# Exact computation of GMM estimators for instrumental variable quantile regression models\*

#### Le-Yu Chen<sup>†</sup>

Institute of Economics, Academia Sinica

### Sokbae Lee<sup>‡</sup>

Department of Economics, Columbia University

Centre for Microdata Methods and Practice, Institute for Fiscal Studies

#### March 2017

#### Abstract

We show that the generalized method of moments (GMM) estimation problem in instrumental variable quantile regression (IVQR) models can be equivalently formulated as a mixed integer quadratic programming problem. This enables exact computation of the GMM estimators for the IVQR models. We illustrate the usefulness of our algorithm via Monte Carlo experiments and an application to demand for fish.

**Keywords**: generalized method of moments, instrumental variable, quantile regression, endogeneity, mixed integer optimization

**JEL codes**: C21, C26, C61, C63

<sup>\*</sup>This work was supported in part by the European Research Council (ERC-2014-CoG-646917-ROMIA) and by the British Academy (International Partnership and Mobility Scheme Grant, reference number PM140162).

<sup>†</sup>E-mail: lychen@econ.sinica.edu.tw ‡E-mail: sl3841@columbia.edu

## 1 Introduction

The instrumental variable quantile regression (IVQR) and related models have been increasingly popular for studying the impacts of possibly endogenous covariates on the distribution of the outcome of interest. See a recent review by Chernozhukov and Hansen (2013) and references therein for the latest developments in identification, estimation, and inference as well as the list of empirical applications.

The IVQR model admits conditional moment restrictions which can be used to construct the estimating equations for the GMM estimation of the model parameters. However, the sample counterparts of the IVQR estimating equations are discontinuous in the parameters so that the resulting GMM estimation problem becomes a non-convex and computationally non-trivial optimization problem. Honoré and Hu (2004) provided a heuristic for computing the IVQR GMM estimates. Chernozhukov and Hansen (2006) developed the inverse quantile regression (QR) estimator that is not directly a GMM estimator but can be shown to be asymptotically equivalent to the IVQR GMM estimator. Xu and Burer (2017) proposed an alternative algorithm for computing the inverse QR estimator. The Markov chain Monte Carlo (MCMC) based Laplace type estimator of Chernozhukov and Hong (2003) can also be used as an approximation of the IVQR GMM estimator but it requires careful tunning in the MCMC implementation. Kaplan and Sun (2015) proposed a smoothed estimating equation approach which facilitates the GMM computation problem but requires the choice of the smoothing parameter.

In this paper, we are concerned with exact computation of the GMM estimates of the IVQR parameters. As pointed out by Andrews (1997), heuristic algorithms for computation of GMM estimates that do not guarantee to find the exact global optimum or a specific level of approximation to the global optimum may result in extremum estimators which could exhibit statistical behavior that is quite different from that established by theory. This source of computational uncertainty may impact on the empirical results. Hence, as a complement to the previous work on the IVQR computation, our paper provides a method for exact computation of the IVQR estimates within the classical GMM framework.

Our computational algorithm is based on the method of mixed integer optimization (MIO). Specifically, we show that the IVQR GMM estimation problem can be equivalently formulated as a mixed integer quadratic programming (MIQP) problem. Thanks to the developments in MIO solution algorithms and fast computing environments, this reformulation allows us to solve for the exact GMM estimates by using the modern efficient MIO solvers. Well-known numerical solvers such as CPLEX and Gurobi can be used to effectively solve large-scale MIQP problems. See Bertsimas, King, and Mazumder (2016, Section 2.1) for discussions on computational advances in solving the MIO problems. See also

Florios and Skouras (2008), Kitagawa and Tetenov (2015), Bertsimas, King, and Mazumder (2016), and Chen and Lee (2016) for related but distinct work on solving non-convex optimization problems in statistics and econometrics via the MIO approach.

The rest of this paper is organized as follows. In Section 2, we summarize the setup of the IVQR model and the inverse quantile regression method of Chernozhukov and Hansen (2006). In Section 3, we present the MIQP formulation of the IVQR GMM estimation problem. We conduct a simulation study of the performance of the MIQP based GMM estimates in Section 4 and illustrate the application of our computation approach in a real data exercise concerning the demand estimation in Section 5. We then conclude the paper in Section 6.

## 2 The instrumental variable quantile regression model

Let Y be an outcome of interest. We consider the quantile regression model under endogeneity, which is characterized by the structural equation

$$Y = W'\theta(U), \tag{2.1}$$

where U is an unobserved scalar random variable and W = (D, X) is a vector of covariates. The covariates D may not be independent of U. We assume that there is a vector of instrumental variables, denoted as Z, which can be excluded from (2.1) but can influence the endogenous variables D such that  $\dim(Z) \geq \dim(D)$  and

$$U|X,Z \sim \text{Uniform}(0,1)$$
.

Assume that the function  $\theta(\cdot)$  in (2.1) is a measurable function such that the mapping  $\tau \mapsto W'\theta(\tau)$  is strictly increasing in  $\tau$  for almost every realization of W. Under these assumptions, it follows that

$$P(Y \le W'\theta(\tau)|X,Z) = P(U \le \tau|X,Z) = \tau. \tag{2.2}$$

Given a random sample,  $(Y_i, W_i, Z_i)_{i=1}^n$  of n observations, we are interested in the estimation of  $\theta(\tau)$  for some fixed values of  $\tau \in (0, 1)$ .

The model set forth so far is the well known linear IVQR model which has been studied by Chernozhukov and Hansen (2004, 2005, 2006, 2008), Chernozhukov, Hansen, and Jansson (2007, 2009), and Kaplan and Sun (2015) among many others. Note that, when there is no endogenous covariate, this model reduces to the conventional linear quantile regression model of Koenker and Bassett (1978) where W = X = Z. In the presence of endogeneity, Chernozhukov and Hansen (2005) provided further modeling assumptions such that

the quantile-specific parameter vector  $\theta(\tau)$  can be causally interpreted as the structural quantile effect in the setting with counterfactual outcomes.

Chernozhukov and Hansen (2006) developed primitive conditions for the identification of  $\theta(\tau)$ . They also provided an inverse QR algorithm for the estimation of  $\theta(\tau)$ . To describe their algorithm, write  $\theta = (\alpha, \beta)$  such that  $W'\theta(\tau) = D'\alpha(\tau) + X'\beta(\tau)$ . Let  $\Psi_i = \Psi(X_i, Z_i)$  be a vector of transformations of instruments with  $\dim(\Psi_i) \geq \dim(\alpha)$ . Let A be a given positive definite matrix. The Chernozhukov-Hansen inverse QR procedure proceeds as follows. Let

$$\widehat{\alpha}(\tau) \equiv \arg\inf_{\alpha \in \mathcal{A}} \widehat{\gamma}_{\tau}(\alpha)' A \widehat{\gamma}_{\tau}(\alpha), \qquad (2.3)$$

where

$$\left(\widehat{\beta}_{\tau}(\alpha), \widehat{\gamma}_{\tau}(\alpha)\right) \equiv \arg\inf_{(\beta, \gamma) \in \mathcal{B} \times \mathcal{G}} \frac{1}{n} \sum_{i=1}^{n} \rho_{\tau} \left(Y_{i} - D'_{i} \alpha - X'_{i} \beta - \Psi'_{i} \gamma\right), \tag{2.4}$$

 $\mathcal{A}$ ,  $\mathcal{B}$  and  $\mathcal{G}$  are compact parameter spaces, and the check function  $\rho_{\tau}$  is defined by  $\rho_{\tau}(u) = u(\tau - 1\{u < 0\})$  for  $u \in \mathbb{R}$ . The inverse QR estimator is then defined by

$$\widehat{\theta}(\tau) = \left(\widehat{\alpha}(\tau), \widehat{\beta}_{\tau}\left(\widehat{\alpha}(\tau)\right)\right).$$

In the procedure above, the function  $\Psi$  and the matrix A can vary across  $\tau$  and be replaced by their consistent estimates. Moreover, the QR objective function can be weighted across observations. See Chernozhukov and Hansen (2006) for further details.

For implementation, Chernozhukov and Hansen (2006) proposed to solve the outer optimization problem (2.3) by the grid search method. The inner optimization problem (2.4) is the standard quantile regression problem, which can be solved very efficiently. Thus, when  $\dim(\alpha)=1$ , the inverse QR method is computationally appealing because its implementation amounts to solving convex optimization sub-problems within a low-dimensional global search procedure. However, this computational merit diminishes rapidly with the increase of the number of endogenous variables. Instead of performing grid search, Xu and Burer (2017) proposed an alternative method to compute the inverse QR estimator. Their approach is based on exact minimization of the quadratic norm as in (2.3) subject to the optimality conditions for the linear programming formulation of the QR problem of (2.4). Xu and Burer (2017) showed that the resulting computational problem reduces to a quadratic programming problem subject to complementarity constraints for which they developed a branch-and-bound algorithm to compute the exact solution.

## 3 Exact computation of the GMM based IVQR estimator via the mixed integer optimization approach

The conditional moment restriction (2.2) can be used to form estimating equations for the GMM estimation of  $\theta(\tau)$ . That is,

$$E[(1\{Y \le W'\theta(\tau)\} - \tau)L] = 0, \tag{3.1}$$

where L is a vector of instruments consisting of functions of X and Z. As noted by Chernozhukov and Hansen (2006), the inverse QR estimator, which is not directly a GMM estimator, can be shown to be asymptotically equivalent to the GMM estimator with the instruments  $L_{\text{CH}} \equiv [X', \Psi(X, Z)']'$ .

In this paper, we provide an algorithm for directly computing the GMM based IVQR estimator using the orthogonality conditions (3.1). Let  $s_{\tau}(t)$  denote the vector  $(s_{\tau,i}(t))_{i=1}^n$ , where  $s_{\tau,i}(t) \equiv 1$   $\{Y_i \leq W_i't\} - \tau$  for  $i \in \{1, ..., n\}$ . Let G be the n-by-dim(L) matrix whose ith row vector is  $L_i'$ . Let  $\widehat{Q}$  be a given positive definite matrix of dimension dim(L). The GMM based IVQR estimator of  $\theta(\tau)$ , denoted by  $\widehat{\theta}_{GMM}(\tau)$ , is given by

$$\widehat{\theta}_{GMM}(\tau) = \arg\inf_{\theta \in \Theta} s_{\tau}(\theta)' G \widehat{Q} G' s_{\tau}(\theta), \qquad (3.2)$$

where  $\Theta$  is the compact parameter space of  $\theta$ .

We now present our computational algorithm, which is based on the method of mixed integer optimization. We note that the optimization problem (3.2) can be equivalently formulated as the following constrained mixed integer quadratic programming (MIQP) problem:

$$\inf_{e=(e_1,\dots,e_n),\theta\in\Theta} (e-\tau)' G\widehat{Q}G'(e-\tau)$$
(3.3)

subject to

$$e_i(-M_i - \epsilon) < Y_i - W_i'\theta \le (1 - e_i) M_i, \ i \in \{1, ..., n\},$$
 (3.4)

$$e_i \in \{0, 1\}, i \in \{1, ..., n\},$$
 (3.5)

where  $\epsilon$  is a given small and positive real scalar (e.g.  $\epsilon = 10^{-6}$  as in our simulation study), and

$$M_i \equiv \max_{\theta \in \Theta} |Y_i - W_i'\theta|, \ i \in \{1, ..., n\}.$$
 (3.6)

We now explain the equivalence between (3.2) and (3.3). Note that, for a given value of  $\theta \in \Theta$ , the sign constraints (3.4) and the dichotomization constraints (3.5) enforce that  $e_i = 1\{Y_i \leq W_i'\theta\}$  for  $i \in \{1,...n\}$ . Therefore, solving the constrained MIQP problem (3.3) is equivalent to solving the GMM estimation problem (3.2). This equivalence enables

us to employ the modern MIQP solvers to exactly compute the GMM estimator  $\widehat{\theta}_{GMM}(\tau)$ . For the implementation, note that the values  $(M_i)_{i=1}^n$  in the inequality constraints (3.4) can be computed by formulating the maximization problem in (3.6) as linear programming problems, which can be efficiently solved by modern optimization solvers. Hence these values can be easily computed and stored as the input to the MIQP formulation (3.3).

Remark 1. Our MIQP based computational approach can be used to find the exact global solution in the IVQR GMM estimation problem. Modern MIQP solvers employ branch-and-bound type algorithms which maintain along the solution process both the feasible solutions and lower bounds on the optimal objective function value. Therefore, for computationally demanding applications, this feature enables us to solve for an approximate IVQR GMM estimator with a guaranteed approximation error bound, thus facilitating the design of an early stopping rule as described in Chen and Lee (2016, Section 4.3).

We can perform inference on  $\theta(\tau)$  using the GMM estimator  $\widehat{\theta}_{GMM}(\tau)$ . As noted by Chernozhukov, Hansen, and Jansson (2009), we can take

$$\widehat{Q} = \left[\tau (1 - \tau) n^{-1} \sum_{i=1}^{n} L_i L_i'\right]^{-1}$$
(3.7)

as a convenient and natural choice of the GMM weight matrix. By (2.2), this weight matrix equals the inverse of the variance of  $n^{-1/2} \sum_{i=1}^{n} s_{\tau,i}(\theta(\tau)) L_i$  conditional on  $(L_i)_{i=1}^n$ . Let  $\varepsilon_{\tau} \equiv Y - W'\theta(\tau)$ . In the GMM estimation (3.2) with  $\hat{Q}$  given by (3.7), it is straightforward to establish via empirical process theory (see e.g., Pakes and Pollard, 1989) that

$$\sqrt{n}(\widehat{\theta}_{GMM}(\tau) - \theta(\tau)) \stackrel{d}{\longrightarrow} N(0, \Omega),$$
 (3.8)

where the asymptotic variance matrix  $\Omega$  is given by

$$\Omega = \tau (1 - \tau) \left[ \Sigma_{WL} \Sigma_{LL}^{-1} \Sigma_{WL}' \right]^{-1}, \Sigma_{WL} = E \left[ f_{\varepsilon_{\tau}} (0|W, Z) W L' \right], \Sigma_{LL} = E \left[ LL' \right].$$
 (3.9)

We can estimate  $\Sigma_{LL}$  by the sample analog  $\widehat{\Sigma}_{LL} \equiv n^{-1} \sum_{i=1}^{n} L_i L_i'$ . Let  $\widehat{\varepsilon}_{\tau,i} \equiv Y_i - W_i' \widehat{\theta}_{GMM}(\tau)$ . Following Powell (1986),  $\Sigma_{WL}$  can be consistently estimated by

$$\widehat{\Sigma}_{WL} \equiv n^{-1} \sum_{i=1}^{n} \left[ K \left( \widehat{\varepsilon}_{\tau,i} / h_n \right) / h_n \right] W_i L_i', \tag{3.10}$$

where  $K(\cdot)$  is a kernel function and  $h_n$  is a bandwidth sequence satisfying that  $h_n \longrightarrow 0$  and  $\sqrt{n}h_n \longrightarrow \infty$ . Specific rule-of-thumb choices of  $h_n$  can be based on Koenker (1994). See also Chernozhukov and Hansen (2006, Section 3.4) and Chernozhukov and Hansen (2008, Section 4.4) for the estimation of the IVQR variance components. Based on these results, it is therefore straightforward to construct the confidence interval estimates of

## 4 Simulation study

In this section, we study the performance of the GMM estimator  $\widehat{\theta}_{GMM}(\tau)$  in finite-sample simulations. We used the MATLAB implementation of the Gurobi Optimizer (version 7.0) to solve the MIQP problems for all numerical results of this paper. All computations were done on a desktop PC (Windows 7) equipped with 32 GB RAM and a CPU processor (Intel i7-5930K) of 3.5 GHz.

We generated n = 100 observations from the following simple location scale model:

$$Y = 1 + D_1 + D_2 + D_3 + (0.5 + D_1 + 0.25D_2 + 0.15D_3)\varepsilon,$$

$$D_1 = \Phi(Z_1 + v_1), D_2 = 2\Phi(Z_2 + v_2), D_3 = 1.5\Phi(Z_3 + v_3),$$
(4.1)

where  $\Phi$  denotes the cdf of the standard normal random variable,  $Z_1$ ,  $Z_2$  and  $Z_3$  are independent standard normal random variables, and  $(\varepsilon, v_1, v_2, v_3)$  is generated independently of  $(Z_1, Z_2, Z_3)$  from multivariate normal distribution with mean zero and variance 0.25V where

$$V = \begin{bmatrix} 1 & 0.4 & 0.6 & -0.2 \\ 0.4 & 1 & 0 & 0 \\ 0.6 & 0 & 1 & 0 \\ -0.2 & 0 & 0 & 1 \end{bmatrix}.$$

By Skorohod representation, we can rewrite the model (4.1) as

$$Y = \theta_0(U) + \theta_1(U)D_1 + \theta_2(U)D_2 + \theta_3(U)D_3,$$

where  $U = F_{\varepsilon}(\varepsilon)$  with  $F_{\varepsilon}$  being the cdf of the unobservable  $\varepsilon$ , and

$$\theta_0(\tau) = 1 + 0.5 F_{\varepsilon}^{-1}(\tau), \theta_1(\tau) = 1 + F_{\varepsilon}^{-1}(\tau), \theta_2(\tau) = 1 + 0.25 F_{\varepsilon}^{-1}(\tau), \theta_3(\tau) = 1 + 0.15 F_{\varepsilon}^{-1}(\tau).$$

We used 500 simulation repetitions for all simulation experiments. In the GMM estimation, we took  $W = (1, D_1, D_2, D_3)$  and  $L = (1, Z_1, Z_2, Z_3)$ . The GMM weight matrix  $\hat{Q}$  was constructed based on (3.7). We set the parameter space  $\Theta$  in the MIQP problem (3.3) to be the product of the intervals  $[\hat{\theta}_{j,2SLS} - 10\hat{\sigma}_{j,2SLS}, \hat{\theta}_{j,2SLS} + 10\hat{\sigma}_{j,2SLS}]$ , where for  $j \in \{0, 1, 2, 3\}$ ,  $\hat{\theta}_{j,2SLS}$  and  $\hat{\sigma}_{j,2SLS}$ , respectively denote the parameter estimate and its estimated heteroskedasticity-robust standard error from the two-stage least square

<sup>&</sup>lt;sup>1</sup>The MATLAB codes for the computation of  $\widehat{\theta}_{GMM}(\tau)$  are available from the authors. This implementation requires the Gurobi Optimizer, which is freely available for academic purposes.

regression of Y on the covariates W using L as the instruments. The value of  $\epsilon$  in (3.4) was set to be  $10^{-6}$ .

Table 1: MIQP computation time (CPU seconds)

$\overline{\tau}$	mean	min	median	max
0.25	94	37	92	197
0.5	348	104	333	989
0.75	86	33	84	186

We now present the simulation results. First, we report the computational performance of our MIQP algorithm for computing the IVQR GMM estimator. Table 1 gives the summary statistics of the MIQP computation time in CPU seconds across simulation repetitions. From this table, we can see that the MIQP problems (3.3) were solved very efficiently in these simulations which incorporated three endogenous covariates. For the two cases with  $\tau \in \{0.25, 0.75\}$ , the computation time was comparable. Both cases could be easily solved with the mean and median computation time not exceeding 100 seconds and the maximum time below 200 seconds. The case of  $\tau = 0.5$  appeared to be the most computationally demanding but its maximum time remained capped within 17 minutes.

Table 2: Finite-sample performance of the GMM estimator

	mean		median	
	bias	RMSE	bias	MAE
$\theta_0 (0.25)$	0.0109	0.2436	0.0012	0.1643
$\theta_1  (0.25)$	-0.0327	0.3554	-0.0048	0.2309
$\theta_2 (0.25)$	0.0003	0.1642	0.0031	0.1008
$\theta_3 (0.25)$	0.0064	0.2232	-0.0068	0.1522
$\theta_0 (0.5)$	0.0161	0.2498	-0.0037	0.1724
$\theta_1 (0.5)$	-0.0412	0.3241	-0.0316	0.2315
$\theta_2 (0.5)$	-0.0012	0.1561	0.0066	0.1038
$\theta_3 (0.5)$	0.0012	0.2047	0.0031	0.1396
$\theta_0 (0.75)$	0.0187	0.3046	0.0055	0.1849
$\theta_1 (0.75)$	-0.0358	0.3425	-0.0264	0.2235
$\theta_2 (0.75)$	-0.0022	0.1820	0.0062	0.1181
$\theta_3  (0.75)$	0.0035	0.2393	-0.0016	0.1538

We now study the statistical performance of the IVQR GMM estimator. In Table 2, we report the mean and median biases, root mean squared error (RMSE) and median absolute error (MAE) of the GMM estimators  $\hat{\theta}_{GMM}(\tau)$  for  $\tau \in \{0.25, 0.5, 0.75\}$ . From

these results, we find that the GMM estimators performed quite well in terms of estimation bias. Across the three quantile cases, the estimators for  $\theta_1(\tau)$  appeared to have larger dispersion in terms of both RMSE and MAE.

Table 3: Comparison with asymptotic approximation

	standard deviation	asymptotic
	in simulations	standard error
$\theta_0 (0.25)$	0.2434	0.2297
$\theta_1 (0.25)$	0.3539	0.3256
$\theta_2 (0.25)$	0.1642	0.1572
$\theta_3 (0.25)$	0.2231	0.2059
$\theta_0 (0.5)$	0.2493	0.2296
$\theta_1 (0.5)$	0.3215	0.3049
$\theta_2 (0.5)$	0.1561	0.1474
$\theta_3 (0.5)$	0.2047	0.1994
$\theta_0 (0.75)$	0.3040	0.2744
$\theta_1  (0.75)$	0.3406	0.3400
$\theta_2  (0.75)$	0.1820	0.1664
$\theta_3 (0.75)$	0.2393	0.2283

It is also interesting to assess how well the finite-sample behavior of the IVQR GMM estimator can be approximated by asymptotic theory. For this purpose, our exact GMM estimator can be used to eliminate the unquantified uncertainty on the solution inaccuracy that might emerge in a heuristic optimization procedure. In Table 3, we calculated the asymptotic standard error based on the formula (3.9) evaluated at true parameter values of the simulation design. This quantity was then compared to standard deviation of  $\hat{\theta}_{GMM}(\tau)$  in simulations. The results of Table 3 indicate that the finite-sample standard error of the GMM estimator in this simulation setup, though being slightly larger, can be well approximated by the asymptotic standard error.

In practice, for carrying out inference, the asymptotic variance of the GMM estimator has to be estimated. We used the Gaussian kernel in the estimation of  $\Sigma_{WL}$ . The bandwidth sequence  $h_n$  in (3.10) was based on the Hall-Sheather bandwidth choice, which was suggested by Koenker (1994) and also used by Chernozhukov, Hansen, and Jansson (2009). We also checked the sensitivity of the inference results with respect to this bandwidth choice. Specifically, we reported in Table 4 the finite-sample cover probabilities of the 95% confidence interval (CI) estimates for  $\theta(\tau)$ , which were constructed based on the normal approximation theory described in Section 3 with three different bandwidth choices:  $h_n \in \{0.8h_{n,HS}, h_{n,HS}, 1.2h_{n,HS}\}$ , where  $h_{n,HS}$  denotes the Hall-Sheather bandwidth sequence. From Table 4, we find that the coverage probabilities results were not

Table 4: Coverage probabilities (95% CI)

	$0.8h_{n,HS}$	$h_{n,HS}$	$1.2h_{n,HS}$
$\theta_0 (0.25)$	0.930	0.940	0.952
$\theta_1 (0.25)$	0.906	0.914	0.918
$\theta_2 (0.25)$	0.912	0.924	0.934
$\theta_3 (0.25)$	0.916	0.926	0.938
$\theta_0 (0.5)$	0.936	0.944	0.950
$\theta_1 (0.5)$	0.938	0.944	0.952
$\theta_2 \left( 0.5 \right)$	0.950	0.958	0.966
$\theta_3 (0.5)$	0.896	0.916	0.928
$\theta_0 \left( 0.75 \right)$	0.896	0.916	0.928
$\theta_1 (0.75)$	0.922	0.938	0.944
$\theta_2 \left( 0.75 \right)$	0.892	0.896	0.908
$\theta_3 \left( 0.75 \right)$	0.918	0.928	0.942

very sensitive across bandwidth values although the CI estimates were slightly undersized. We also notice that the CI estimates based on taking  $h_n = h_{n,HS}$  or  $h_n = 1.2h_{n,HS}$ performed quite well in terms of overall performance.

## 5 An illustrative empirical example: estimating the demand for fish

We illustrate usefulness of our method for exact computation of the IVQR GMM estimator in an empirical study of the demand for fish. We used the dataset constructed by Graddy (1995) on the transactions of whiting in the Fulton fish market in New York. The data were also previously studied in Chernozhukov and Hansen (2008) and Chernozhukov, Hansen, and Jansson (2009) to illustrate the econometric methods developed for quantile regression models with endogeneity. In what follows, we mainly focused on analyzing the results estimated by the MIQP approach and comparing them to the inverse QR estimation results.

The data consist of 111 observations on the price and quantity of whiting transactions aggregated by day. The outcome variable Y is the logarithm of total amount of whitings sold on each day and the endogenous explanatory variable D is the logarithm of the average daily price. The exogenous explanatory variables include the indicators (Monday, Tuesday, Wednesday and Thursday) for days of the week. The instrumental variables are indicators (Stormy and Mixed) for weather conditions at sea. These instruments capture the wave height and wind speed, which should affect the supplied quantity of fish and hence the price in the market but should not influence the demand for fish. See

Graddy (1995, 2006) for further details on the operation of the Fulton fish market, and the data and variables used for this study.

Following Chernozhukov, Hansen, and Jansson (2009), we considered the simple demand equation

$$Y = \theta_0(U) + \theta_1(U) D \tag{5.1}$$

for the estimation of  $\theta_1$ , the price elasticity of the demand, which may vary across the demand level U. We also augmented the specification (5.1) by incorporating the day effect variables as additional controls, and then performed the estimation. Table 5 presents the estimation results for  $\theta_1(\tau)$  under these two different specifications. For GMM estimation results, we took L=(1,Stormy,Mixed) as instruments and configurated the MIQP setting in the same fashion as in Section 4. We used the Gaussian kernel and the Hall-Sheather bandwidth choice for estimating the standard deviation of the GMM estimator and constructing the 95% CI for  $\theta_1(\tau)$ . We also performed some sensitivity check and found that the results were not very sensitive to the bandwidth choice. Moreover, we also extracted the inverse QR and the corresponding 95% asymptotic CI estimation results provided by Chernozhukov, Hansen, and Jansson (2009, Table 1) on the same estimating model specifications and listed them in Table 5 for comparison.

We now summarize the results in Table 5. First, we find that, for both model specifications, the point estimates of the demand elasticity were all negative but the magnitudes varied across quantile indices. Moreover, both the GMM and inverse QR estimates of  $\theta_1(\tau)$  were of similar values under the basic specification (5.1). When the day effect variables were included as additional controls, the values of  $\hat{\theta}_1(\tau)$  across these two estimation methods differed to a larger extent in the case of  $\tau = 0.25$ . Furthermore, we note that the CI results based on both the GMM and inverse QR methods indicate that the negativity of  $\theta_1(\tau)$  was significant for  $\tau \in \{0.25, 0.75\}$  but we could not reject the case of  $\theta_1(\tau)$  being zero at  $\tau = 0.5$ .

## 6 Conclusions

In this paper, we have proposed a mixed integer quadratic programming approach for estimating the IVQR model within the GMM framework. One possible application of our approach is panel data quantile regression for group-level treatments (Chetverikov, Larsen, and Palmer, 2016). To deal with group-level unobservables, the estimation procedure in Chetverikov, Larsen, and Pa (2016) consists of group-by-group quantile regression followed by two-stage least squares. They mention (in their footnote 10) that the latter step could be replaced by an IV median regression, if one is willing to replace the usual assumption that the group-level errors are uncorrelated with instruments with median uncorrelation (Komarova, Severini, and Tamer,

Table 5: IVQR estimation of demand elasticity

	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	
Specificate	ion (5.1)			
Estimatio	n method: GMM via	the MIQP implement	ntation	
$\widehat{ heta}_1( au)$	-1.0880	-0.8876	-0.9755	
std. dev.	0.4773	0.5056	0.3027	
95% CI	(-2.0234, -0.1525)	(-1.8787, 0.1034)	(-1.5689, -0.3822)	
<b></b>				
^	n method: Inverse QI	₹		
- ( )	-1.3680	-0.8860	-1.2685	
	0.5704		0.3911	
95% CI	(-2.486, -0.250)	(-1.802, 0.030)	(-2.035, -0.502)	
Specificati	Specification (5.1) augmented with day fixed effects			
Estimatio	Estimation method: GMM via the MIQP implementation			
^				
	-0.6915	-0.7152	-1.0904	
	0.3253			
95% CI	(-1.3290, -0.0540)	(-1.6616, 0.2312)	(-1.5735, -0.6074)	
Estimation method: Inverse QR				
	•		4.4500	
- ( )	-1.3635		-1.1790	
	0.5304		0.3653	
95% CI	(-2.403, -0.324)	(-1.457, 0.267)	(-1.895, -0.463)	

2012). This alternative step can be computed using our computation algorithm. It is an interesting topic for future research to fully develop this alternative to IV quantile regression for group-level treatments.

Our approach is limited to GMM estimators for parametric IVQR models. One may consider semiparametric models with endogeneity. For example, Chen, Linton, and Van Keilegom (2003) considered partially linear median regression with some endogenous regressors as one of their examples. Their proposed estimator consists of a two-step procedure: in the first step, nonparametric median regression is carried out given the parameter of interest and in the second step, GMM estimation is implemented with the first step estimates as inputs. Our proposed algorithm is not directly applicable because of the first nonparametric step. It is another interesting topic for future research to develop an algorithm to compute this kind of two-step semiparametric quantile IV estimators.

## References

- Andrews, D. W. (1997): "A stopping rule for the computation of generalized method of moments estimators," *Econometrica*, pp. 913–931.
- Bertsimas, D., A. King, and R. Mazumder (2016): "Best subset selection via a modern optimization lens," *Annals of Statistics*, 44(2), 813–852.
- Chen, L.-Y., and S. Lee (2016): "Best Subset Binary Prediction," arXiv:1610.02738.
- CHEN, X., O. LINTON, AND I. VAN KEILEGOM (2003): "Estimation of semiparametric models when the criterion function is not smooth," *Econometrica*, 71(5), 1591–1608.
- 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.
- ——— (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 C. HANSEN (2013): "Quantile Models with Endogeneity," Annual Review of Economics, 5(1), 57–81.
- Chernozhukov, V., C. Hansen, and M. Jansson (2007): "Inference approaches for instrumental variable quantile regression," *Economics Letters*, 95(2), 272–277.
- ———— (2009): "Finite sample inference for quantile regression models," *Journal of Econometrics*, 152(2), 93–103.
- CHERNOZHUKOV, V., AND H. HONG (2003): "An MCMC approach to classical estimation," *Journal of Econometrics*, 115(2), 293–346.
- CHETVERIKOV, D., B. LARSEN, AND C. PALMER (2016): "IV Quantile Regression for Group-Level Treatments, With an Application to the Distributional Effects of Trade," *Econometrica*, 84(2), 809–833.
- FLORIOS, K., AND S. SKOURAS (2008): "Exact computation of max weighted score estimators," *Journal of Econometrics*, 146(1), 86–91.

- GRADDY, K. (1995): "Testing for imperfect competition at the Fulton fish market," *The RAND Journal of Economics*, pp. 75–92.
- ———— (2006): "Markets: the Fulton fish market," The Journal of Economic Perspectives, 20(2), 207–220.
- Honoré, B. E., and L. Hu (2004): "On the performance of some robust instrumental variables estimators," *Journal of Business and Economic Statistics*, 22(1), 30–39.
- Kaplan, D. M., and Y. Sun (2015): "Smoothed estimating equations for instrumental variables quantile regression," *Econometric Theory*, pp. 1–53.
- KITAGAWA, T., AND A. TETENOV (2015): "Who should be Treated? Empirical Welfare Maximization Methods for Treatment Choice," Cemmap Working Paper CWP10/15.
- Koenker, R. (1994): "Confidence intervals for regression quantiles," in *Asymptotic statistics*, pp. 349–359. Springer.
- Koenker, R., and G. Bassett (1978): "Regression quantiles," *Econometrica*, pp. 33–50.
- Komarova, T., T. Severini, and E. Tamer (2012): "Quantile uncorrelation and instrumental regressions," *Journal of Econometric Methods*, 1(1), 2–14.
- PAKES, A., AND D. POLLARD (1989): "Simulation and the asymptotics of optimization estimators," *Econometrica*, pp. 1027–1057.
- POWELL, J. L. (1986): "Censored regression quantiles," *Journal of econometrics*, 32(1), 143–155.
- Xu, G., and S. Burer (2017): "A branch-and-bound algorithm for instrumental variable quantile regression," *Mathematical Programming Computation*, pp. 1–27.