{\theta}(\tau)) = (\tau - \mathbb{1}(\epsilon_i(\tau) < 0))$*, where* $\boldsymbol{\theta} = (\boldsymbol{\alpha}(\tau), \boldsymbol{\beta}(\tau))$*;*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}(\cdot) - \boldsymbol{\theta}(\cdot)) = -\mathbf{J}(\cdot)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} l_i(\cdot, \boldsymbol{\theta}(\cdot))\boldsymbol{\Psi}_i(\cdot) + o_p(1) \implies \mathbf{b}(\cdot) \tag{15}$$

where $\mathbf{b}(\cdot)$ *is a mean zero Gaussian process with covariance function*

$$\mathbf{E}\mathbf{b}(\tau)\mathbf{b}(\tau')' = \mathbf{J}(\tau)^{-1}\mathbf{S}(\tau, \tau')[\mathbf{J}(\tau')^{-1}]'$$

and

$$\begin{aligned}\mathbf{\Psi}_i(\tau) &= (\mathbf{\Phi}_i(\tau)', \mathbf{X}_i')' \\ \mathbf{J}(\tau) &= \mathbf{E}\left[f_{\epsilon(\tau)}(0|\mathbf{X}, \mathbf{D}, \mathbf{Z})\mathbf{\Psi}(\tau)[\mathbf{D}', \mathbf{X}']\right] \\ \mathbf{S}(\tau, \tau') &= (\min(\tau, \tau') - \tau\tau')\mathbf{E}\mathbf{\Psi}(\tau)\mathbf{\Psi}(\tau)'\end{aligned}$$

where $f_{\epsilon(\tau)}(0|\mathbf{X}, \mathbf{D}, \mathbf{Z})$ *is the conditional density of* $\epsilon(\tau)$ *evaluated at zero.*

For a discussion on the intuition behind Theorem 2, see Section A.2

REMARK 1. *(Remark 3 in Chernozhukov and Hansen (2006)) A basic implication of Theorem 2 is that for a given probability index* τ

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}(\tau)) \to N(0, \mathbf{J}(\tau)^{-1}\mathbf{S}(\tau, \tau)[\mathbf{J}(\tau)^{-1}]') \tag{16}$$

Also, for any finite collection of quantile indices $\tau_j, j \in T$

$$\{\sqrt{n}(\widehat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}(\tau))\}_{j\in T} \to N(0, \{\mathbf{J}(\tau_k)^{-1}\mathbf{S}(\tau_k, \tau_l)[\mathbf{J}(\tau_l)^{-1}]'\}_{k,l\in T}) \tag{17}$$

REMARK 2. *(Remark 4 in Chernozhukov and Hansen (2006)) The components in the variance matrix in (16) and (17) can be obtained by its sample counterparts:*

$$\widehat{\mathbf{S}}(\tau, \tau') = (\min(\tau, \tau') - \tau\tau')\frac{1}{n}\sum_{i=1}^{n}\widehat{\mathbf{\Psi}}_i\widehat{\mathbf{\Psi}}_i' \tag{18}$$

The estimator for $\mathbf{J}(\tau)$ *takes the form*

$$\frac{1}{nh_n}\sum_{i}^{n} K\left(\frac{-\epsilon_i(\tau)}{h_n}\right)\widehat{\mathbf{\Psi}}_i[\mathbf{D}_i', \mathbf{X}_i'] \tag{19}$$

where $K(\cdot)$ *is a Kernel function and* h_n *is the bandwidth.*

In practice, we can use Kernel as in Stata command `kdensity`. For example, the Gaussian Kernel is

$$K(z) = \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}$$

For the bandwidth, we can use either the Silverman's rule of thumb bandwidth or the bandwidth used Koenker (2005, Page 81).

The Silverman's rule of thumb bandwidth is

$$h_s = 0.9 \min\left(\widehat{\sigma(\epsilon)}, \frac{M}{1.349}\right) n^{-\frac{1}{5}} \tag{20}$$

where $\widehat{\sigma(\epsilon)}$ is the standard deviation of ϵ and M is the interquartile range of ϵ.

The Koenker bandwidth is

$$h_k = \min\left(\widehat{\sigma(\epsilon)}, \frac{M}{1.349}\right) \left(\Phi^{-1}(\tau + h_1) - \Phi^{-1}(\tau - h_1)\right) \tag{21}$$

where h_1 can be one of the bandwidth in Hall and Sheather (1988) and Bofinger (1975). In particular,

$$h_{hs} = n^{-1/3}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)^{2/3} \left[\frac{3}{2} \times \frac{\phi\left\{\Phi^{-1}(\tau)\right\}^2}{2\Phi^{-1}(\tau)^2 + 1}\right]^{1/3} \tag{22}$$

$$h_b = n^{-1/5} \left[\frac{9}{2} \times \frac{\phi\left\{\Phi^{-1}(\tau)\right\}^4}{\left\{2\Phi^{-1}(\tau)^2 + 1\right\}^2}\right]^{1/5} \tag{23}$$

4.2 Robust inference approach

In the presence of weak instruments the inference is unreliable. Chernozhukov and Hansen (2008) proposed an inference approach robust to the weak instruments. In addition, this approach allows evaluating the quality of grid specification so that the grid covers the true value of $\boldsymbol{\alpha}$ with a predefined probability level.

PROPOSITION 1. *(Proposition 1 in Chernozhukov and Hansen (2008)) When $\boldsymbol{\alpha} = \boldsymbol{\alpha}(\tau)$,*

$$W_n[\boldsymbol{\alpha}(\tau)] \rightarrow_d \chi^2(dim(\boldsymbol{\gamma}))$$

and for the confidence region $CR_p[\boldsymbol{\alpha}(\tau)] = \{\boldsymbol{\alpha} \in A : W_n(\boldsymbol{\alpha}) < c_p\}$, where $P(\chi^2(dim(\boldsymbol{\gamma})) < c_p) = p$,

$$P\{\boldsymbol{\alpha}(\tau) \in CR_p[\boldsymbol{\alpha}(\tau)]\} = P\{W_n[\boldsymbol{\alpha}(\tau)] < c_p\} = p \tag{24}$$

Intuitively, $W_n(\boldsymbol{\alpha}(\tau))$ is the Wald statistic for testing whether the coefficients for the instruments are zero ($\hat{\boldsymbol{\gamma}} = 0$). When $\boldsymbol{\alpha}$ equals the true value $\boldsymbol{\alpha}(\tau)$, $W(\cdot)$ is χ^2 distributed with the degree of freedom of dimension of $\boldsymbol{\gamma}$. Thus a valid confidence interval for $\boldsymbol{\alpha}$ can be constructed by the inversion of the Wald statistic. That is

$CR_p[\boldsymbol{\alpha}(\tau)] = \{\boldsymbol{\alpha} \in A : W_n(\boldsymbol{\alpha}) < c_p\}$ cover the true value of $\boldsymbol{\alpha}$ with probability approaching p

The confidence region is a byproduct of the inverse quantile regression. Algorithm 2 is an extension of algorithm 1 so that the robust confidence region is computed. Following Chernozhukov and Hansen (2008), the robust confidence interval CR_p is also called dual confidence interval, abbreviated as dual CI.

Algorithm 2: Inverse quantile regression with robust inference

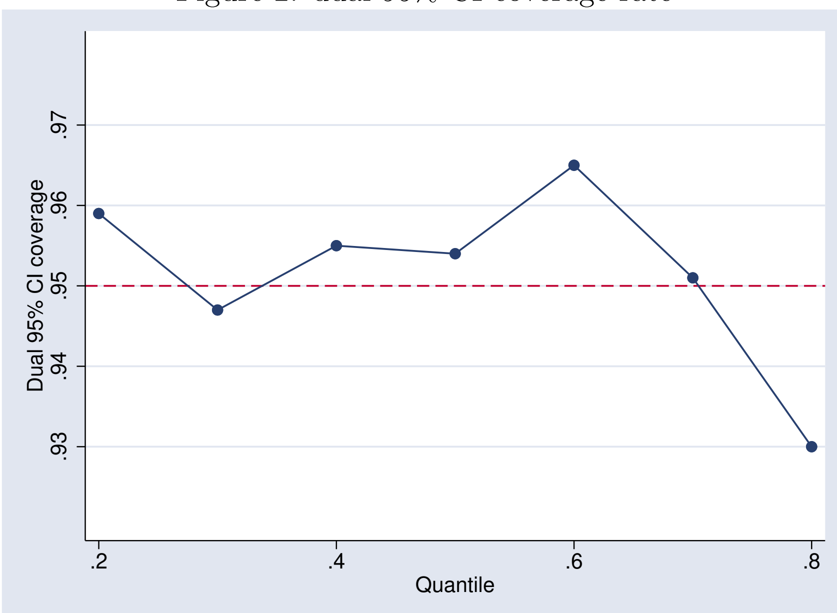
1. Define a sequence of $\boldsymbol{\alpha}$'s, denoted by $A = \{\boldsymbol{\alpha}_j\}_{j=1}^J$.
2. Define the quantile level τ.
3. For each variable in $\mathbf{D}$, compute its linear projection on the space spanned by $\mathbf{X}$ and $\mathbf{Z}$. Denote the predicted $\mathbf{D}$ as $\widehat{\boldsymbol{\Phi}}$.
4. For j from 1 to J, make the following loop.
   (a) Run regular τ-th quantile regression of $Y - \mathbf{D}'\boldsymbol{\alpha}_j$ on $\mathbf{X}$ and $\widehat{\boldsymbol{\Phi}}$. Denote the estimate for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ as $\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}_j)$ and $\hat{\boldsymbol{\gamma}}(\boldsymbol{\alpha}_j)$, respectively. Also denote the $\hat{\boldsymbol{\Omega}}(\boldsymbol{\alpha}_j)$ as the estimated variance matrix for $\sqrt{N}(\hat{\boldsymbol{\gamma}}(\boldsymbol{\alpha}_j) - \boldsymbol{\gamma}(\boldsymbol{\alpha}_j))$.
   (b) Compute the norm of $\hat{\boldsymbol{\gamma}}(\boldsymbol{\alpha}_j)$ as $W(\boldsymbol{\alpha}_j) = N\hat{\boldsymbol{\gamma}}(\boldsymbol{\alpha}_j)'\hat{\boldsymbol{\Omega}}(\boldsymbol{\alpha}_j)^{-1}\hat{\boldsymbol{\gamma}}(\boldsymbol{\alpha}_j)$
5. The IQR estimate for $\boldsymbol{\alpha}$ is $\hat{\boldsymbol{\alpha}}_{IQR}$ such that $W(\hat{\boldsymbol{\alpha}}_{IQR})$ is the smallest among $\{W(\boldsymbol{\alpha}_j)\}_{j=1}^J$. Formally,
$$\hat{\boldsymbol{\alpha}}_{IQR} = \arg\min_{\boldsymbol{\alpha} \in A} W(\boldsymbol{\alpha})$$
6. Given $\hat{\boldsymbol{\alpha}}_{IQR}$, the estimate for $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}_{IQR})$.
7. The dual confidence region for $\boldsymbol{\alpha}(\tau)$, CR_p, can be computed as $CR_p[\boldsymbol{\alpha}(\tau)] = \{\boldsymbol{\alpha}_j : W_n(\boldsymbol{\alpha}_j) < c_p\}$, where $P(\chi^2(dim(\boldsymbol{\gamma})) < c_p) = p$. By default, $p = 0.95$. The upper and lower bound of $CR_p(\boldsymbol{\alpha}(\tau)$ may be used as endpoints of confidence interval for $\boldsymbol{\alpha}(\tau)$
8. If CR_p is empty, $\mathbf{A}$ is not a valid grid specification because the true value of $\boldsymbol{\alpha}(\tau)$ is not covered by $\mathbf{A}$ with a big probability.

4.3 Initial grid

The initial grid points are computed using the two-stage-quantile regression, extending the two-stage-median regression in Amemiya (1982). The quality of the initial grid can be evaluated using the dual CI. The implementation requires that the initial grid interval be wider than the dual CI. Otherwise, `ivqregress iqr` will error out.

A simulation is run to evaluate the effectiveness of the dual CI. In particular, in a repeated sampling setting, we compute the coverage rate of the dual CI. In each repetition, we compute the coverage indicator, which is either 1 if the true value is within the dual CI or 0 otherwise. The coverage is the average of the indicator variable over 1000 repetitions. The sample size is set to 1000. Given the default 95% level, the coverage rate of the dual CI should be close to 0.95. Figure 5 plots the coverage rate of dual CI at different quantile estimates. As expected, the dual CI can cover the true value with a probability close to 95%.

Figure 2: dual 95% CI coverage rate



The two-stage-quantile regression is computed in the following steps.

1. Compute $\hat{\mathbf{\Phi}}(\tau)$, which is the linear projection of $\mathbf{D}$ on $\mathbf{X}$ and $\mathbf{Z}$.

2. Run a quantile regression of Y on $\mathbf{X}$ and $\hat{\mathbf{\Phi}}(\tau)$. Denote $\tilde{\boldsymbol{\alpha}}$ as the point estimates for the coefficient on $\hat{\mathbf{\Phi}}(\tau)$ and $\tilde{s}$ as its standard errors. $\tilde{s}$ is computed by assuming the error term is normally distributed.

3. Compute the lower and upper bound of the grid. The lower bound is $lb = \tilde{\boldsymbol{\alpha}} - 4\tilde{s}$, and the upper bound is $ub = \tilde{\boldsymbol{\alpha}} + 4\tilde{s}$.

4. By default, the grid points are 30 equally spaced points between lb and ub. We can also specify the number of grid points in option `ngrid()`.

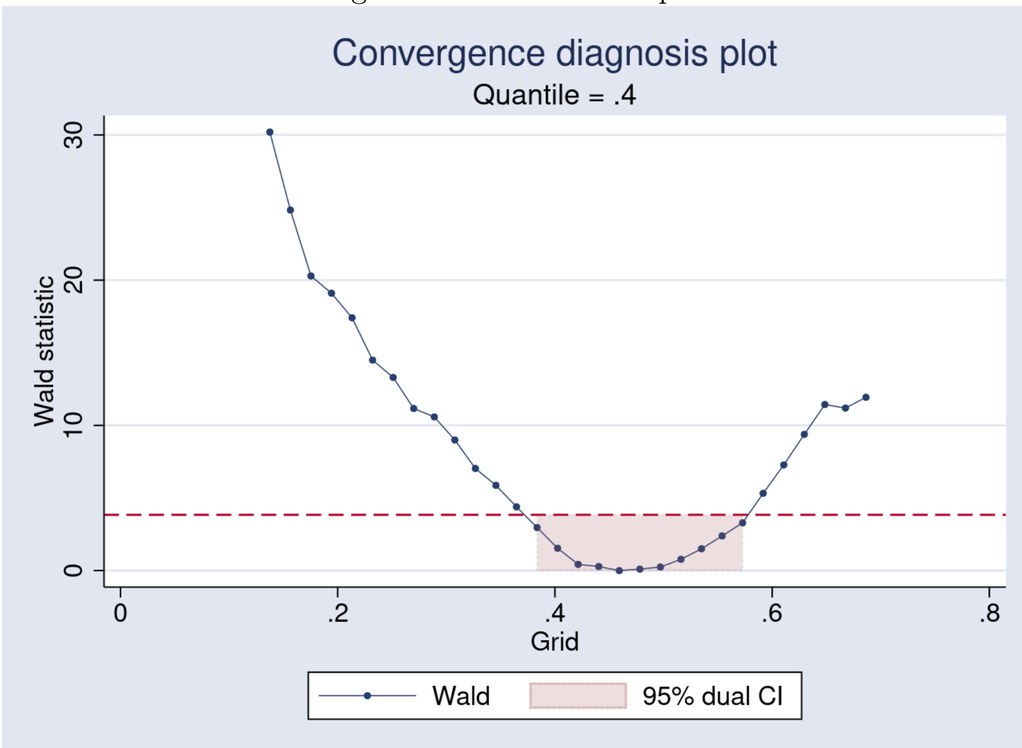
We show one example to illustrate the connection between the dual CI and the grid. To fix the idea, we fit an IVQR model using the simulated data. First, we fit the inverse quantile regression of y on the endogenous $d1$ and exogenous $x1$ and $x2$, and use $z1$ and $z2$ as instruments.

```
. ivqregress iqr y x1 x2 (d1 = z1 z2),  quantile(40) noadaptive

Initial grid
    quantile = 0.40: .........10.........20.........30

IV .4 quantile regression                               Number of obs =   1,000
Estimator: Inverse quantile regression                  Wald chi2(3)  = 3534.04
                                                        Prob > chi2   =  0.0000

------------------------------------------------------------------------------
             |               Robust
           y | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
          d1 |    .459274   .0502662     9.14   0.000      .360754    .5577939
          x1 |   1.393901   .0334069    41.72   0.000     1.328425    1.459377
          x2 |   1.427885   .0321419    44.42   0.000     1.364888    1.490882
       _cons |  -.3458003   .0411389    -8.41   0.000     -.426431   -.2651695
------------------------------------------------------------------------------
Endogenous: d1
 Exogenous: x1 x2 z1 z2
```

By default, the initial grid is computed by the two-stage-quantile regression. We now plot the Wald statistic for each point in the grid.

```
. estat waldplot
```

Figure 3: Wald statistic plot



In the above graph, the Wald statistics are plotted for each point in the grid. The critical value of the Wald statistic is shown as a red horizontal dash line. The dual CI is the grid points such that the corresponding Wald statistic is smaller than the critical value, shown in the red shaded area.

From the above graph, we can visually inspect two features regarding the estimation. First, we see that the point estimate results in the minimum of the Wald statistic. Second, we see that the point estimate is within the 95% dual CI and the initial grid interval is wider than the dual CI. Thus, we are 95% or more confident that the initial grid interval contains the true value of the parameter.

To see the numerical values of the dual CI, we use `estat dualci`.

```
. estat dualci

Dual confidence interval                            Number of obs = 1,000

                            Robust                              Dual
           y  Coefficient  std. err.      z    P>|z|     [95% conf. interval]

          d1     .459274   .0502662     9.14   0.000     .3835663    .5728356
```

4.4 Adaptive grid search

Sometimes, a point estimate is found in a flat region, which implies some degree of weak identification of the parameter. The weak identification can be caused by the weak identified model or the coarse grid points. The adaptive grid search can help alleviate the weak identification issue caused by the coarse grid points. However, the adaptive grid search can not help if the model is intrinsically weakly identified.

The adaptive grid search is conducted in the following steps.

1. Based on the initial grid points, compute the inverse quantile regression and the dual CI.

2. Construct a new grid using the dual CI, that is, construct several equally spaced points between the lower and upper bound in the dual CI.

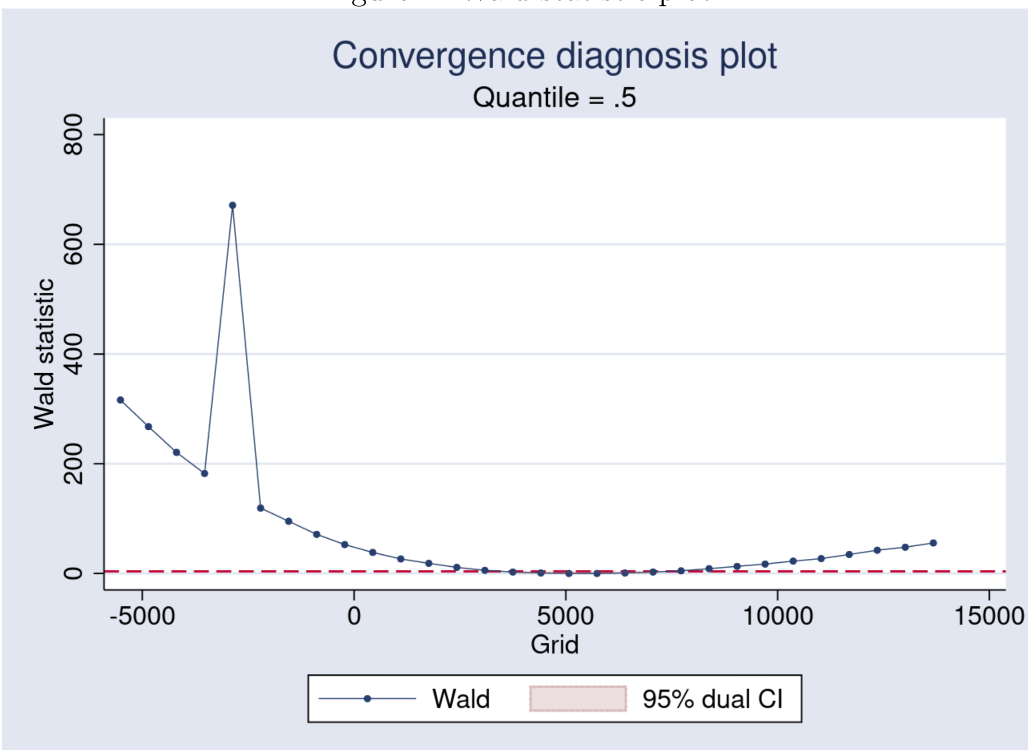
3. Compute the inverse quantile regression using the grid in step 2.

Here is one example. First, we fit a IVQR model using the original grid and plot the Wald statistics.

```
. use assets2, clear
(Excerpt from Chernozhukov and Hanson (2004) Rev. of Economics and Statistics)

.
. ivqregress iqr assets (i.p401k = i.e401k) income age familysize ///
>         i.married i.ira i.pension i.ownhome educ, quantile(50) noadaptive

Initial grid
    quantile = 0.50: .........10.........20.........30

IV median regression                                   Number of obs =   9,913
Estimator: Inverse quantile regression                 Wald chi2(9)  = 1312.24
                                                       Prob > chi2   =  0.0000

------------------------------------------------------------------------------
             |               Robust
      assets | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
     1.p401k |   5073.007   551.8459     9.19   0.000     3991.409    6154.605
      income |    .159635   .0124353    12.84   0.000     .1352622    .1840077
         age |   100.7959   8.584905    11.74   0.000     83.96984     117.622
  familysize |  -203.4073   54.42447    -3.74   0.000    -310.0773   -96.73727
             |
     married |
    Married  |  -1348.987   227.0043    -5.94   0.000    -1793.907   -904.0668
             |
         ira |
        Yes  |   22630.45   1016.463    22.26   0.000     20638.21    24622.68
             |
     pension |
Receives ..  |  -708.9357   210.4489    -3.37   0.001    -1121.408   -296.4633
             |
     ownhome |
```

```
         Yes    -39.62138   154.8029    -0.26   0.798    -343.0294    263.7867
        educ    -98.81898   32.18152    -3.07   0.002    -161.8936   -35.74435
       _cons    -5026.207    570.501    -8.81   0.000    -6144.368   -3908.045

Endogenous: 0b.p401k 1.p401k
 Exogenous: income age familysize 0b.married 1.married 0b.ira 1.ira
            0b.pension 1.pension 0b.ownhome 1.ownhome educ 0b.e401k 1.e401k

.
. estat waldplot, name(a)
```

Figure 4: Wald statistic plot

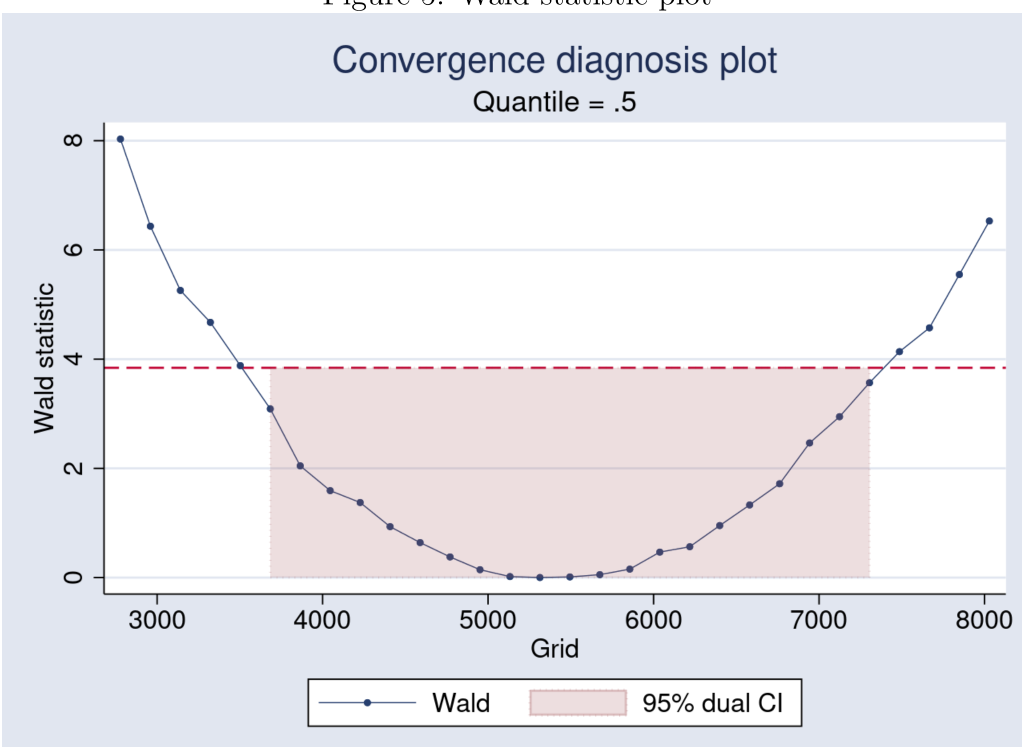


The Wald plot shows that the minimum is found in a flat region. Now, we refit the IVQR model but using the adaptive grid search.

```
. use assets2, clear
(Excerpt from Chernozhukov and Hanson (2004) Rev. of Economics and Statistics)

.
. ivqregress iqr assets (i.p401k = i.e401k) income age familysize ///
>          i.married i.ira i.pension i.ownhome educ, quantile(50)

Initial grid
    quantile = 0.50: .........10.........20.........30

Adaptive grid
    quantile = 0.50: .........10.........20.........30

IV median regression                                    Number of obs =    9,913
Estimator: Inverse quantile regression                  Wald chi2(9)  = 1289.75
                                                        Prob > chi2   =  0.0000
```

```
------------------------------------------------------------------------------
             |               Robust
      assets | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
     1.p401k |   5313.397   573.2818     9.27   0.000     4189.786    6437.009
      income |   .1577512   .0124889    12.63   0.000     .1332735    .1822289
         age |   99.96526   8.561923    11.68   0.000      83.1842    116.7463
  familysize |  -197.8251   54.36773    -3.64   0.000    -304.3838   -91.26627
             |
     married |
    Married  |  -1359.124   227.3366    -5.98   0.000    -1804.696   -913.5528
             |
         ira |
        Yes  |   22629.61   1022.706    22.13   0.000     20625.15    24634.08
             |
     pension |
 Receives .. |  -693.8347   210.6176    -3.29   0.001    -1106.638   -281.0317
             |
     ownhome |
        Yes  |  -30.29657   154.7265    -0.20   0.845     -333.555    272.9618
        educ |  -96.43983   32.09465    -3.00   0.003    -159.3442   -33.53547
       _cons |  -4998.673   570.1315    -8.77   0.000     -6116.11   -3881.236
------------------------------------------------------------------------------
Endogenous: 0b.p401k 1.p401k
 Exogenous: income age familysize 0b.married 1.married 0b.ira 1.ira
            0b.pension 1.pension 0b.ownhome 1.ownhome educ 0b.e401k 1.e401k

.
. estat waldplot, name(b)
```

Figure 5: Wald statistic plot



Clearly, the adaptive grid search helps to identify the point estimate in a more convex region. By default, `ivqregress iqr` is doing a robust and adaptive grid search.

5 Smoothed estimating equation estimator

The basic idea of smoothd estimating equation (SEE) estimator for the IVQR model is to replace the indicator function in the moment condition with a smooth function. To be precise, we replace the moment condition in Equation (8) with

$$\frac{1}{n}\sum_{i=1}^{n}\left[\tau - \widetilde{\mathbb{1}}(Y_i - \mathbf{D}_i'\boldsymbol{\alpha} - \mathbf{X}_i'\boldsymbol{\beta} \leq 0)\right]\boldsymbol{\Psi}_i(\mathbf{X}, \mathbf{Z}) = 0 \tag{25}$$

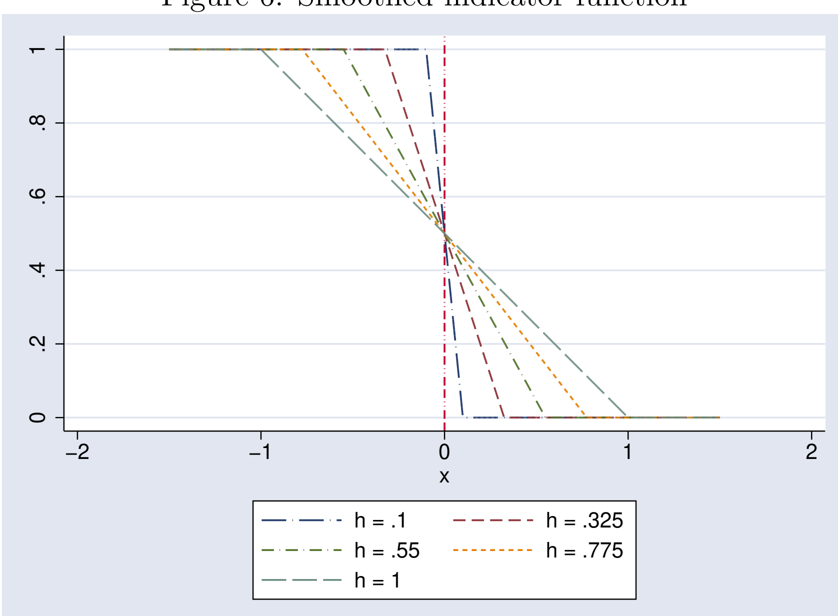
where $\widetilde{\mathbb{1}}()$ is defined as

$$\widetilde{\mathbb{1}}(v) = \begin{cases} 1 & \text{if } v \leq -1 \\ 0 & \text{if } v \geq 1 \\ \frac{1-v}{2} & \text{if } -1 < v < 1 \end{cases} \tag{26}$$

Mathematically, $\widetilde{\mathbb{1}}() = \max\{0, \min\{1, \frac{1-v}{2}\}\}$.

Figure 6 plot the $\widetilde{\mathbb{1}}(v/h)$ with different bandwidth h. We see that smaller bandwidth approximate the indicator function better.

Figure 6: Smoothed indicator function



Here are some remarks on the smoothed EE estimator.

- In practice, $\mathbf{\Psi}(\mathbf{X},\mathbf{Z}) = (\mathbf{X}', \mathbf{\Phi}(\mathbf{X},\mathbf{Z})')'$, and $\mathbf{\Phi}(\mathbf{X},\mathbf{Z})$ can be computed as the linear projection of $\mathbf{D}$'s on the space spanned by $\mathbf{X}$ and $\mathbf{Z}$. So, the over-identified model can be transformed to a just-identified model.
- Given a bandwidth h, the solution to Equation (25) can be found via mata function `solvenl()`.
- The SEE estimator does not suffer from the curse of dimensionality as seen in the inverse quantile regression estimator.

5.1 Optimal bandwidth

The optimal bandwidth minimize the asymptotic MSE of the smoothed estimating equations. For details, see Proposition 2 in Kaplan and Sun (2017).

The optimal bandwidth h^*_{SEE} is defined as

$$h^*_{SEE} = \left(\frac{(r!)^2 \left[1 - \int_{-1}^{1} G^2(u)du\right] f_U(0)}{2r\left[\int_{-1}^{1} G'(v)v^r dv\right]^2 \left[f_U^{(r-1)}(0)\right]^2} \frac{p}{n} \right)^{\frac{1}{2r-1}} \tag{27}$$

where

- $G(u) = \widetilde{\mathbb{1}}(-u)$
- r means $G(u)$ is a bounded rth order kernel such that $\int_{-1}^{1} v^k G'(v)dv = 0$ for $k = 1, 2, \ldots, r-1$ and $\int_{-1}^{1} |v^r G'(v)dv| < \infty$. In the case of $\widetilde{\mathbb{1}}()$, $r = 2$.
- p is the sum of the dimension of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$.
- $U = Y - \mathbf{D}'\boldsymbol{\alpha} - \mathbf{X}'\boldsymbol{\beta}$ and $f_U(0)$ is the PDF of U at the point 0.
- $f_U^{(1)}(0)$ is the derivatives of the PDF of U evaluated at the point 0.

Given the definition of $\widetilde{\mathbb{1}}()$ in Equation (26), we can further simplify the expression of h^*_{SEE}. In particular,

$$1 - \int_{-1}^{1} G^2(u)du = 1 - \int_{-1}^{1} \frac{(1+v)^2}{4} dv = 1 - \frac{(1+v)^3}{12}|_{-1}^{1} = \frac{1}{3}$$
$$\left[\int_{-1}^{1} G'(v)v^r dv\right]^2 = \left[\int_{-1}^{1} \frac{v^2}{2} dv\right]^2 = \left[\frac{v^3}{6}|_{-1}^{1}\right]^2 = \left[\frac{1}{6} - \frac{-1}{6}\right]^2 = \frac{1}{9}$$

Plug in these two values into h^*_{SEE}, we have

$$
\begin{aligned}
h^*_{SEE} &= \left(\frac{(2!)^2 \frac{1}{3} f_U(0)}{2 * 2 \frac{1}{9} \left[f_U^{(2-1)}(0) \right]^2} \frac{p}{n} \right)^{\frac{1}{2*2-1}} \\
&= \left(\frac{3 f_U(0)}{\left[f'_U(0) \right]^2} \frac{p}{n} \right)^{\frac{1}{3}} \\
&= \left(\frac{3p}{n} \right)^{\frac{1}{3}} \left(\frac{f_U(0)}{[f'_U(0)]^2} \right)^{\frac{1}{3}}
\end{aligned} \tag{28}
$$

Notice that the term $\left(\frac{3p}{n}\right)^{\frac{1}{3}}$ can be used as an initial value of the bandwidth. We can nonparametrically or parametrically estimate $f_U(0)$ and $f'_U(0)$.

5.2 Nonparametrically estimated optimal bandwidth

The optimal bandwidth in Equation (28) requires the evaluation of $f_U(0)$ and $f'_U(0)$. Both terms can be estimated using a Kernel-based method.

For $f_U(0)$, given an initial estimate of residuals $\hat{u}_i = Y_i - \mathbf{D}'_i \hat{\boldsymbol{\alpha}}(\tau) - \mathbf{X}'_i \hat{\boldsymbol{\beta}}(\tau)$, the kernel density estimator is

$$
\frac{1}{ns} \sum_{i=1}^{n} K\left(\frac{-\hat{u}_i}{s}\right) \tag{29}
$$

where s is the kernel bandwidth and $K()$ is the Gaussian Kernel.

Following the point-wise optimal bandwidth approach in DasGupta (2008) and the Gaussian reference approach in Silverman (1998), the optimal s^* is

$$
s^* = \left(\frac{1}{2n\sqrt{\pi}} \right)^{1/5} \sigma \left\{ \phi(\Phi^{-1}(\tau))[\left(\Phi^{-1}(\tau)\right)^2 - 1]^2 \right\}^{-1/5} \tag{30}
$$

where $\phi()$ is standard Normal density function and $\Phi()$ is standard Normal CDF. σ can be replaced by the sample deviation of the $\hat{u}$ or the standard normal interquartile range, whichever is smaller.

For the density derivative $f'_U(0)$, the kernel estimator is

$$
\frac{1}{nb^2} \sum_{i=1}^{n} K'(-\hat{u}_i/b) \tag{31}
$$

, where $K'()$ is the derivative of the Gaussian Kernel.

Following Wand and Jones (1995), the optimal bandwidth b^* can be estimated as

$$b^* = n^{-1/7}\sigma \left(\frac{3/(4\sqrt{\pi})}{\phi(\Phi^{-1}(\tau))\left[\Phi^{-1}(\tau)\right]^2\left[3-[\Phi^{-1}(\tau)]^2\right]^2} \right)^{1/7} \tag{32}$$

5.3 Parametrically estimated optimal bandwidth

The optimial bandwidth in Equation (28) can also be estimated parametrically. Assuming the residuals u follows a specific distribution, the terms $f_U(0)$ and $f'_U(0)$ can be computed.

Suppose u follows a Gaussian distribution, the optimal bandwidth can be simplified as

$$h^*_{SEE} = n^{-1/3}\sigma \left(\frac{3k_d}{\{\Phi^{-1}(\tau)\}^2\phi(\Phi^{-1}(\tau))} \right)^{1/3} \tag{33}$$

5.4 Detection of “bad” bandwidth

Given any bandwidth h, the SEE estimator may or may not converge. A “bad” bandwidth means that the SEE estimator does not converge or the SEE estimator converges, but the point estimates are out of 95% dual confidence interval. In the first scenario, the non-convergence issue can be detected easily. We can employ the “robust” inference approach in Chernozhukov and Hansen (2008) in the second scenario. For a detailed discussion, see Section 4.2.

In particular, we run an auxiliary quantile regression of $Y - \mathbf{D}'\hat{\alpha}_{SEE}$ on $\mathbf{X}$ and $\mathbf{\Phi}(\mathbf{X}, \mathbf{Z})$, where $\hat{\alpha}_{SEE}$ is SEE estimates for $\boldsymbol{\alpha}$. Denote the W_h as the Wald statistic for the coefficient on $\mathbf{\Phi}(\mathbf{X}, \mathbf{Z})$. If W_h is greater than the critical value c_p, the point estimates for the endogenous coefficients are out of the 95% confidence interval.

5.5 Variance-covariance estimation

The SEE estimator has the same (first-order) asymptotic distribution and covariance structure as the inverse quantile regression estimator. See the second equation in Theorem 5 in de Castro et al. (2019) for the derivation. It turns out that the SEE estimator in Kaplan and Sun (2017) is a special case of the de Castro et al. (2019) when assuming the linear IVQR model. The formulas for the asymptotic variance estimator are the same as in Remarks 1 and 2.

Bootstrap can also be used to compute the standard errors.

6 General inference

We outline the inference procedures in Chernozhukov and Hansen (2006) that are of particular interest in the IVQR model.

It is convenient to write the hypothesis in the following null hypothesis:

$$\mathbf{R}(\tau)(\boldsymbol{\theta}(\tau) - \mathbf{r}(\tau)) = 0 \quad \text{for each } \tau \in T \tag{34}$$

where $\mathbf{R}(\tau)$ is a $q \times p$ matrix of rank q with q is smaller than the $dim(\boldsymbol{\theta})$, and $\mathbf{r}(\tau) \in R^p$. This form is different from the classical setting because $\boldsymbol{\theta}()$ and $\mathbf{r}()$ are functions, which need to be estimated in some cases.

Based on the IVQR model estimates $\hat{\boldsymbol{\theta}}(\cdot)$, We focus on the basic inference process

$$\mathbf{v}_n(\cdot) = \mathbf{R}(\cdot)(\hat{\boldsymbol{\theta}}(\cdot) - \hat{\mathbf{r}}(\cdot)) \tag{35}$$

where $\hat{\mathbf{r}}()$ are either a vector of constant or an estimate from the classical quantile regression.

The Kolmogorov-Smirnov (KS) statistic $S_n = f(\sqrt{n}\mathbf{v}_n(\cdot))$ is a function of $\mathbf{v}_n(\cdot)$, which is

$$S_n = \sqrt{n} sup_{\tau \in T} ||\mathbf{v}_n(\tau)||_{\hat{\mathbf{\Lambda}}(\tau)} \tag{36}$$

where $||\mathbf{x}||_{\mathbf{A}} = \sqrt{\mathbf{x}'\mathbf{A}\mathbf{x}}$, the choice of $\mathbf{\Lambda}(\tau)$ will be discussed later.

To describe the limit distribution of S_n, we need the following assumption in addition to Assumption 1.

ASSUMPTION 2. *(Assumption 3 in Chernozhukov and Hansen (2006))*

1. $\mathbf{R}(\cdot)(\boldsymbol{\theta}(\cdot) - \mathbf{r}(\cdot)) = \mathbf{g}(\cdot)$, *where the functions* $\mathbf{g}(\tau)$, $\mathbf{R}(\tau)$, *and* $\mathbf{r}(\tau)$ *are continuous and either*
   - *(a)* $\mathbf{g}(\tau) = 0$ *for all* τ
   - *(b) or* $\mathbf{g}(\tau) \neq 0$ *for some* τ.
2. $\sqrt{n}(\widehat{\boldsymbol{\theta}(\cdot)} - \boldsymbol{\theta}(\cdot)) \to \mathbf{b}(\cdot)$ *and* $\sqrt{n}(\widehat{\mathbf{r}(\cdot)} - \mathbf{r}(\cdot)) \to \mathbf{c}(\cdot)$, *where* $\mathbf{b}(\cdot)$ *and* $\mathbf{c}(\cdot)$ *are zero means Gaussian functions that may have different laws under the null and the alternative.*

Assumption 2.2 is satisfied by the inverse quantile regression estimator and the smoothed estimating equation estimator regarding $\hat{\boldsymbol{\theta}}(\cdot)$. Correspondingly, the assumption regarding $\hat{\mathbf{r}(\cdot)}$ is satisfied by the regular quantile regression.

Theorem 3 describes the limit distribution of S_n under the null and the alternative.

THEOREM 3. *(Theorem 4 in Chernozhukov and Hansen (2006))*

1. *Under Assumptions 1 and 2.1a , and 2.2* $S_n \to S = f(\mathbf{v}(\cdot))$*, where* $\mathbf{v}(\cdot) = \mathbf{R}(\cdot)(\mathbf{b}(\cdot) - \mathbf{c}(\cdot))$*. If* $\mathbf{v}(\cdot)$ *has nodegenerate covariance kernel, then for* $\alpha < 1/2$*,* $P(S_n > c(1-\alpha)) \to \alpha = P(f(\mathbf{v}(\cdot)) > c(1-\alpha))$*, where* $c(1-\alpha)$ *is chosen such that* $P(f(\mathbf{v}(\cdot) > c(1-\alpha)) = \alpha$

2. *Under Assumptions 1 and 2.1b ,and 2.2* $S_n \to \infty$ *and* $P_n(S_n > c(1-\alpha)) \to 1$*.*

Theorem 3 is not operational because it does not provide the critical value $c(1-\alpha)$. Following Chernozhukov and Hansen (2006), the critical values can be obtained by resampling scores.

6.1 Critical values by resampling scores

Suppose that we have linear representation for the inference process:

$$\sqrt{n}(\mathbf{v}_n(\cdot) - \mathbf{g}(\cdot)) = -\frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i(\cdot) + o_p(1) \tag{37}$$

where $\mathbf{z}_i(\cdot)$ will be defined below.

Given a sample of the estimated scores $\{\hat{\mathbf{z}}_i(\tau), i \leq n, \tau \in T\}$, consider the following steps.

1. Construct B_n randomly chosen subsets of $\{1, \ldots, n\}$ of size b. Denote subsets as I_i, $i \leq B_n$. Denote $\mathbf{v}_{j,b,n}(\cdot)$ the inference process computed over the jth subset of data, I_j, that is
$$\mathbf{v}_{j,b,n}(\tau) = \frac{1}{b}\sum_{i \in I_j} \hat{\mathbf{z}}_i(\tau)$$
and define $S_{j,b,n} = f(\sqrt{b}\mathbf{v}_{j,b,n}(\cdot))$ as
$$S_{j,b,n} = sup_{\tau \in T}\sqrt{b}||\mathbf{v}_{j,b,n}||_{\hat{\mathbf{\Lambda}}(\tau)}$$

2. Define, for $S = f(\mathbf{v}(\cdot))$, $\Gamma(x) = P(S \leq x)$. Estimate $\Gamma(x)$ by
$$\hat{\Gamma}_{b,n}(x) = 1/B_n \sum_{j=1}^{B_n} \mathbb{1}(S_{j,b,n} \leq x)$$
The critical value is obtained as the $1-\alpha$th quantile of $\hat{\Gamma}_{b,n}(x)$. That is, $c_{b,n}(1-\alpha) = inf\{c : \hat{\Gamma}_{b,n}(c) \geq 1-\alpha)\}$

3. The level α test rejects the null hypothesis when $S_n > c_{b,n}(1-\alpha)$.

6.2 Some interesting testing examples

In the context of IVQR model, the following tesing examples are of particular interest.

- **Hypothesis of no effect**: the null hypothesis is that the treatment has no impact on the outcome: $\alpha(\tau) = 0$. In this case,

$$\begin{aligned} H_0 : &\quad \alpha(\tau) = 0 \quad \text{for all } \tau \in T \\ H_1 : &\quad \alpha(\tau) \neq 0 \quad \text{for some } \tau \in T \end{aligned}$$

So, $R(\tau)$, $r(\cdot)$ and z_i are defined as below

$$\begin{aligned} R(\tau) &= \begin{bmatrix} 1 & 0 & \dots & \dots & \dots & 0 \\ 0 & 1 & 0 & \dots & \dots & 0 \\ \vdots & & & & & \\ 0 & \dots & 1 & 0 & \dots & 0 \end{bmatrix}_{k_d \times p} \\ r(\cdot) &= 0 \\ z_i(\tau) &= R(\tau)[J(\tau)^{-1} l_i(\tau, \theta(\tau))\Psi_i(\tau)] \end{aligned}$$

where

$$\begin{aligned} l_i(\tau, \theta(\tau)) &= [\tau - \mathbb{1}\left(Y_i < D_i'\alpha(\tau) + X_i'\beta(\tau)\right)] \\ \Psi(\tau) &= [\Phi(\tau)', X']' \end{aligned}$$

- **Hypothesis of constant effect**: The hypothesis of a constant effect is that the treatment only affects the location of the outcome Y. That is, $\alpha(\tau) = \alpha$ for all $\tau \in T$. In this case,

$$\begin{aligned} H_0 : &\quad \alpha(\tau) = \alpha \quad \text{for all } \tau \in T \\ H_1 : &\quad \alpha(\tau) \neq \alpha \quad \text{for some } \tau \in T \end{aligned}$$

The definition of $R(\tau)$ is same as in the test of no effect. $\hat{r} = \hat{\theta}(\tau_0)$, where τ_0 can either be 0.5 or some quantile index specified in the model.

The scores for this inference process is

$$z_i(\tau) = R(\tau)\left[J(\tau)^{-1} l_i(\tau, \theta(\tau))\Psi_i(\tau) - J(\tau_0)^{-1} l_i(\tau_0, \theta(\tau_0))\Psi_i(\tau_0)\right]$$

- **Dominance hypothesis**: The test of dominance tests whether the effects are unambiguously beneficial, that is $\alpha(\tau) > 0$ for all $\tau \in T$. One may use the one-sided KS statistics

$$S_n = \sqrt{n} sup_{\tau \in T} \quad max(-\alpha(\tau), 0)$$

In this case,

$$\begin{aligned} H_0 : &\quad \alpha(\tau) > 0 \quad \text{for all } \tau \in T \\ H_1 : &\quad \alpha(\tau) \leq 0 \quad \text{for some } \tau \in T \end{aligned}$$

$r(\cdot) = 0$, and the scores for this inference process is

$$z_i(\tau) = R(\tau)[J(\tau)^{-1} l_i(\tau, \theta(\tau))\Psi_i(\tau)]$$

which is the same as in the test of no effect.

- **Exogeneity Hypothesis**: If D are exogenous, we can estimate the model by the regular quantile regression and denote $\eta(\tau)$ as the quantile regression estimates. The difference between $\theta(\tau)$ and $\eta(\tau)$ can be used to formulate a Hausman test of exogeneity. In this case, the null and alternative is defined as

$$\begin{aligned} H_0 : &\quad \alpha(\tau) = \eta(\tau)_{k_d} \quad \text{for all } \tau \in T \\ H_1 : &\quad \alpha(\tau) \neq \eta(\tau)_{k_d} \quad \text{for some } \tau \in T \end{aligned}$$

$r(\cdot) = \eta(\tau)$ and the scores for the inference process is given by

$$z_i(\tau) = R(\tau) \left[J(\tau)^{-1} l_i(\tau, \theta(\tau))\Psi_i(\tau) - H(\tau)^{-1} d_i(\tau, \eta(\tau)) \right]$$

where

$$d_i(\tau, \eta(\tau)) = \left[\tau - \mathbb{1}\left(y_i < \tilde{X}_i' \eta(\tau) \right) \right] \tilde{X}_i \quad \text{with } \tilde{X}_i = (D_i', X_i')'$$

and

$$H(\tau) = E(f_\epsilon(0|\tilde{X})\tilde{X}\tilde{X}')$$

Notice that $H(\tau)$ and $d_i(\tau, \eta(\tau))$ is analogous to the definition of $J(\tau)$ and $l_i(\tau, \theta(\tau))\Psi_i(\tau)$, respectively. The only difference is to replace the $\Phi(\tau)$ in $\Psi(\tau)$ with D. It is intuitive because under the null D are exogenous, so the instrument for D is itself.

6.3 Implementation details

We provide some details on the estimation of $\Lambda(\tau)$, $J(\tau)$ and $H(\tau)$, and the block size b_n.

- $\Lambda(\tau)$

$$\Lambda^*(\tau) = [\Omega^*(\tau)]^{-1} = [Var(z_i(\tau))]^{-1}$$

where $\Omega^*(\tau)$ can be estimated as

$$\widehat{\Omega}^*(\tau) = \frac{1}{n}\sum_{i=1}^{n} \hat{z}_i(\tau)\hat{z}_i(\tau)'$$

Notice that $\Omega(\tau) = 0$ when $r(\cdot) = \theta(\tau_0)$ in the testing of constant effect, so we can need to set $\widehat{\Omega}^*(\tau) = I$ in this case.

- $J(\tau)$

The estimator for $J(\tau)$ takes the form

$$\frac{1}{nh_n}\sum_{i}^{n} K\left(\frac{-\epsilon_i(\tau)}{h_n}\right)\widehat{\Psi}_i[D_i', X_i']$$

where $\epsilon_i(\tau) = y_i - (D_i', X_i')\theta(\tau)$, $K(\cdot)$ is a Kernel function, and h_n is the bandwidth.

In practice, we can use Gaussian Kernel

$$K(z) = \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}$$

and the Silverman's rule of thumb bandwidth

$$h_n = 0.9\min\left(\widehat{\sigma(\epsilon)}, \frac{M}{1.349}\right)n^{-\frac{1}{5}}$$

where $\widehat{\sigma(\epsilon)}$ is the standard deviation of ϵ and M is the interquartile range of ϵ.

- $H(\tau)$

The estimator for $H(\tau)$ takes the form

$$\frac{1}{nh_n}\sum_{i}^{n} K\left(\frac{-\upsilon_i(\tau)}{h_n}\right)\tilde{X}_i\tilde{X}_i'$$

where $\upsilon_i(\tau) = y_i - \tilde{X}_i'\hat{\eta}(\tau)$, and the definition of $K()$ and h_n is the same as in the estimation of $J(\tau)$.

- b_n

  Following Chernozhukov and Hansen (2006), $b_n = 5n^{2/5}$. Notice that $b_n^2/n \to 0$, the bootstrap of the scores can be done with replacement.

7 Syntax of `ivqregress`

`ivqregress` fits a linear IV quantile regression model using the inverse quantile regression estimator or the smoothed estimating equation estimator.

7.1 Syntax

inverse quantile regression estimator:

> `ivqregress iqr` *depvar* [*varlist*$_1$] (*varname*$_2$ = *varlist*$_{iv}$) [`if`] [`in`] [, *options* *IQR_options*]

smoothed estimating equation estimator:

> `ivqregress smooth` *depvar* [*varlist*$_1$] (*varlist*$_2$ = *varlist*$_{iv}$) [`if`] [`in`] [, *options* *SMOOTH_options*]

where

- *varlist*$_1$ specifies the exogenous variables.
- *varname*$_2$ or *varlist*$_2$ specify the endogenous variables.
- *varlist*$_{iv}$ specifies the instrumental variables.
- Factor variables are allowed for *varlist*$_1$, *varlist*$_2$, or *varname*$_2$.

The options are

options	Description
Model	
quantile(numlist)	estimate quantiles specified in numlist; default is quantile(0.5)
SE/Robust	
vce(*vcespec*)	technique used to estimate standard errors; the default is vce(robust)
Optimization	
[no]log	suppress or display the iteration log
verbose	display a verbose iteration log
Reporting	
level(#)	set confidence level; default is level(95)
display_options	control display formats

Notes:

- The prefixes collect, by, statsby, rolling, bootstrap and xi are allowed.

- The *vcespec* can be only one of the following

  robust [, *robust_options*]

  or

  bootstrap [, *bootstrap_options*]

  where *boostrap_options* are options for bootstrap. See help vce_option.

  The default is robust.

- The *robust_options* is

kernel(*kernel*)	use a nonparametric kernel density estimator; default is epanechnikov
bwidth(#|*bwrule*)	bandwidth used by the kernel density estimator; default is Silverman's rule of thumb

- The *kernel* is

kernel	Description
`epanechnikov`	Epanechnikov kernel function; the default
`epan2`	alternative Epanechnikov kernel function
`biweight`	biweight kernel function
`cosine`	cosine trace kernel function
`gaussian`	Gaussian kernel function
`parzen`	Parzen kernel function
`rectangle`	rectangle kernel function
`triangle`	triangle kernel function

- The *bwrule* is

kernel	Description
`silverman`	Silverman rule of thumb; the default
`hsheather`	Hall–Sheather's bandwidth
`bofinger`	Bofinger's bandwidth

The *IQR_options* are

IQR_options	Description
`ngrid(`$\#_g$`)`	use $\#_g$ grid points; default is `grid(30)`
`bound(`$\#_{min}$ $\#_{max}$ `[, at(`$\#_q$`)])`	specify the lower and the upper bound for the grid in the $\#_q$-th quantile estimation; may be repeated

Notes:

- The option `bound()` can be specified only once for each quantile specified in option `quantile()`. If `at()` is not specified, the bound is applied to all the quantiles specified in option `quantile()`.

The *SMOOTH_options* are

SMOOTH_options	Description
`bwidth(`$\#_b$ `[, at(`$\#_q$`)])`	specify bandwidth $\#_b$ to smooth the estimating equation for the $\#_q$-th quantile estimation; the default is the theoretical optimal bandwidth; may be repeated
`nosearchbwidth`	do not search for feasible bandwidth if the initial bandwidth is not feasible; default is searching for feasible bandwidth
`iterate(#)`	perform maximum of # iterations when solving estimating equation; default is iterate(100)
`tolerance(#)`	tolerance for the coefficient vector; default is `tolerance(1E-9)`
`ztolerance(#)`	tolerance used to determine whether the proposed solution to a zero-finding problem is sufficiently close to zero; default is `ztolerance(1E-9)`

Notes:

- The option `bwidth()` can be specified only once for each quantile specified in option `quantile()`. If `at()` is not specified, the bandwidth is applied to all the quantiles specified in option `quantile()`.

7.2 Quick start

Basic

Use inverse quantile regression to estimate IV median regression of `y` on exogenous `x1` and endogenous `d1` with instruments `z1` and `z2`

```
ivqregress iqr y x1 (d1 = z1 z2)
```

As above, but estimate the 0.75 quantile

```
ivqregress iqr y x1 (d1 = z1 z2), quantile(0.75)
```

As above, but estimate the 0.1, 0.2, ..., 0.9 quantiles

```
ivqregress iqr y x1 (d1 = z1 z2), quantile(10(10)90)
```

Use the smoothed estimating equation estimator to estimate the 0.6 quantile regression of `y` on exogenous `x1` and endogenous `d1` and `d2` with instruments `z1` and `z2`

```
ivqregress smooth y x1 (d1 d2 = z1 z2), quantile(0.6)
```

As above, but estimate the 0.1, 0.2, ..., 0.9 quantiles

```
ivqregress smooth y x1 (d1 d2 = z1 z2), quantile(10(10)90)
```

Advanced options for inverse quantile regression estimator

Use 50 grid points in the inverse quantile regression to estimate the IV 0.5 and 0.75 quantile regression

```
ivqregress iqr y x1 (d1 = z1 z2), ngrid(50) quantile(50 75)
```

As above, but construct grid points between 1 and 5 for all the quantiles

```
ivqregress iqr y x1 (d1 = z1 z2), ngrid(50) quantile(50 75) bound(1 5)
```

As above, but construct grid using different bounds for different quantiles

```
ivqregress iqr y x1 (d1 = z1 z2), ngrid(50) quantile(50 75)          \\\
              bound(1 5, at(50)) bound(2 6, at(75))
```

Advanced options for smoothed estimating equation estimator

Use 2 as the bandwidth in the smoothed estimating equation estimator to estimate the IV 0.5 and 0.75 quantile regression

```
ivqregress smooth y x1 (d1 d2 = z1 z2), quantile(50 75) bwidth(2)
```

As above, but use different bandwidths for different quantiles

```
ivqregress smooth y x1 (d1 d2 = z1 z2), quantile(50 75)          \\\
              bwidth(2, at(50)) bwidth(1, at(75))
```

8 Post-estimation of ivqregress

8.1 Overview

The following postestimation commands are of particular interest after `ivqregress`.

Commands	Description
`estat coefplot`	plot coefficients and their confidence intervals at different quantiles
`estat endogeffects`	process test of no effect, constant effect, stochastic dominance, and endogeneity
*`estat dualci`	provide the dual-confidence interval for the endogenous variables
*`estat waldplot`	plot Wald statistics corresponding to each grid point

Note :

- `estat waldplot` and `estat dualci` are allowed only after `ivqregress iqr`.

The following postestimation commands are also available after `ivqreg`

Commands	Description
`contrast`	contrasts and ANOVA-style joint tests of estimates
`estat summarize`	summary statistics for the estimation sample
`estat vce`	variance-covariance matrix of the estimators (VCE)
`estimates`	cataloging estimation results
`forecast`	dynamic forecasts and simulations
`lincom`	point estimates, standard errors, testing, and inference for linear combinations of coefficients
`margins`	marginal means, predictive margins, marginal effects, and average marginal effects
`marginsplot`	graph the results from margins (profile plots, interaction plots, etc.)
`nlcom`	point estimates, standard errors, testing, and inference for nonlinear combinations of coefficients
`predict`	predictions and their SEs, residuals, etc.
`predictnl`	point estimates, standard errors, testing, and inference for generalized predictions
`pwcompare`	pairwise comparisons of estimates
`test`	Wald tests of simple and composite linear hypotheses
`testnl`	Wald tests of nonlinear hypotheses

8.2 Syntax of `estat coefplot`

`estat coefplot` plots the estimated coefficients and their confidence intervals after `ivqreg`.

The syntax is

`estat coefplot` [*varname*] [, *options*]

Notes:

1. *varname* specifies the coefficient for the variable in the model. By default, it plots the coefficients for the first endogenous variables specified in the model.

2. The options are

Options	Description
Main	
`noci`	do not plot the confidence intervals
`no2sls`	do not plot the 2SLS estimates
Scatter plot	
marker_options	change look of markers (color, size, etc.)
connect_options	change look of lines or connecting method
CI plot	
`ciopts(`*area_options*`)`	affect rendition of the pointwise confidence interval
Reference line	
`lineopts(`*cline_options*`)`	affect rendition of reference line identifying the 2SLS estimates
Others	
twoway_options	titles, legends, axes, added lines and text, regions, name, aspect ratio, etc.

where *marker_options* and *connect_options* are defined in `scatter`; *area_options* and *twoway_options* are defined `rarea` and `twoway`, respectively.

8.3 Syntax of `estat endogeffects`

`estat endogeffects` tests the quantile process with different hypotheses for the coefficients on the endogenous treatment variables. In particular, `estat endogeffects` provides a test for four different hypotheses. The null hypotheses are

1. The hypothesis of no effect: the treatment has no effect on the outcome.
2. The hypothesis of constant effect: the treatment effect does not vary at different quantiles.
3. The dominance hypothesis: the effects are unambiguously beneficial.
4. The exogeneity hypothesis: the treatment is exogenous.

The syntax is

`estat endogeffects` [*varlist*] [, *options*]

where *varlist* is the endogenous variables. By default, *varlist* refers to all the endogenous variables specified in `ivqregress` when fitting the model.

The options are

Options	Description
`level(#)`	confidence level of a test; default is 0.95
`reps(#)`	perform # bootstrap replications; default is reps(100)
`rseed(#)`	set random-number seed to #
`all`	test four hypotheses; the default
`noeffect`	test of no effect
`constant`	test of constant effect
`dominance`	test of stochastic dominance
`exogeneity`	test of exogeneity

8.4 Syntax of `estat waldplot`

`estat waldplot` plots the Wald statistic corresponding to each grid point after `ivqregress iqr`.

The syntax is

`estat waldplot` [, *options*]

Notes:

1. The options are the following:

Options	Description
Main	
`quantile(#)`	plot Wald statistics for the #-th quantile estimation
`level(#)`	set confidence level; default is `level(95)`
Scatter plot	
marker_options	change look of markers (color, size, etc.)
connect_options	change look of lines or connecting method
Dual CI plot	
`ciopts(`*area_options*`)`	affect rendition of the dual confidence interval plot
Others	
twoway_options	titles, legends, axes, added lines and text, regions, name, aspect ratio, etc.

where *marker_options* and *connect_options* are defined in `scatter`; *area_options* and *twoway_options* are defined `rarea` and `twoway`, respectively.

8.5 Syntax of `estat dualci`

`estat dualci` computes the dual confidence interval for the endogenous variables after `ivqregress iqr`. `estat dualci` implements the dual confidence interval method proposed in Chernozhukov and Hansen (2008). The dual confidence interval is robust to the weak instruments, and it is usually wider than the traditional CI.

The syntax is

estat dualci [, level(#) *display_options*]

display_options: noci, nopvalues, noomitted, vsquish, noemptycells, baselevels, allbaselevels, nofvlabel, fvwrap(#), fvwrapon(*style*), cformat(*%fmt*), pformat(*%fmt*), sformat(*%fmt*), and nolstretch; see [R] Estimation options.

8.6 Syntax of predict

The syntax for predict is

predict [*type*] *newvar* [*if*] [*in*] [, equation(*eqno*) *statistic*]

statistic	Description
Main	
xb	linear prediction; the default
residuals	residuals

These statistics are available both in and out of sample; type predict ... if e(sample) ... if wanted only for the estimation sample.

The options are

- xb, the default, calculates the linear prediction.
- residuals calculates the residuals, that is, y - xb.
- equation(*eqno*) specifies the equation to which you are making the calculation.

  equation() is filled in with one *eqno* for the xb and residuals options. equation(#1) would mean that the calculation is to be made for the first equation, equation(#2) would mean the second, and so on. You could also refer to the equations by their names. equation(q10) would refer to the equation named q10 and equation(q50) to the equation named q50.

  If you do not specify equation(), results are the same as if you had specified equation(#1).

8.7 Syntax of margins

Syntax for margins

margins [*marginlist*] [, *options*]

margins [*marginlist*] , predict(*statistic* ...) [*options*]

statistic	Description
`xb`	linear prediction; the default
`residuals`	not allowed with margins

Statistics not allowed with margins are functions of stochastic quantities other than `e(b)`. For the full syntax, see [R] `margins`.

9 Examples

We want to estimate the effect of 401(k) participation (`p401k`) on different conditional quantiles of net financial assets (`asset`). We use data reported by Chernozhukov and Hansen (2004). These data are from a sample of households in 1990 Survey of Income and Program Participation (SIPP). For the head of household we have data on: income (`income`), age (`age`), number of people in the family (`familysize`), years of education (`educ`), marital status (`married`), whether to participate in the IRA (`ira`), whether to receive pension benefit (`pension`), and whether own a home (`ownhome`).

We suspect the 401(k) participation is endogenous because it may depend on unobserved factors such as saving preference that also impacts financial assets. Using the 401(k) eligibility (`e401k`) as an instrument for the 401(k) participation, we use `ivqregress` to estimate how the `p401k` affect the entire range of `asset`'s conditional distribution. One concern about using `e401k` as an instrument is that choosing to work for a company that offers the 401(k) program is not randomly assigned. Poterba et al. (1995) suggest that after conditioning on the income, we can take working for a company that offers the 401(k) plan as exogenous.

The IVQR model we want to estimate is

$$\texttt{asset}_i = \texttt{p401k}_i \alpha(U) + \mathbf{covariates}_i' \boldsymbol{\beta}(U), \tag{38}$$

where the distribution of U conditional on instrument `e401k` and covariates is assumed to be uniform between 0 and 1. The covariates are the continuous variables `income`, `age`, `familysize`, and `educ`, and the categorical variables `i.married`, `i.ira`, `i.pension`, and `i.ownhome`. As discussed above, `e401k` is the instrument for `p401k`. The coefficients $\alpha(U)$ and $\boldsymbol{\beta}(U)$ are random because they depend on the unobserved random variable U, which is uniformly distributed. In practice, U can be considered a ranking variable for the asset. When U is set to a fixed level τ, we are estimating an IVQR model at a specific quantile index τ. For example, when $\tau = 0.5$, we estimate how the 401(k) participation affects the median of net financial assets conditional on other covariates.

There are two objectives in the analysis:

1. Estimate the conditional quantile function of the two potential outcomes, which are the net financial asset when everyone or no one in the population participates in the 401(k) plan, respectively. In particular, the τ-th conditional quantile of the asset when everyone participates in the 401(k) plan is

$$\texttt{asset}_{\text{participate in 401(k)}} = \alpha(\tau) + \mathbf{covariates}' \boldsymbol{\beta}(\tau)$$

Similarly, the τ-th conditional quantile of the asset when no one participate in the 401(k) plan is

$$\texttt{asset}_{\text{no 401(k)}} = \textbf{covariates}'\boldsymbol{\beta}(\tau)$$

2. Estimate the quantile treatment effects of 401k participation on net financial assets. By definition, the τ-th quantile treatment effect is

$$\texttt{asset}_{\text{participate in 401(k)}} - \texttt{asset}_{\text{no 401(k)}} = \alpha(\tau)$$

   Thus, the coefficient $\alpha(\tau)$ can fully summarize the quantile treatment effect of `p401k` on `asset`.

We will show four examples that use `ivqregress` to estimate the quantile treatment effect of 401(k) participation on net financial assets.

- Example 1 estimates the median treatment effect of `p401k` on `asset` using the inverse quantile regression estimator. We will illustrate the use of `ivqregress` for estimation, how to interpret the estimates, use `estat dualci` to obtain confidence interval robust to the weak instrument, and how to use `margins` to get the conditional quantile function of the potential outcome.

- Example 2 is similar to Example 1 except that we use the smoothed estimating equation estimator for estimation. We compare the results between Example 1 and 2 and explain the difference.

- In Example 3, we use `ivqregress` to estimate the IVQR model at a range of quantile indexes, use `estat coefplot` to visualize quantile treatment effects, and use `estat endogeffects` to test some hypotheses of particular interest in the context of the IVQR model.

- Finally, Example 4 takes a closer look at the optimization procedure underhood the IQR estimator, and uses `estat waldplot` to diagnose the IQR estimator if a non-convergence issue emerges.

9.1 Example 1: IV median regression with the IQR estimator

In this example, we use the inverse quantile regression estimator to estimate the effect of 401(k) participation on the conditional median of the net financial asset.

```
. use assets2, clear
(Excerpt from Chernozhukov and Hanson (2004) Rev. of Economics and Statistics)
```

```
. ivqregress iqr assets (i.p401k = i.e401k) income age familysize        ///
>         i.married i.ira i.pension i.ownhome educ

Initial grid
    quantile = 0.50: .........10.........20.........30

Adaptive grid
    quantile = 0.50: .........10.........20.........30

IV median regression                                  Number of obs =   9,913
Estimator: Inverse quantile regression                Wald chi2(9)  = 1289.75
                                                      Prob > chi2   =  0.0000

------------------------------------------------------------------------------
             |               Robust
      assets | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
     1.p401k |   5313.397   573.2818     9.27   0.000     4189.786    6437.009
      income |   .1577512   .0124889    12.63   0.000     .1332735    .1822289
         age |   99.96526   8.561923    11.68   0.000      83.1842    116.7463
  familysize |  -197.8251   54.36773    -3.64   0.000    -304.3838   -91.26627
             |
     married |
    Married  |  -1359.124   227.3366    -5.98   0.000    -1804.696   -913.5528
             |
         ira |
        Yes  |   22629.61   1022.706    22.13   0.000     20625.15    24634.08
             |
     pension |
Receives ..  |  -693.8347   210.6176    -3.29   0.001    -1106.638   -281.0317
             |
     ownhome |
        Yes  |  -30.29657   154.7265    -0.20   0.845     -333.555    272.9618
        educ |  -96.43983   32.09465    -3.00   0.003    -159.3442   -33.53547
       _cons |  -4998.673   570.1315    -8.77   0.000     -6116.11   -3881.236
------------------------------------------------------------------------------
Endogenous: 0b.p401k 1.p401k
 Exogenous: income age familysize 0b.married 1.married 0b.ira 1.ira
            0b.pension 1.pension 0b.ownhome 1.ownhome educ 0b.e401k 1.e401k

. estimates store est_iqr
```

We specify `iqr` to use the inverse quantile regression. The dependent variable is `asset`. The endogenous variable `i.p401k` and the instrument `i.e401k` are specified in parenthesis, the other covariates follow as a regular `varlist`. `ivqregress` estimate the IV median regression by default.

The coefficient for `p401k` is \$5313. It means participation in 401(k) would increase the median of net financial asset by \$5313 conditional on other covariates, relative to a senario where no one participates. We store the estimation result as `est_iqr` for later use.

After `ivqregress iqr`, we can also use `estat dualci` to obtain the dual confidence interval robust to weak instrument for the coefficient on the endogenous variables. The dual confidence interval is usually wider than the regular confidence interval, but it provides a more robust inference if the instrument is weak. In this example, we see that dual 95% CI is [\$3684, \$7305], which is wider than the regular 95% CI [\$4178, \$6449].

```
. estat dualci

Dual confidence interval                              Number of obs = 9,913
```

assets	Coefficient	Robust std. err.	z	P>|z|	Dual [95% conf. interval]	
p401k						
1	5313.397	573.2818	9.27	0.000	3683.916	7304.986

The coefficients on each variable summarize the quantile treatment effects of the respective variable on the net financial asset. If we want to know the exact quantity of the conditional median for each potential outcome, we need to use `margins`. In particular, we want to know the median of financial assets when everyone does or does not participate in 401(k) conditional on other covariates. We specify `i.p401k` right after `margins` to tell `margins` to obtain median of the asset under 401(k) participation or no participation. The option `at` specifies the values of other covariates when computing the median. In particular, the continuous variables such as `income`, `age`, `familysize`, `educ` are fixed at the sample mean, and people are assumed to be married, participate in the IRA, receive pension benefits, and own a home.

```
. margins i.p401k, at((mean) income age familysize educ   ///
>         married = 1 ira = 1 pension = 1 ownhome = 1)

Adjusted predictions                                  Number of obs = 9,913
Model VCE: Robust

Expression: Linear prediction, predict()
At: income     =  37208.4 (mean)
    age        = 41.05891 (mean)
    familysize = 2.865328 (mean)
    married    =        1
    ira        =        1
    pension    =        1
    ownhome    =        1
    educ       = 13.20629 (mean)
```

	Margin	Delta-method std. err.	z	P>|z|	[95% conf. interval]	
p401k						
0	23681.37	1007.612	23.50	0.000	21706.49	25656.26
1	28994.77	1123.076	25.82	0.000	26793.58	31195.96

The results show that the conditional median of assets when everyone participates in 401(k) is \$28,995. In contrast, the conditional median of assets when no one participates in 401(k) is only \$23,681. The difference between these two medians is \$5313, which is the quantile treatment effect of `p401k` and is the same as the coefficient's value.

9.2 Example 2: IV median regression with the SEE estimator

In this example, we use `ivqregress` to estimate the IV median regression as in Example 1 but using the smoothed estimating equation estimator. We type `smooth` after `ivqregress` to use this estimator. The model specification is the same as in Example 1. The estimation result is stored as `est_smooth` for later use.

```
. ivqregress smooth assets (i.p401k = i.e401k) income age familysize
>    ///
>         i.married i.ira i.pension i.ownhome educ

Fitting smoothed IV quantile regression ...

Quantile = .5
Step 1:   bandwidth =  1302.9736    GMM criterion Q(b) =  2.617e-08
Step 2:   bandwidth =  6079.6881    GMM criterion Q(b) =  2.391e-12
Step 3:   bandwidth =  1438.3068    GMM criterion Q(b) =  8.068e-13

IV median regression                                 Number of obs =   9,913
Estimator: Smoothed estimating equations             Wald chi2(9)  = 1243.05
                                                     Prob > chi2   =  0.0000

------------------------------------------------------------------------------
             |               Robust
      assets | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
     1.p401k |   5364.468   573.3728     9.36   0.000     4240.678    6488.258
      income |   .1679934    .013419    12.52   0.000     .1416925    .1942942
         age |   113.6318   9.352867    12.15   0.000     95.30052    131.9631
  familysize |  -228.7766   57.61072    -3.97   0.000    -341.6916   -115.8617
             |
     married |
    Married  |   -1362.56   238.5988    -5.71   0.000    -1830.205   -894.9153
             |
         ira |
        Yes  |   22402.04   1043.504    21.47   0.000     20356.81    24447.27
             |
     pension |
Receives ..  |   -713.996    220.476    -3.24   0.001    -1146.121   -281.8709
             |
     ownhome |
        Yes  |  -12.71396   161.3703    -0.08   0.937     -328.994    303.5661
        educ |  -102.2889   34.18527    -2.99   0.003    -169.2908   -35.28701
       _cons |  -5672.645   619.7049    -9.15   0.000    -6887.244   -4458.045
------------------------------------------------------------------------------
Endogenous: 0b.p401k 1.p401k
 Exogenous: income age familysize 0b.married 1.married 0b.ira 1.ira
            0b.pension 1.pension 0b.ownhome 1.ownhome educ 0b.e401k 1.e401k

. estimates store est_smooth
```

The interpretation of the coefficient estimates is the same as in Example 1. For example, the coefficient for `p401k` is \$5364. So participation in 401(k) would increase the median of net financial assets by \$5364 conditional on other covariates, relative to a senario where no one participates.

Now we can compare the coefficient on `p401k` between the SEE estimator and the IQR estimator in Example 1.

```
. estimates table est_iqr est_smooth, keep(i.p401k) se

------------------------------------------
    Variable |   est_iqr     est_smooth
-------------+----------------------------
       p401k |
           1 |  5313.3974     5364.468
             |  573.28183    573.37279
------------------------------------------
                              Legend: b/se
```

We see that the point estimate in these two estimators are similar but not the same. It is normal to see the different results between IQR and SEE estimators because these two estimators approximate the original exact estimating equation in different ways. On the one hand, the IQR estimator tries to find the solution by an exhaust grid search. The estimation result critically depends on the range and finesse of grid points. On the other hand, the SEE estimator uses a Kernel method to smooth the original estimating equation. Its result depends on how well the smoothed estimating equation approximates the original equation, mainly controlled by the bandwidth.

Both the IQR and SEE estimators have their advantage and weakness. The IQR estimator is numerically stable, and it allows computing the dual confidence interval robust to weak instruments (use `estat dualci`). However, the IQR becomes computational intensive when there is more than one endogenous variable. Thus, `ivqregress iqr` allows only one endogenous variable. In contrast, the SEE estimator can handle multiple endogenous variables within a reasonable computation time. However, it does not allow `estat dualci` for robust inference. In practice, if there is only one endogenous variable in the model, we recommend using both estimators, comparing the results, and using IQR estimator as a benchmark as it can provide valid inference even the instrument is weak. If there is more than one endogenous variable, we can only use `ivqregress smooth`.

As in Example 1, we can use `margins` to obtain the conditional quantile function of the potential outcome.

```
. margins i.p401k, at((mean) income age familysize educ   ///
>         married = 1 ira = 1 pension = 1 ownhome = 1)

Adjusted predictions                                    Number of obs = 9,913
Model VCE: Robust

Expression: Linear prediction, predict()
At: income     =  37208.4 (mean)
    age        = 41.05891 (mean)
    familysize = 2.865328 (mean)
    married    =        1
    ira        =        1
    pension    =        1
```

```
    ownhome    =        1
    educ       = 13.20629 (mean)
```

	Margin	Delta-method std. err.	z	P>|z|	[95% conf. interval]	
p401k						
0	23550.11	1026.286	22.95	0.000	21538.63	25561.6
1	28914.58	1142.314	25.31	0.000	26675.69	31153.47

The results show that the conditional median of assets when everyone participates in 401(k) is \$28,915. In contrast, the conditional median of assets when no one participates in 401(k) is only \$23,550.

9.3 Example 3: IVQR at different quantiles

In the first two examples, we estimate the 401(k) participation (`p401k`) treatment effect on the conditional median of net financial assets (`asset`). From the policy designer's point of view, we may be more interested in estimating the treatment effect of `p401k` on other conditional quantiles of `asset`. For example, we can ask questions like 1) how the 401(k) participation affect the lower quantile of asset 2) are the 401(k) participation unambiguously beneficial for both lower and upper conditional quantiles of asset. In addition, we also want to know whether the 401(k) participation is endogenous in our model. In this example, we will show how to use `ivqregress` to estimate the IVQR model at different quantiles and how to use the post-estimation tools to answer the above questions.

First, we use the IQR estimator to estimate the model at different quantiles. In particular, we specify option `quantile(10(10)90)` to estimate the IVQR model at the 10th, 20th, ..., 90th quantiles.

```
. ivqregress iqr assets (i.p401k = i.e401k) income age familysize        ///
>         i.married i.ira i.pension i.ownhome educ, quantile(10(10)90)

Initial grid
    quantile = 0.10: .........10.........20.........30
    quantile = 0.20: .........10.........20.........30
    quantile = 0.30: .........10.........20.........30
    quantile = 0.40: .........10.........20.........30
    quantile = 0.50: .........10.........20.........30
    quantile = 0.60: .........10.........20.........30
    quantile = 0.70: .........10.........20.........30
    quantile = 0.80: .........10.........20.........30
    quantile = 0.90: .........10.........20.........30

Adaptive grid
    quantile = 0.10: .........10.........20.........30
    quantile = 0.20: .........10.........20.........30
    quantile = 0.30: .........10.........20.........30
```

```
   quantile = 0.40: .........10.........20.........30
   quantile = 0.50: .........10.........20.........30
   quantile = 0.60: .........10.........20.........30
   quantile = 0.70: .........10.........20.........30
   quantile = 0.80: .........10.........20.........30
   quantile = 0.90: .........10.........20.........30

IV quantile regression                                Number of obs =   9,913
Estimator: Inverse quantile regression                Wald chi2(81) = 5121.46
                                                      Prob > chi2   =  0.0000
```

assets	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
q10						
1.p401k	3240.08	475.6184	6.81	0.000	2307.885	4172.275
income	.0303072	.0123138	2.46	0.014	.0061725	.0544419
age	131.5908	15.13725	8.69	0.000	101.9223	161.2592
familysize	-329.2838	123.4665	-2.67	0.008	-571.2737	-87.29385
married						
Married	-1504.648	380.0373	-3.96	0.000	-2249.508	-759.7886
ira						
Yes	7864.15	344.2198	22.85	0.000	7189.492	8538.809
pension						
Receives ..	63.88643	326.6017	0.20	0.845	-576.2412	704.0141
ownhome						
Yes	969.6861	300.4319	3.23	0.001	380.8503	1558.522
educ	-301.1635	52.02897	-5.79	0.000	-403.1384	-199.1885
_cons	-7455.806	1192.112	-6.25	0.000	-9792.302	-5119.311
q20						
1.p401k	3446.347	334.4227	10.31	0.000	2790.89	4101.803
income	.0730526	.0083383	8.76	0.000	.0567097	.0893954
age	117.4232	10.09197	11.64	0.000	97.64336	137.2031
familysize	-311.518	77.31495	-4.03	0.000	-463.0525	-159.9835
married						
Married	-1023.821	245.1418	-4.18	0.000	-1504.29	-543.3515
ira						
Yes	8728.335	389.1614	22.43	0.000	7965.593	9491.077
pension						
Receives ..	8.450874	211.6773	0.04	0.968	-406.429	423.3307
ownhome						
Yes	247.107	193.7332	1.28	0.202	-132.6031	626.817
educ	-198.8986	36.72851	-5.42	0.000	-270.8852	-126.9121
_cons	-6064.723	762.7341	-7.95	0.000	-7559.654	-4569.791
q30						
1.p401k	3674.434	318.7578	11.53	0.000	3049.68	4299.188
income	.0900328	.0080609	11.17	0.000	.0742337	.1058319
age	106.4245	8.502461	12.52	0.000	89.76	123.089

familysize	-217.4507	57.05271	-3.81	0.000	-329.272	-105.6295
married						
Married	-1021.046	209.9362	-4.86	0.000	-1432.513	-609.5787
ira						
Yes	11974.65	566.1735	21.15	0.000	10864.98	13084.33
pension						
Receives ..	-149.6646	187.9519	-0.80	0.426	-518.0435	218.7144
ownhome						
Yes	118.1594	157.3384	0.75	0.453	-190.2182	426.537
educ	-122.6666	32.44148	-3.78	0.000	-186.2508	-59.08249
_cons	-5631.287	617.0855	-9.13	0.000	-6840.752	-4421.821
q40						
1.p401k	4196.127	369.6983	11.35	0.000	3471.532	4920.722
income	.1230721	.0104554	11.77	0.000	.1025799	.1435643
age	93.83839	7.963134	11.78	0.000	78.23093	109.4458
familysize	-225.3647	51.96267	-4.34	0.000	-327.2097	-123.5198
married						
Married	-1191.624	211.8386	-5.63	0.000	-1606.82	-776.4277
ira						
Yes	16997.44	803.6711	21.15	0.000	15422.28	18572.61
pension						
Receives ..	-511.7032	194.9456	-2.62	0.009	-893.7895	-129.6168
ownhome						
Yes	102.3659	148.1471	0.69	0.490	-187.997	392.7288
educ	-112.4069	30.98445	-3.63	0.000	-173.1353	-51.67845
_cons	-4787.913	553.4111	-8.65	0.000	-5872.579	-3703.247
q50						
1.p401k	5313.397	573.2818	9.27	0.000	4189.786	6437.009
income	.1577512	.0124889	12.63	0.000	.1332735	.1822289
age	99.96526	8.561923	11.68	0.000	83.1842	116.7463
familysize	-197.8251	54.36773	-3.64	0.000	-304.3838	-91.26627
married						
Married	-1359.124	227.3366	-5.98	0.000	-1804.696	-913.5528
ira						
Yes	22629.61	1022.706	22.13	0.000	20625.15	24634.08
pension						
Receives ..	-693.8347	210.6176	-3.29	0.001	-1106.638	-281.0317
ownhome						
Yes	-30.29657	154.7265	-0.20	0.845	-333.555	272.9618
educ	-96.43983	32.09465	-3.00	0.003	-159.3442	-33.53547
_cons	-4998.673	570.1315	-8.77	0.000	-6116.11	-3881.236
q60						
1.p401k	7006.205	801.4258	8.74	0.000	5435.439	8576.97

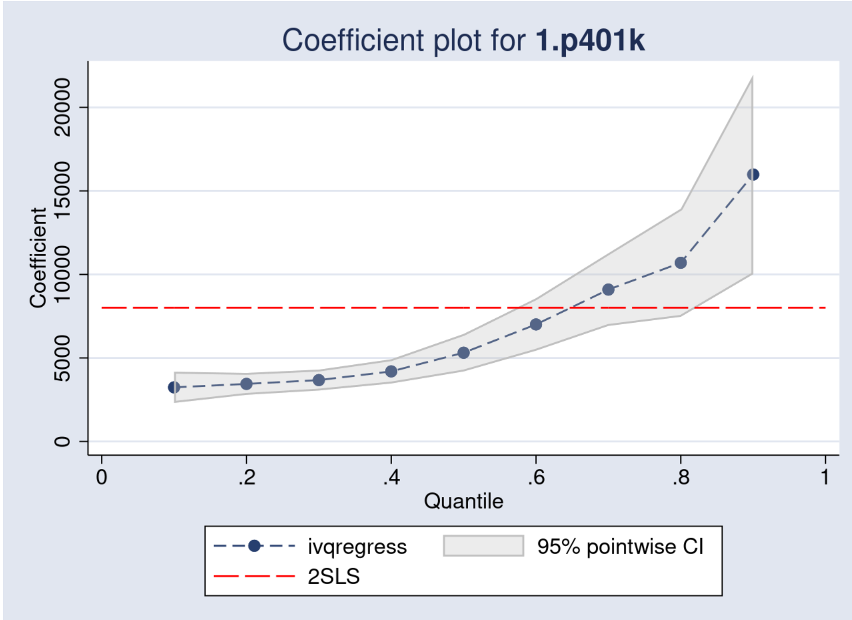
income	.2327564	.0174037	13.37	0.000	.1986458	.2668671
age	135.4321	11.38565	11.89	0.000	113.1166	157.7475
familysize	-262.5927	65.82424	-3.99	0.000	-391.6058	-133.5795
married						
Married	-1716.762	269.9874	-6.36	0.000	-2245.927	-1187.596
ira						
Yes	30301.55	1241.557	24.41	0.000	27868.15	32734.96
pension						
Receives ..	-988.7325	261.4987	-3.78	0.000	-1501.261	-476.2044
ownhome						
Yes	-122.2135	193.9046	-0.63	0.529	-502.2595	257.8324
educ	-118.7153	40.43096	-2.94	0.003	-197.9585	-39.47208
_cons	-6290.287	688.2098	-9.14	0.000	-7639.154	-4941.421
q70						
1.p401k	9093.469	1109.745	8.19	0.000	6918.408	11268.53
income	.3459585	.0226207	15.29	0.000	.3016228	.3902942
age	191.2876	16.53737	11.57	0.000	158.875	223.7003
familysize	-242.6605	86.05014	-2.82	0.005	-411.3157	-74.00534
married						
Married	-2470.874	352.4949	-7.01	0.000	-3161.751	-1779.996
ira						
Yes	39365.32	1608.07	24.48	0.000	36213.56	42517.08
pension						
Receives ..	-1796.514	344.5429	-5.21	0.000	-2471.806	-1121.222
ownhome						
Yes	-4.058645	262.5795	-0.02	0.988	-518.7051	510.5878
educ	-143.4298	51.12786	-2.81	0.005	-243.6386	-43.22104
_cons	-8637.647	928.462	-9.30	0.000	-10457.4	-6817.894
q80						
1.p401k	10699.12	1651.062	6.48	0.000	7463.098	13935.14
income	.5103271	.0293056	17.41	0.000	.4528892	.5677649
age	280.7892	24.10894	11.65	0.000	233.5366	328.0419
familysize	-400.8973	117.2019	-3.42	0.001	-630.6089	-171.1858
married						
Married	-2902.662	480.5005	-6.04	0.000	-3844.426	-1960.899
ira						
Yes	48875.79	2297.873	21.27	0.000	44372.04	53379.54
pension						
Receives ..	-3072.814	502.6944	-6.11	0.000	-4058.077	-2087.551
ownhome						
Yes	235.2409	402.9808	0.58	0.559	-554.5869	1025.069
educ	-130.398	68.51364	-1.90	0.057	-264.6823	3.886266
_cons	-11871.24	1257.402	-9.44	0.000	-14335.7	-9406.775

```
q90          |
     1.p401k |   15983.42   3046.028     5.25   0.000     10013.32    21953.53
      income |   .8247356   .0570029    14.47   0.000      .713012    .9364593
         age |   485.8734   48.99224     9.92   0.000     389.8504    581.8965
  familysize |  -646.4962    185.913    -3.48   0.001    -1010.879   -282.1134
             |
     married |
    Married  |  -3265.007   753.4701    -4.33   0.000    -4741.782   -1788.233
             |
         ira |
        Yes  |   68543.44   4952.261    13.84   0.000     58837.18    78249.69
             |
     pension |
Receives ..  |  -4656.177   869.4887    -5.36   0.000    -6360.343    -2952.01
             |
     ownhome |
        Yes  |   400.1957   680.2776     0.59   0.556     -933.124    1733.515
        educ |    48.4205   106.2844     0.46   0.649    -159.8931    256.7341
       _cons |  -20594.85   2260.983    -9.11   0.000     -25026.3   -16163.41
------------------------------------------------------------------------------
Endogenous: 0b.p401k 1.p401k
 Exogenous: income age familysize 0b.married 1.married 0b.ira 1.ira
            0b.pension 1.pension 0b.ownhome 1.ownhome educ 0b.e401k 1.e401k
```

The results show the estimates for the effect of 401(k) participation on each conditional quantile of the asset. The interpretation of the coefficient is similar as in Example 1 except we are looking at different conditional quantiles. For example, for the Equation `q30`, the estimates for coefficient on `p401k` is \$3674, thus participation of 401(k) would increase the 30th conditional quantile of net financial assets by \$3674.

In addition to looking at the exact numerical estimates from the coefficient table, we can also use `estat coefplot` to visualize the trend of `p401k`'s treatment effect from the lower to upper quantile. We specify `1.p401k` after `estat coefplot` to only show the coefficient plot for the variable `1.p401k`.

```
. estat coefplot 1.p401k
```



The dots in the plot show the point estimates of `p401k`'s treatment effect on different conditional quantiles of `asset`, and the grey bound show the 95% confidence interval. We see that there is an upward trend of `p401k`'s treatment effect. At lower level quantiles such as the 10th, 20th, ..., 40th quantiles, the treatment effect is relatively flat. However, we see the treatment effect increase significantly in the upper-level quantiles.

`estat coefplot` is a good way to visualize the treatment effect's trend. If we want to test some hypotheses regarding the trend and the model statistically, we can use `estat endogeffects`. For example, we are interested in testing the following hypotheses.

- **No effect**: The 401(k) participation does not affect net financial asset for all the estimated quantiles;
- **Constant effect**: The 401(k) participation's treatment effect is constant for the different conditional quantiles of asset;
- **Dominance**: The 401(k) participation is unambiguously beneficial for all the estimated quantiles of asset;
- **Exogeneity**: The 401(k) participation is exogenous.

```
. estat endogeffects

Tests for endogenous effects          Replications = 100
─────────────────────────────────────────────────────────
Null hypothesis  │     KS statistic    95% Critical value
─────────────────┼───────────────────────────────────────
No effect        │           11.271                 2.658
Constant effect  │            5.395                 2.650
Dominance        │            0.000                 2.390
Exogeneity       │            4.145                 2.386
─────────────────────────────────────────────────────────
Note: If the KS statistic < critical value, there is
      insufficient evidence to reject the null
      hypothesis.
```

`estat endogeffects` show the Kolmogorov-Smirnov statistic and the 95% critical value for each hypothesis. Therefore, we reject the null hypothesis if the test statistic is greater than the critical value. Otherwise, we can not reject the null hypothesis.

In particular, we see that the test statistics are greater than the critical values in testing the hypotheses of no effect, constant effect, and exogeneity. Thus, with 95% confidence level, we can reject these three hypotheses. In other words, we accept the hypotheses that the 401(k) participation has some effect, treatment is not constant across different quantiles, and 401(k) participation is endogenous. In contrast, we can not reject the dominance hypothesis. Thus, we accept the hypothesis that 401(k) participation is unambiguously beneficial for all the estimated quantiles of assets.

The test results are consistent with the result of the coefficient plot produced by `estat coefplot`, where we see that the treatment effects are positive (dominance and no effect hypotheses) and upward trended (constant effect hypothesis).

Finally, for reference, we can also use the SEE estimator to estimate the model.

```
. ivqregress smooth assets (i.p401k = i.e401k) income age familysize      ///
>          i.married i.ira i.pension i.ownhome educ, quantile(10(10)90)

Fitting smoothed IV quantile regression ...

Quantile = .1
Step 1:   bandwidth =  1327.0069    GMM criterion Q(b) =  9.224e-11
Step 2:   bandwidth =  1311.3131    GMM criterion Q(b) =  1.995e-10

Quantile = .2
Step 1:   bandwidth =  1272.5204    GMM criterion Q(b) =  2.089e-10
Step 2:   bandwidth =  1237.7195    GMM criterion Q(b) =  3.075e-19

Quantile = .3
Step 1:   bandwidth =  1504.4065    GMM criterion Q(b) =  5.407e-13
Step 2:   bandwidth =  1486.4224    GMM criterion Q(b) =  1.136e-10

Quantile = .4
Step 1:   bandwidth =  1362.7753    GMM criterion Q(b) =  5.511e-17
Step 2:   bandwidth =  1362.6479    GMM criterion Q(b) =  8.561e-16

Quantile = .5
Step 1:   bandwidth =  1302.9736    GMM criterion Q(b) =  2.617e-08
```

```
Step 2:   bandwidth =  6079.6881    GMM criterion Q(b) =  2.391e-12
Step 3:   bandwidth =  1438.3068    GMM criterion Q(b) =  8.068e-13

Quantile = .6
Step 1:   bandwidth =  1533.5129    GMM criterion Q(b) =  2.679e-18
Step 2:   bandwidth =  1520.1182    GMM criterion Q(b) =  1.141e-19

Quantile = .7
Step 1:   bandwidth =  2044.8617    GMM criterion Q(b) =  1.391e-10
Step 2:   bandwidth =  1977.2482    GMM criterion Q(b) =  1.827e-11

Quantile = .8
Step 1:   bandwidth =  2503.7256    GMM criterion Q(b) =  3.623e-10
Step 2:   bandwidth =  2458.6714    GMM criterion Q(b) =  2.317e-10

Quantile = .9
Step 1:   bandwidth =  3560.2178    GMM criterion Q(b) =  4.301e-12
Step 2:   bandwidth =  3529.3557    GMM criterion Q(b) =  2.932e-10

IV quantile regression                             Number of obs =   9,913
Estimator: Smoothed estimating equations           Wald chi2(81) = 4932.84
                                                   Prob > chi2   =  0.0000
```

assets	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
q10						
1.p401k	3191.667	486.2193	6.56	0.000	2238.695	4144.639
income	.0318585	.0123707	2.58	0.010	.0076124	.0561046
age	128.9268	15.42632	8.36	0.000	98.69178	159.1618
familysize	-329.8374	125.4774	-2.63	0.009	-575.7687	-83.90615
married						
Married	-1480.013	386.4611	-3.83	0.000	-2237.463	-722.5635
ira						
Yes	7914.049	342.9506	23.08	0.000	7241.878	8586.22
pension						
Receives ..	-5.356704	334.9869	-0.02	0.987	-661.919	651.2056
ownhome						
Yes	1043.279	308.722	3.38	0.001	438.1945	1648.363
educ	-289.8807	53.06713	-5.46	0.000	-393.8904	-185.8711
_cons	-7631.313	1214.725	-6.28	0.000	-10012.13	-5250.496
q20						
1.p401k	3503.744	338.8383	10.34	0.000	2839.633	4167.854
income	.0737261	.0084716	8.70	0.000	.057122	.0903302
age	114.9688	10.38239	11.07	0.000	94.61965	135.3179
familysize	-277.8925	78.67289	-3.53	0.000	-432.0885	-123.6964
married						
Married	-1160.725	253.6528	-4.58	0.000	-1657.876	-663.5752
ira						
Yes	8799.905	388.3753	22.66	0.000	8038.703	9561.106
pension						
Receives ..	-33.33779	218.144	-0.15	0.879	-460.8921	394.2165

ownhome						
Yes	386.2308	201.4194	1.92	0.055	-8.543996	781.0057
educ	-194.0516	37.98876	-5.11	0.000	-268.5082	-119.595
_cons	-6264.968	792.0489	-7.91	0.000	-7817.356	-4712.581
q30						
1.p401k	3754.908	320.9631	11.70	0.000	3125.832	4383.984
income	.0939826	.0083408	11.27	0.000	.0776348	.1103303
age	103.8314	8.712147	11.92	0.000	86.75593	120.9069
familysize	-250.4947	59.95479	-4.18	0.000	-368.0039	-132.9855
married						
Married	-1028.643	217.4311	-4.73	0.000	-1454.8	-602.4861
ira						
Yes	12008.63	563.5555	21.31	0.000	10904.08	13113.18
pension						
Receives ..	-179.5281	192.0513	-0.93	0.350	-555.9418	196.8855
ownhome						
Yes	195.7323	162.634	1.20	0.229	-123.0246	514.4891
educ	-134.7013	33.4085	-4.03	0.000	-200.1808	-69.22189
_cons	-5536.814	637.1917	-8.69	0.000	-6785.686	-4287.941
q40						
1.p401k	4326.754	371.7419	11.64	0.000	3598.153	5055.354
income	.1288469	.0105877	12.17	0.000	.1080955	.1495983
age	99.89601	8.289602	12.05	0.000	83.64869	116.1433
familysize	-231.3411	53.94265	-4.29	0.000	-337.0668	-125.6155
married						
Married	-1212.951	216.8328	-5.59	0.000	-1637.935	-787.966
ira						
Yes	16874.38	801.2841	21.06	0.000	15303.89	18444.86
pension						
Receives ..	-493.1742	198.6221	-2.48	0.013	-882.4663	-103.8821
ownhome						
Yes	105.4536	152.7777	0.69	0.490	-193.9852	404.8925
educ	-114.4753	32.09266	-3.57	0.000	-177.3758	-51.57484
_cons	-5216.625	581.4362	-8.97	0.000	-6356.219	-4077.031
q50						
1.p401k	5364.468	573.3728	9.36	0.000	4240.678	6488.258
income	.1679934	.013419	12.52	0.000	.1416925	.1942942
age	113.6318	9.352867	12.15	0.000	95.30052	131.9631
familysize	-228.7766	57.61072	-3.97	0.000	-341.6916	-115.8617
married						
Married	-1362.56	238.5988	-5.71	0.000	-1830.205	-894.9153
ira						
Yes	22402.04	1043.504	21.47	0.000	20356.81	24447.27

pension						
Receives ..	-713.996	220.476	-3.24	0.001	-1146.121	-281.8709
ownhome						
Yes	-12.71396	161.3703	-0.08	0.937	-328.994	303.5661
educ	-102.2889	34.18527	-2.99	0.003	-169.2908	-35.28701
_cons	-5672.645	619.7049	-9.15	0.000	-6887.244	-4458.045
q60						
1.p401k	6964.18	799.1829	8.71	0.000	5397.811	8530.55
income	.2422267	.0180009	13.46	0.000	.2069457	.2775078
age	145.0532	11.88882	12.20	0.000	121.7515	168.3549
familysize	-271.8402	68.28584	-3.98	0.000	-405.678	-138.0024
married						
Married	-1790.19	278.6729	-6.42	0.000	-2336.379	-1244.001
ira						
Yes	30029.76	1251.554	23.99	0.000	27576.76	32482.76
pension						
Receives ..	-1063.919	269.4261	-3.95	0.000	-1591.984	-535.8533
ownhome						
Yes	-79.57029	198.2018	-0.40	0.688	-468.0387	308.8981
educ	-128.4236	41.84504	-3.07	0.002	-210.4384	-46.40883
_cons	-6708.442	714.0485	-9.39	0.000	-8107.951	-5308.932
q70						
1.p401k	9002.846	1108.915	8.12	0.000	6829.412	11176.28
income	.3555392	.0229067	15.52	0.000	.310643	.4004354
age	203.3279	17.59732	11.55	0.000	168.8378	237.818
familysize	-314.0023	89.13006	-3.52	0.000	-488.694	-139.3106
married						
Married	-2396.634	359.7017	-6.66	0.000	-3101.636	-1691.631
ira						
Yes	38962.04	1621.653	24.03	0.000	35783.66	42140.42
pension						
Receives ..	-1882.868	352.3168	-5.34	0.000	-2573.396	-1192.34
ownhome						
Yes	-19.74676	271.0796	-0.07	0.942	-551.053	511.5595
educ	-163.099	52.90533	-3.08	0.002	-266.7915	-59.40646
_cons	-8753.875	954.0692	-9.18	0.000	-10623.82	-6883.934
q80						
1.p401k	10658.02	1665.467	6.40	0.000	7393.765	13922.27
income	.5172628	.0298155	17.35	0.000	.4588255	.5757
age	293.9692	24.78395	11.86	0.000	245.3935	342.5448
familysize	-407.2737	119.8248	-3.40	0.001	-642.1259	-172.4214
married						
Married	-3077.77	491.0029	-6.27	0.000	-4040.118	-2115.422
ira						

```
         Yes |   48410.11   2296.042    21.08   0.000     43909.95    52910.27
             |
     pension |
 Receives .. |  -3049.023   515.0161    -5.92   0.000    -4058.436    -2039.61
             |
     ownhome |
         Yes |   272.4814   412.1642     0.66   0.509    -535.3457    1080.308
        educ |  -131.0776   69.87187    -1.88   0.061    -268.0239    5.868784
       _cons |  -12294.73   1284.692    -9.57   0.000    -14812.68   -9776.781
-------------+----------------------------------------------------------------
q90          |
     1.p401k |   15525.23   3035.965     5.11   0.000     9574.848    21475.61
      income |   .8311508   .0574108    14.48   0.000     .7186277    .9436738
         age |   486.9876   51.61654     9.43   0.000      385.821    588.1541
  familysize |  -586.2617   193.5936    -3.03   0.002    -965.6983   -206.8252
             |
     married |
     Married |  -3877.165   781.2296    -4.96   0.000    -5408.347   -2345.983
             |
         ira |
         Yes |   67888.86   4902.106    13.85   0.000     58280.91    77496.81
             |
     pension |
 Receives .. |  -4829.506   898.9147    -5.37   0.000    -6591.346   -3067.665
             |
     ownhome |
         Yes |   715.6272   722.8727     0.99   0.322    -701.1773    2132.432
        educ |    14.5293   110.8781     0.13   0.896    -202.7878    231.8464
       _cons |  -19953.21   2326.698    -8.58   0.000    -24513.45   -15392.96
------------------------------------------------------------------------------

Endogenous: 0b.p401k 1.p401k
 Exogenous: income age familysize 0b.married 1.married 0b.ira 1.ira
            0b.pension 1.pension 0b.ownhome 1.ownhome educ 0b.e401k 1.e401k
```

As seen above, after `ivqregress smooth`, we can use `estat coefplot` to visualize the treatment effect and `estat endogeffects` to test some hypotheses of particular interest in the context of the IVQR model. To save space, we will not list the result here.

9.4 Example 4: Diagnose the IQR estimator

In this example, we will take a closer look at the IQR estimator and show how to diagnose the non-convergence issue if it happens. Let us first briefly talk about the algorithm in the IQR estimator in the context of the 401(k) example. Intuitively, the IQR estimator can be divided into the following steps.

1. Define a series of $A = \{\alpha_j\}_{j=1}^{J}$, where J is the number of grid points and it can be specified via option `ngrid()`, and α_j is a candidate solution for the coefficient on `1.p401k`.

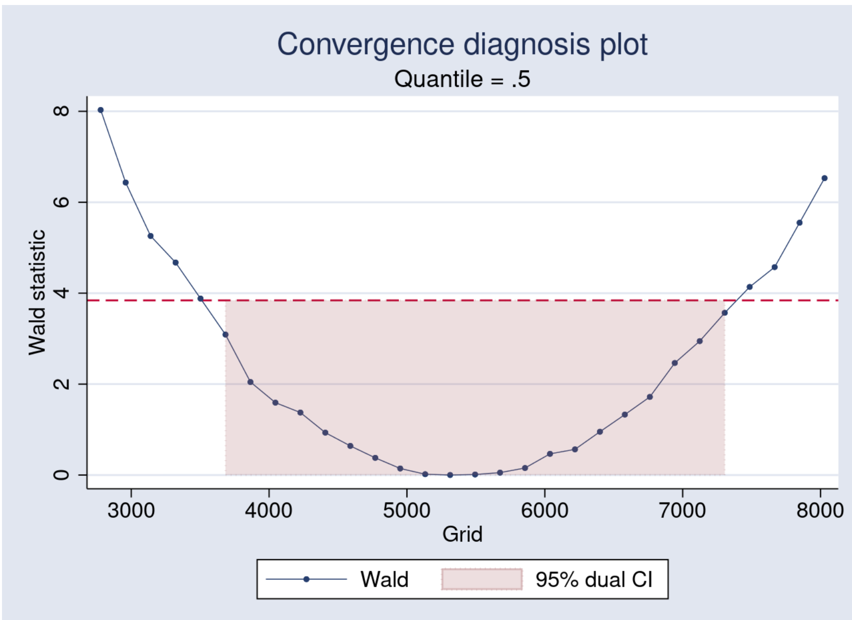
2. For each α_j, run the quantile regression of $\texttt{asset} - \texttt{1.p401k} * \alpha_j$ on the covariates and some transformations of instruments. Denote $\boldsymbol{\gamma}_j$ as the coefficients on the instruments

and $\mathbf{W}_j$ as the Wald statistic for $\boldsymbol{\gamma}_j$.

3. The IQR solution for α (the coefficients on `1.p401k`) is the α_j such that $\boldsymbol{\gamma}_j$ is closest to zero. In other words, the solution chooses α_j such that the Wald statistic $\mathbf{W}_j$ is the smallest.

We can use `estat waldplot` to visualize the above procedure. Using the estimation result in Example 1, we first restore the result `est_iqr` and then use `estat waldplot` to draw the Wald statistics corresponding to each grid points.

```
. estimates restore est_iqr
(results est_iqr are active now)

. estat waldplot, name(a)
```



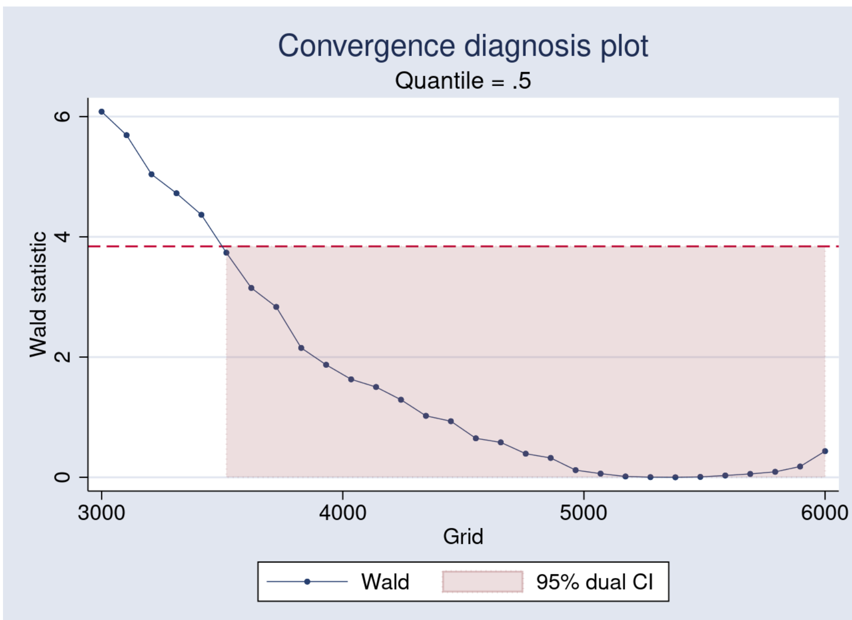
The horizontal axis shows the grid points for α, and the vertical axis shows the values of Wald statistics. Each dot in the plot shows the Wald statistic corresponding to each grid point. The red dashed line is the 95% critical value of the Wald test. Thus, only the Wald statistics below the red dashed line will not reject the hypothesis that $\boldsymbol{\gamma}_j$ equals zero. Respectively, the 95% dual confidence interval is the α's such that the Wald statistics are below the critical value. See Example 1 for the use of `estat dualci` to show the numerical values of the dual CI.

By default, `ivqregress iqr` uses the dual CI to generate the lower and upper bound for the grid points to make sure that the grid covers the true value of parameter α with a big probability. Sometimes, we may want to customize the bounds. For example, suppose we want to search grid points between 3000 and 6000. We can use option `bound()` for this purpose.

```
. cap noi ivqregress iqr assets (i.p401k = i.e401k) income age familysize  ///
>         i.married i.ira i.pension i.ownhome educ, bound(3000 6000)

Initial grid
    quantile = 0.50: .........10.........20.........30

convergence not achieved
    The grid interval should be wider than the 95% dual confidence interval.
    Try to set a wider bound using option bound(). Use estat waldplot for
    diagnosis.
```

We see that `ivqregress` errors out with "convergence not achieved". The reason is that the specified bound is too narrow to cover the true value of the parameter with a 95% probability. We can use `estat waldplot` to further visualize the issue.

```
. estat waldplot, name(b)
```



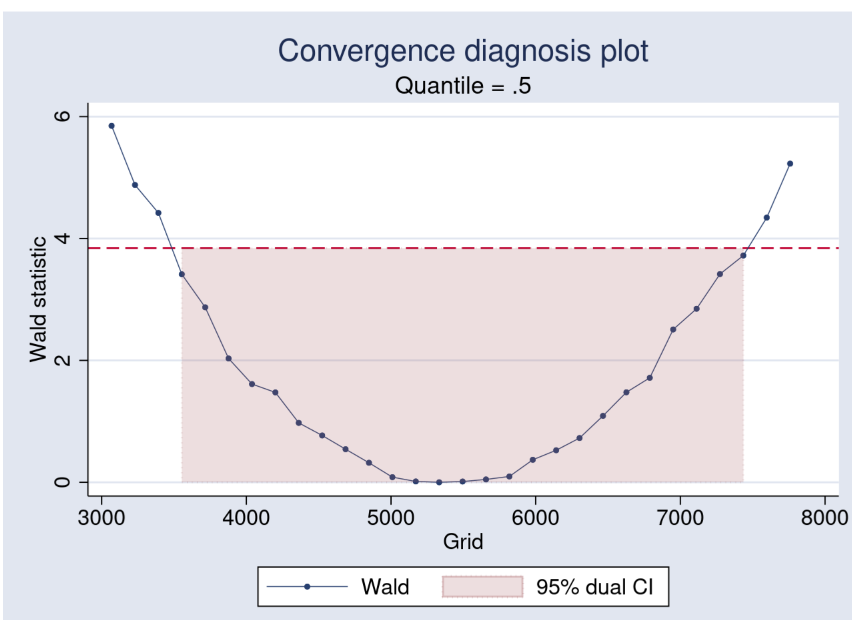
The graph shows that the upper bound 6000 is too small because we need the Wald statistics to intersect with the 95% critical value at both the lower and upper bound. Now, we can increase the upper bound and see if the IQR estimator converges. For example, we increase the upper bound to 8000.

```
. cap noi ivqregress iqr assets (i.p401k = i.e401k) income age familysize  ///
>         i.married i.ira i.pension i.ownhome educ, bound(3000 8000)

Initial grid
    quantile = 0.50: .........10.........20.........30

Adaptive grid
    quantile = 0.50: .........10.........20.........30

IV median regression                                   Number of obs =    9,913
```

```
Estimator: Inverse quantile regression                     Wald chi2(9)  = 1290.41
                                                           Prob > chi2   =  0.0000

------------------------------------------------------------------------------
             |               Robust
      assets | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
     1.p401k |   5332.937   574.5175     9.28   0.000     4206.903    6458.971
      income |    .157381    .012478    12.61   0.000     .1329246    .1818374
         age |   99.78981   8.553978    11.67   0.000     83.02432    116.5553
  familysize |  -199.6165    54.3519    -3.67   0.000    -306.1442   -93.08872
             |
     married |
     Married |  -1351.309   227.0824    -5.95   0.000    -1796.382   -906.2357
             |
         ira |
         Yes |   22631.85   1022.023    22.14   0.000     20628.72    24634.98
             |
     pension |
 Receives .. |  -694.1447    210.533    -3.30   0.001    -1106.782   -281.5077
             |
     ownhome |
         Yes |  -30.67158   154.6947    -0.20   0.843    -333.8676    272.5244
        educ |  -96.30363    32.0715    -3.00   0.003    -159.1626   -33.44465
       _cons |  -4983.758   569.4043    -8.75   0.000     -6099.77   -3867.746
------------------------------------------------------------------------------
Endogenous: 0b.p401k 1.p401k
 Exogenous: income age familysize 0b.married 1.married 0b.ira 1.ira
            0b.pension 1.pension 0b.ownhome 1.ownhome educ 0b.e401k 1.e401k
```

Now, the IQR estimator converges. We can redraw the Wald plot to confirm that the proposed grid points interval is indeed wider than the dual confidence interval.

```
. estat waldplot, name(c)
```



A Appendix

A.1 Proof for Theorem 1

Proof.

$$
\begin{aligned}
P(Y \leq q(\mathbf{D}, \mathbf{X}, \tau) | \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}) &\overset{(1)}{=} P(q(\mathbf{D}, \mathbf{X}, U_{\mathbf{D}}) \leq q(\mathbf{D}, \mathbf{X}, \tau) | \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}) \\
&\overset{(2)}{=} P(U_{\mathbf{D}} \leq \tau | \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}) \\
&\overset{(3)}{=} \int P(U_{\mathbf{D}} \leq \tau | \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}, V = v) dP(V = v | \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}) \\
&\overset{(4)}{=} \int P(U_{\delta(\mathbf{x}, \mathbf{z}, v)} \leq \tau | \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}, V = v) dP(V = v | \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}) \\
&\overset{(5)}{=} \int P(U_{\mathbf{d}} \leq \tau | \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}, V = v) dP(V = v | \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}) \\
&\overset{(6)}{=} P(U_0 \leq \tau | \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}) \\
&\overset{(7)}{=} \tau
\end{aligned}
$$

Equality (1) by the definition of Y in A5. Equality (2) follows the decreasing feature of $q()$ in A1. Equality (3) is by the definition of conditional probability. Equality (4) is by the definition of $\mathbf{D}$ in A3. Equality (5) holds because, conditional on $(\mathbf{X}, \mathbf{Z}, v)$, $\mathbf{D}$ is a constant, and the distribution of $U_{\mathbf{d}}$ is identical. Here, we assume $\mathbf{D} = 0$ in this case. $\mathbf{D}$ can be any fixed value in $\mathbb{D}$. Equality (6) is by definition. Finally, equality (7) holds because $U_{\mathbf{d}} \sim U(0, 1)$ in A1 and $U_{\mathbf{d}}$ is independent of $\mathbf{X}$ and $\mathbf{Z}$ in A2.

$$
\begin{aligned}
P(U_{\mathbf{D}} \leq \tau | \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}) &= P(U_{\delta(\mathbf{X}, \mathbf{Z}, V)} \leq \tau | \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}) \\
&= \int P(U_{\delta(\mathbf{X}, \mathbf{Z}, v)} \leq \tau | \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}, V = v) dP(V = v | \mathbf{X}, \mathbf{Z}) \\
&= \int P(U_0 \leq \tau | \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}, V = v) dP(V = v | \mathbf{X}, \mathbf{Z}) \\
&= P(U_0 \leq \tau | \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z})
\end{aligned}
$$

So conditional on $\mathbf{X}$ and $\mathbf{Z}$, $U_{\mathbf{D}}$ has the same distribution as $U_{\mathbf{d}}$ for a fixed value of $\mathbf{d}$.

□

From the proof for Theorem 1, we see that rank similarity assumption in A4 is the fundamental assumption to convert the conditional distribution of $U_{\mathbf{D}}$ to $U_{\mathbf{d}}$ with fixed value $\mathbf{d}$. Here, we discuss the nuisance underlying this assumption.

First, rank invariance is a particular case of rank similarity. Namely, the rank invariance assumes $U_{\mathbf{d}}$ are the same across different values in $\mathbb{D}$. While convenient and supported by some applications, the rank invariance assumption may be too strong in practice.

Second, rank similarity means that given an assignment of treatment, the rank $U_{\mathbf{d}}$ is identically distributed. For example, among the individuals who have $\mathbf{X} = \mathbf{x}$, $\mathbf{Z} = \mathbf{z}$, and $D = 1$, the distribution of U_0 and U_1 are the same. In this formulation, we implicitly assume that one selects the treatment without knowing the potential outcomes. However, the individuals may know the distribution of the potential outcomes given $(\mathbf{X}, \mathbf{Z}, v)$ but not the exact value of U_0 for each observation.

A.2 Discussion on Theorem 2

For a rigorous proof for Theorem 2, see Chernozhukov and Hansen (2006). Here I provide a intuitive interpretation. IQR estimator approximately solves the GMM moment condition in Equation 8, and thus the variance of IQR estimator can be understood via the GMM approach. We take the first-order taylor expansion of the sample analog of moment 9 at the true value $\boldsymbol{\theta}(\cdot)$. The term, $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} l_i(\cdot, \boldsymbol{\theta}(\cdot))\boldsymbol{\Psi}_i(\cdot)$, is the moment condition at the true value. The term, $\mathbf{J}(\cdot)$ is the gradient of the moment with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. The term, $\mathbf{S}(\tau, \tau')$ is the variance for the moment

$$\mathbf{E}\left[l(\cdot)\boldsymbol{\Psi}(\cdot)l(\cdot)\boldsymbol{\Psi}(\cdot)'\right]$$

This variance has a block-by-block structure such that the block for the (τ, τ') is $\mathbf{S}(\tau, \tau')$.

$$\begin{aligned}
&\mathbf{E}\left[l(\tau)\boldsymbol{\Psi}(\tau)l(\tau')\boldsymbol{\Psi}(\tau')'\right] \\
&\quad = \mathbf{E}\left[l(\tau)l(\tau')\boldsymbol{\Psi}(\tau)\boldsymbol{\Psi}(\tau')'\right] \\
&\quad = \mathbf{E}\left\{\mathbf{E}\left[l(\tau)l(\tau')\boldsymbol{\Psi}(\tau)\boldsymbol{\Psi}(\tau')'|\mathbf{X},\mathbf{Z}\right]\right\} \\
&\quad = \mathbf{E}\left\{\mathbf{E}\left[l(\tau)l(\tau')|\mathbf{X},\mathbf{Z}\right]\boldsymbol{\Psi}(\tau)\boldsymbol{\Psi}(\tau')'\right\} \\
&\quad = \mathbf{E}\left\{\mathbf{E}\left[(\tau - \mathbb{1}(\epsilon(\tau) \leq 0))(\tau' - \mathbb{1}(\epsilon(\tau') \leq 0)|\mathbf{X},\mathbf{Z}\right]\boldsymbol{\Psi}(\tau)\boldsymbol{\Psi}(\tau')'\right\} \\
&\quad = \mathbf{E}\left\{\mathbf{E}\left[(\tau\tau' - \tau\mathbb{1}(\epsilon(\tau') \leq 0) - \mathbb{1}(\epsilon(\tau) \leq 0)\tau' + \mathbb{1}(\epsilon(\tau) \leq 0)\mathbb{1}(\epsilon(\tau') \leq 0)|\mathbf{X},\mathbf{Z}\right]\boldsymbol{\Psi}(\tau)\boldsymbol{\Psi}(\tau')'\right\}
\end{aligned}$$

notice that $\mathbf{E}(\mathbb{1}(\epsilon(\tau) \leq 0) = \tau$ by the main implications in IVQR model in Equation 7. So

$$= \mathbf{E}\left\{\mathbf{E}\left[(\tau\tau' - \tau\tau' - \tau\tau' + \mathbb{1}(\epsilon(\tau) \leq 0)\mathbb{1}(\epsilon(\tau' \leq 0))|\mathbf{X},\mathbf{Z}\right]\boldsymbol{\Psi}(\tau)\boldsymbol{\Psi}(\tau')'\right\}$$

smaller value of τ also means greater residuals ϵ, so $\mathbb{1}(\epsilon(\tau) \leq 0)\mathbb{1}(\epsilon(\tau' \leq 0))$ is simplified to $\min(\tau, \tau')$

$$
\begin{aligned}
&= \mathbf{E}\left\{\mathbf{E}\left[(\min(\tau,\tau') - \tau\tau')|\mathbf{X},\mathbf{Z}\right]\boldsymbol{\Psi}(\tau)\boldsymbol{\Psi}(\tau')'\right\} \\
&= (\min(\tau,\tau') - \tau\tau')\mathbf{E}\left[\boldsymbol{\Psi}(\tau)\boldsymbol{\Psi}(\tau')'\right]
\end{aligned}
$$

For the gradient, $\mathbf{J}(\cdot)$,

$$
\begin{aligned}
\mathbf{J}(\tau) &= \frac{\partial \mathbf{E}(l(\tau)\boldsymbol{\Psi}(\tau))}{\partial \boldsymbol{\theta}'} \\
&= \mathbf{E}\left[\mathbf{E}\left(\frac{\partial l(\tau)\boldsymbol{\Psi}(\tau)}{\partial \boldsymbol{\theta}'}|\mathbf{X},\mathbf{D},\mathbf{Z}\right)\right] \\
&= \mathbf{E}\left[\mathbf{E}\left(\frac{\partial l(\tau)}{\partial \boldsymbol{\theta}'}|\mathbf{X},\mathbf{D},\mathbf{Z}\right)\boldsymbol{\Psi}(\tau)\right] \\
&= \mathbf{E}\left[\frac{\partial \mathbf{E}(l(\tau)|\mathbf{X},\mathbf{D},\mathbf{Z})}{\partial \boldsymbol{\theta}'}\boldsymbol{\Psi}(\tau)\right] \\
&= \mathbf{E}\left[\left(\frac{\partial F_\epsilon(0|\mathbf{X},\mathbf{D},\mathbf{Z})}{\partial \boldsymbol{\theta}'}\right)\boldsymbol{\Psi}(\tau)\right] \\
&= \mathbf{E}\left[f_\epsilon(0|\mathbf{X},\mathbf{D},\mathbf{Z})\boldsymbol{\Psi}(\tau)[\mathbf{D}',\mathbf{X}']\right]
\end{aligned}
$$

A.3 Decentralization estimators

One major drawback of the IQR approach is that it suffers from the curse of dimensionality. When there are more than two endogenous variables, the grid search computing time becomes slow. Recently, Kaido and Wüthrich (2021) proposes to recast the original problem into finding a Nash equilibrium game solution. The resulting estimator is called the "decentralization estimator" because it iteratively divides the original problem into small sub-problems, a form of decentralization.

Another advantage of the decentralization estimator is that it does not require choosing extra tuning parameter as in the smoothing GMM approach.

Kaido and Wüthrich (2021) proposes three different decentralization algorithms: the simultaneous contraction-based algorithm, the sequential contraction-based algorithm, and the root-finding algorithm. The simulations in Kaido and Wüthrich (2021) show that the sequential contraction-based estimator is more stable and faster than the other two alternatives, so we will only focus on this method.

A.3.1 Sequential contraction-based algorithm

We start by partition the parameter vector $\theta = (\beta', \alpha')'$ into $J = k_d + 1$ subvectors, where $\theta_1 = \beta$ and $\theta_j = \alpha_{j-1}$ for $j = 2, \ldots, J$. Let θ_{-j} denote all the elements in θ except θ_j.

We define the following quantile regression objective functions:

$$Q_1 = \mathbf{E}\left[\rho_\tau\left(Y - X'\theta_1 - D_1\theta_2 - D_2\theta_3 - \ldots - D_{k_d}\theta_J\right)\right] \tag{39}$$

$$Q_j = \mathbf{E}\left[\rho_\tau\left(Y - X'\theta_1 - D_1\theta_2 - D_2\theta_3 - \ldots - D_{k_d}\theta_J\right)\frac{Z_{j-1}}{D_{j-1}}\right] \qquad \text{for } j = 2, \ldots, J \tag{40}$$

where $\rho_\tau(u) = u(\tau - \mathbb{1}(u < 0))$ is the "check function". We assume that Z_{j-1}/D_{j-1} is well defined and positive. In practice, we can transform the instruments Z and add a large constant to D to satisfy this condition. See Section A.3.2 for more discussion.

Suppose there are J players, and each player j has control over θ_j and take θ_{-j} as given. The players solve the following optimization problems:

$$\min_{\tilde{\theta}_1 \in R^{k_x}} Q_1(\tilde{\theta}_1, \theta_{-1}) \tag{41}$$

$$\min_{\tilde{\theta}_j \in R} Q_j(\tilde{\theta}_j, \theta_{-j}) \tag{42}$$

Notice that each problem is just an ordinary weighted quantile regression problem.

The solution for Q_j satisfies the first-order condition such that

$$\frac{\partial Q_1}{\partial \theta_1} = \mathbf{E}\left[(\tau - \mathbb{1}(Y - X'\theta_1 - D_1\theta_2 - D_2\theta_3 - \ldots - D_{k_d}\theta_J))X\right] = 0 \tag{43}$$

$$\frac{\partial Q_j}{\partial \theta_j} = \mathbf{E}\left[(\tau - \mathbb{1}(Y - X'\theta_1 - D_1\theta_2 - D_2\theta_3 - \ldots - D_{k_d}\theta_J))Z_{j-1}\right] = 0 \qquad j = 2, \ldots, J \tag{44}$$

These FOC conditions together form the moment conditions in Equation (**??**) implied by the linear IVQREG model.

Let $L_j(\theta_{-j})$ as the minimizer for Q_j. The solution satisfies

$$L_j(\theta^*_{-j}) = \theta^*_j$$

which means θ^* satisfies the moment condition in Equation (**??**) simultaneously. θ^* is also the Nash equilibrium of the game.

The sequential contraction-based algorithm is as follows.

Algorithm 3: Sequential contraction-based algorithm

1. Define the starting value $\theta^{(0)} = (\theta_1^{(0)}, \ldots, \theta_J^{(0)})$ as the solution of the two-stage-least-square estimator for the linear IV model.

2. Define the maximum number of iteration K.

3. Define the numerical tolerance level e_N.

4. For $k = 1$ to K, do loop

   (a) $\theta_1^{(k)} = L_1(\theta_{-1}^{(k-1)})$

   (b) For $j = 2, \ldots, J$, do loop

   $$\theta_j^{(k)} = M_j(\theta_{-1}^{(k)}) = L_j\left(\theta_1^{(k)}, \theta_2^{(k)}, \ldots, \theta_{j-1}^{(k)}, \theta_{\{-1,-2,\ldots,-(j-1)\}}^{(k-1)}\right)$$

   (c) Compute the relative difference between $\theta_{-1}^{(k)}$ and $\theta_{-1}^{(k-1)}$, denote it as δ.

   (d) If $\delta \leq e_N$ or $k > K$, exit the loop.

5. The solution is $\hat{\theta} = (\hat{\theta}_1, \theta_{-1}^{(k)})$, where $\hat{\theta}_1 = L_1(\theta_{-1}^{(k)})$.

To conduct inference, we can employ a regular bootstrap.

Algorithm 4: bootstrap the decentralized estimator

1. Using the original sample, estimate $\hat{\theta}$ as in Algorithm (3).

2. For $b = 1, \ldots, B$, do

   (a) Draw a bootstrap sample ${W_{i=1}^N}^{(b)}$ with replacement.

   (b) Using ${W_{i=1}^N}^{(b)}$ and Algorithm (3), estimate $\widetilde{\theta^{(b)}}$

3. Let

   $$F_B(x) = \frac{1}{B}\sum_{b=1}^{B} \mathbb{1}(\sqrt{N}(\widetilde{\theta^{(b)}} - \hat{\theta}) \leq x)$$

   Use $F_B(x)$ as an approximation to the distribution of $\sqrt{N}(\hat{\theta} - \theta^*)$

A.3.2 Reparameterization

The sequential contraction-based algorithm requires that D_j/Z_j is positive and well defined. Therefore, we need to reparameterize the model to satisfy this condition.

For the instruments, we can transform Z_j such as

$$\tilde{Z}_j = \exp Z_j/(1+\exp Z_j)$$

So $\tilde{Z}_j$ will always be positive. Given the original moment condition in Equation (**??**) , any transformation of Z_j should also satisfy this condition.

For the endogenous variable D_j, we can add to it a large constant to make the sum always positive. In particular, suppose $D_j \in (D_{min}, D_{max})$ for each $j = 1, \ldots, k_d$, we add D_j to a constant c such that $c > |D_{min}|$. Denote $\tilde{D}_j = D_j + c$.

Suppose the original model is

$$q(D, X, \tau) = \beta_0 + D'\alpha + X'\beta$$

After reparameterization, the model becomes

$$q(D, X, \tau) = \beta_0 - c\sum_{j=1}^{k_d} \alpha_j + (D+c)'\alpha + X'\beta$$

So the estimated constant is $\tilde{\beta}_0 = \hat{\beta}_0 - c\sum_{j=1}^{k_d} \hat{\alpha}_j$, and the estimate for the original constant term is $\hat{\beta}_0 = \tilde{\beta}_0 + c\sum_{j=1}^{k_d} \hat{\alpha}_j$.

A.3.3 Overidentified case

So far, we assume the model is just identified, that is $k_z = k_d$. When there are more instruments than the endogenous variables, we can use the instrument $\tilde{D}$, which is the linear projection of D on the space spanned by Z and X. Then the objective function in the sequential contraction-based algorithm becomes:

$$Q_1 = \mathbf{E}\left[\rho_\tau\left(Y - X'\theta_1 - D_1\theta_2 - D_2\theta_3 - \ldots - D_{k_d}\theta_J\right)\right]$$
$$Q_j = \mathbf{E}\left[\rho_\tau\left(Y - X'\theta_1 - D_1\theta_2 - D_2\theta_3 - \ldots - D_{k_d}\theta_J\right)\frac{\widetilde{D_{j-1}}}{D_{j-1}}\right] \qquad \text{for } j = 2, \ldots, J$$

where $\widetilde{D_{j-1}} = \Psi(\Psi'\Psi)^{-1}\Psi' D_{j-1}$, and $\Psi = (X, Z)$.

To achieve further efficiency gain, we can use the method in Remark 5 of Chernozhukov and Hansen (2006). In practice, this method also involves nonparametrically estimating the error term's density function at 0, which inevitably requires a choice of the bandwidth. For simplicity, we will not pursue this path.

References

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

Bofinger, E. 1975. Estimation of a Density Function using Order Statistics. *Australian Journal of Statistics* 17: 1–7.

de Castro, L., A. F. Galvao, D. M. Kaplan, and X. Liu. 2019. Smoothed GMM for quantile models. *Journal of Econometrics* 213: 121–144. URL `https://doi.org/10.1016/j.jeconom.2019.04.008`.

Chen, L. Y., and S. Lee. 2018. Exact computation of GMM estimators for instrumental variable quantile regression models. *Journal of Applied Econometrics* 33(4): 553–567.

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(3): 735–751.

———. 2005. an IV model of quantile treatment effects. *Econometrica* 73(1): 245–261.

———. 2006. Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics* 132(2): 491–525.

———. 2008. Instrumental variable quantile regression: A robust inference approach. *Journal of Econometrics* 142(1): 379–398.

Chernozhukov, V., and H. Hong. 2003. An MCMC approach to classical estimation. *Journal of Econometrics* 115(2): 293–346.

DasGupta, A. 2008. *Asymptotic Theory of Statistics and Probability*. Springer.

Hall, P., and S. J. Sheather. 1988. On the Distribution of a Studentized Quantile. *Journal of the Royal Statistical Society. Series B (Methodological)* 50: 381–391.

Kaido, H., and K. Wüthrich. 2021. Decentralization estimators for instrumental variable quantile regression models. *Quantitative Economics* 12(2): 443–475.

Kaplan, D. M., and Y. Sun. 2017. Smoothed estimating equations for instrumental variables quantile regression. *Econometric Theory* 33(1): 105–157.

Koenker, R. 2005. *Quantile Regression*. Econometric Society Monographs, Cambridge University Press. URL `https://books.google.com/books?id=hdkt7V4NXsgC`.

Koenker, R., and G. Bassett. 1978. Regression quantiles. *Econometrica: journal of the Econometric Society* 33–50.

Poterba, J. M., S. F. Venti, and D. A. Wise. 1995. Do 401(k) contributions crowd out other personal saving? *Journal of Public Economics* 58(1): 1–32.

Silverman, B. 1998. *Density estimation for statistics and data analysis*. Chapman & Hall/CRC.

Wand, M., and M. Jones. 1995. *Kernel Smoothing*. Chapman & Hall/CRC.