# Convergence Rates of GMM Estimators with Nonsmooth Moments under Misspecification

**Byunghoon Kang, Seojeong Lee, and Juha Song**

The asymptotic behavior of generalized method of moments (GMM) estimators depends critically on whether the underlying moment condition model is correctly specified. Hong and Li (2024) showed that GMM estimators with nonsmooth (nondirectionally differentiable) moment functions are at best $n^{1/3}$ consistent under misspecification. Through simulations, we verify the decelerated convergence rate of GMM estimators in such cases. For the two-step GMM estimator with an estimated weight matrix, our results align with the theory. However, for the one-step GMM estimator with the identity weight matrix, the convergence rate remains $\sqrt{n}$ even under severe misspecification.

*Keywords*: Generalized method of moments, Nondifferentiable moment, Instrumental variables quantile regression
*JEL Classification*: C13, C15, C21

Byunghoon Kang, Department of Economics, Lancaster University, UK; (E-mail): b.kang1@lancaster.ac.uk

Seojeong Lee (Corresponding author), Department of Economics, Seoul National University, South Korea; (E-mail): s.jay.lee@snu.ac.kr

Juha Song, Department of Economics, Seoul National University, South Korea; (E-mail): yuyuhee2@snu.ac.kr

Lee acknowledges that this work was supported by the New Faculty Startup Fund from Seoul National University.

[**Seoul Journal of Economics** 2025, Vol. 38, No.1]
DOI: 10.22904/sje.2025.38.1.002

## I. Introduction

The generalized method of moments (GMM) is a unifying estimation framework that efficiently combines information about the average behavior of economic variables to estimate parameters of interest defined by the underlying moment condition model. We focus on overidentified, nonsmooth (nondifferentiable), and potentially misspecified moment functions. Additionally, we investigate the convergence rates of one-step and two-step GMM estimators through extensive simulations.

Under standard assumptions, the convergence rate of GMM estimators is $\sqrt{n}$ regardless of whether the moment function is differentiable (Hansen 1982, Newey and McFadden 1994). However, when the moment condition is misspecified, the convergence rate can decelerate. While GMM estimators remain $\sqrt{n}$ consistent and asymptotically normal when the moment function is smooth (Hall and Inoue 2003) or directionally differentiable (Kang and Lee 2024), they become $n^{1/3}$ consistent when the moment function is nondirectionally differentiable (Hong and Li 2024).

Although not specific to GMM, other studies have also demonstrated that the convergence rate of an estimator can change under misspecification. Koo and Seo (2015) investigated this phenomenon in the context of structural break models. When the model is correctly specified or exhibits weak misspecification, such as an incorrect number of breaks, the estimator for the break location converges rapidly to the true breakpoint at a rate of $n^{-1}$. However, under strong misspecification, such as when the true regression function is neither linear nor time invariant , the oracle property no longer holds. The convergence rate of the breakpoint estimator drops significantly, thus reaching $n^{1/3}$ at most.

Similarly, Hidalgo, Lee, and Seo (2019) examined threshold models and found that the convergence rate of the breakpoint estimator is reduced to $n^{1/3}$ if the model is continuous (*e.g.*, a kink model) but the true restriction is not imposed during estimation.

Although Hong and Li (2024) theoretically established a the slower convergence rate for nonsmooth moments under misspecification, they did not provide explicit simulation results.[1] We address this gap by conducting an extensive simulation study to complement the theoretical results. We employ two simulation designs based on those in Hong and Li (2024): (i) a simple location model and (ii) an instrumental variable quantile regression (IVQR) model.

For the two-step efficient GMM estimator or the one-step GMM estimator with an estimated weight matrix, our simulation results align with the theoretical predictions. The outcome shows that the variance of the GMM estimator decreases at a rate of $n^{-2/3}$.[2] However, for the one-step GMM estimator with the identity weight matrix, we observe that the convergence rate remains $\sqrt{n}$ even under severe misspecification. This result is unexpected because Theorem 1 of Hong and Li (2024) establishes the cubic-root convergence rate for GMM estimators with a fixed weight matrix.

## II. Model and Estimator

The moment condition is given by

$$E\left[g(X_i, \theta_0)\right] = 0$$

for a unique $\theta_0$, where $g(X_i, \theta)$ is a known function of the random variables $X_i$ and the parameter of interest $\theta$. The moment condition is just identified if $\dim(\theta) = \dim(g(x, \theta))$ and overidentified if $\dim(\theta) < \dim(g(x, \theta))$. An overidentified moment condition model is misspecified if

$$E\left[g(X_i, \theta)\right] \neq 0, \quad \forall \theta.$$

Notably, this type of moment misspecification can only happen in overidentified moment condition models. When the model is misspecified, the parameter of interest is set as the minimizer of the population GMM criterion function, which is referred to as the pseudo-true value. We assume that the pseudo-true value is unique. Additional

[1] They only provide simulation results comparing the finite sample performance of their proposed rate-adaptive bootstrap with the standard bootstrap.

[2] $\hat{\theta}$ denotes the GMM estimator and $V \equiv \mathrm{var}\left(n^{1/3}(\hat{\theta} - \theta_0)\right) > 0$. Then, $\mathrm{var}(\hat{\theta}) = V/n^{2/3}$.

details can be found in Kang and Lee (2024).

For observations indexed by $i = 1,...,n$, the one-step GMM estimator is defined as

$$\hat{\theta}_1 = \arg\min_{\theta} g_n(\theta)' W_n \, g_n(\theta),$$

where $g_n(\theta) = n^{-1} \sum_{i=1}^{n} g(X_i, \theta)$, $W_n$ is a positive definite weight matrix, which takes the form of $n^{-1} \sum_{i=1}^{n} W(X_i)$, and $W(X_i)$ does not depend on any unknown parameter. We consider two common choices: 1) $W_n = I$, identity matrix, and 2) $W(X_i) = (Z_i Z_i')^{-1}$, where $Z_i$ is the instrument vectors.

The two-step efficient GMM is defined as

$$\hat{\theta}_2 = \arg\min_{\theta} g_n(\theta)' \hat{W}_n \, g_n(\theta),$$

where $\hat{W}_n = \hat{W}_n(\hat{\theta}_1)$ and

$$\hat{W}_n(\theta) = \left( \frac{1}{n} \sum_{i=1}^{n} g(X_i, \theta) g(X_i, \theta)' \right)^{-1}.$$

## III. Nonsmooth Location Model

First, we consider a simple location model with i.i.d. data following Hong and Li (2024). Specifically, the data are generated as

$$y_i = \theta_0 + \varepsilon_i, \qquad i = 1,...,n,$$

where $\varepsilon_i \sim N(0, 2^2)$ and $\theta_0 = 0$. The baseline moment function is defined as

$$g_1(y_i, \theta) = \begin{bmatrix} (y_i \leq \theta) - \tau \\ y_i - \theta \end{bmatrix}.$$

Given that the distribution of $y_i$ is symmetric, the moment condition is correctly specified if $\tau = 0.5$. If $\tau \neq 0.5$, then the model is misspecified. We examine the effect of a misspecified nonlinear moment function (when $\tau \neq 0.5$) by considering the following set of moment functions:

$$g_2(y_i, \theta) = \begin{bmatrix} 1(y_i \leq \theta) - \tau \\ y_i - \theta \\ (y_i - \theta)^2 - 4 \end{bmatrix}.$$

The last equation of $g_2(y_i, \theta)$ imposes a condition that the variance of $y_i$ is four provided that $\theta$ is the mean. When $\tau \neq 0.5$, no parameter satisfies the first two equations of the moment condition simultaneously. Thus, the moment condition is misspecified. In this case, the pseudo-true value differs from the mean. The last equation does not hold either.

Next, we add potentially misspecified parameter-free moment functions (if $\tau \neq 0.5$):

$$g_3((y_i, x_i')', \theta) = \begin{bmatrix} 1(y_i \leq \theta) - \tau \\ y_i - \theta \\ (y_i - \theta)^2 - 4 \\ x_i - (0.5 - \tau) \end{bmatrix},$$

where $x_i \in \mathbb{R}^5$ is generated as

$$\begin{pmatrix} \varepsilon_i \\ x_i \end{pmatrix} \sim N\left(0_{6\times 1}, \begin{bmatrix} 1 & 0.5 & 0.4 & 0.3 & 0.2 & 0.1 \\ 0.5 & 1 & 0.5 & 0.4 & 0.3 & 0.2 \\ 0.4 & 0.5 & 1 & 0.5 & 0.4 & 0.3 \\ 0.3 & 0.4 & 0.5 & 1 & 0.5 & 0.4 \\ 0.2 & 0.3 & 0.4 & 0.5 & 1 & 0.5 \\ 0.1 & 0.2 & 0.3 & 0.4 & 0.5 & 1 \end{bmatrix}\right).$$

Given that $E[x_i] = 0$, the last set of equations in $g_3((y_i, x_i')', \theta)$ is misspecified if $\tau \neq 0.5$.

We compare the finite sample behavior of the GMM estimator using nonsmooth moments with those using smooth moments. Thus, we consider the following moment condition:

$$g_4((y_i, x_i')', \theta) = \begin{bmatrix} y_i - \theta \\ (y_i - \theta)^2 - 4 \\ x_i - (0.5 - \tau) \end{bmatrix},$$

where the variables are generated as specified above.

For each set of moment conditions, we generate $n$ = 200, 400, 800, 1600, 3200, 6400 observations and estimate $\theta$ by one-step GMM with the identity matrix as the weight matrix and the two-step efficient GMM using the one-step GMM as the preliminary estimator. The number of Monte Carlo repetitions is 10,000. The simulation is conducted in MATLAB, where minimization is conducted using the *fminunc* function. The initial value of $\theta$ in the minimization problem is randomly generated from $U[-1,1]$.

Table 1 shows the variance of the one-step and two-step GMM estimators under each set of the moment conditions with increasing sample size. The variance of the $\sqrt{n}$-convergent estimator decays faster ($n^{-1}$) than the $n^{1/3}$-convergent estimator ($n^{-2/3}$).

When the model is correctly specified ($\tau$ = 0.5), the variance of the estimator decreases inversely proportional to the sample size, which is consistent with theoretical predictions. This outcome holds regardless of the moment condition used and whether the estimator is one-step or two-step GMM.

By contrast , under misspecification ($\tau \neq 0.5$), the variance of the estimator does not decrease at the same rate as the increasing sample size. However, the simulation results show otherwise. For the one-step GMM estimator $\hat{\theta}_1$ with the identity matrix as the weight matrix, the variance decreases at the same rate as under correct specification. This finding differs from the theoretical results of Hong and Li (2024), who established a cubic-root convergence rate for the GMM estimator when a fixed weight matrix is used. In particular, Hong and Li (2024) considered a similar simulation setting (except for the variance of $\varepsilon_i$) to demonstrate the superior finite-sample coverage of their adaptive bootstrap confidence intervals (CIs) compared with standard bootstrap CIs. The results in Table 1 suggest that the improved performance of the adaptive bootstrap reported in their Table 1 may stem from factors other than the convergence rate, such as the recentering procedure when calculating the bootstrap GMM estimator.

Meanwhile, the variance of the two-step efficient GMM estimator $\hat{\theta}_2$ decreases at a slower rate than $n^{-1}$, which is approximately $n^{-2/3}$. Adding a nonlinear moment condition ($g_1$ versus $g_2$) does not significantly affect the convergence rate. However, adding misspecified moment conditions ($g_2$ versus $g_3$) generally slows the convergence rate further.

We illustrate the difference in the variance decay rates by normalizing

**TABLE 1**
VARIANCE OF THE GMM ESTIMATOR

| $g_1$ | | $\hat{\theta}_1$ | | | $\hat{\theta}_2$ | | |
|---|---|---|---|---|---|---|---|
| | $\tau$ | 0.1 | 0.3 | 0.5 | 0.1 | 0.3 | 0.5 |
| $n$ | 200 | 0.0205 | 0.0207 | 0.0202 | 0.1115 | 0.1187 | 0.0289 |
| | 400 | 0.0104 | 0.0103 | 0.0100 | 0.0787 | 0.0886 | 0.0166 |
| | 800 | 0.0054 | 0.0054 | 0.0050 | 0.0471 | 0.0570 | 0.0090 |
| | 1600 | 0.0027 | 0.0027 | 0.0025 | 0.0261 | 0.0319 | 0.0046 |
| | 3200 | 0.0014 | 0.0014 | 0.0012 | 0.0152 | 0.0178 | 0.0024 |
| | 6400 | 0.0008 | 0.0007 | 0.0006 | 0.0089 | 0.0103 | 0.0013 |
| $g_2$ | | $\hat{\theta}_1$ | | | $\hat{\theta}_2$ | | |
| | $\tau$ | 0.1 | 0.3 | 0.5 | 0.1 | 0.3 | 0.5 |
| $n$ | 200 | 0.0367 | 0.0383 | 0.0351 | 0.1097 | 0.1199 | 0.0315 |
| | 400 | 0.0166 | 0.0162 | 0.0132 | 0.0795 | 0.0893 | 0.0168 |
| | 800 | 0.0083 | 0.0071 | 0.0052 | 0.0487 | 0.0570 | 0.0089 |
| | 1600 | 0.0041 | 0.0034 | 0.0025 | 0.0275 | 0.0317 | 0.0045 |
| | 3200 | 0.0019 | 0.0017 | 0.0012 | 0.0155 | 0.0174 | 0.0024 |
| | 6400 | 0.0010 | 0.0008 | 0.0006 | 0.0090 | 0.0102 | 0.0013 |
| $g_3$ | | $\hat{\theta}_1$ | | | $\hat{\theta}_2$ | | |
| | $\tau$ | 0.1 | 0.3 | 0.5 | 0.1 | 0.3 | 0.5 |
| $n$ | 200 | 0.0372 | 0.0386 | 0.0357 | 0.1493 | 0.1268 | 0.0230 |
| | 400 | 0.0171 | 0.0166 | 0.0135 | 0.1202 | 0.0926 | 0.0115 |
| | 800 | 0.0084 | 0.0072 | 0.0053 | 0.0911 | 0.0570 | 0.0055 |
| | 1600 | 0.0042 | 0.0035 | 0.0025 | 0.0681 | 0.0274 | 0.0029 |
| | 3200 | 0.0020 | 0.0017 | 0.0012 | 0.0445 | 0.0131 | 0.0015 |
| | 6400 | 0.0010 | 0.0008 | 0.0006 | 0.0235 | 0.0083 | 0.0008 |
| $g_4$ | | $\hat{\theta}_1$ | | | $\hat{\theta}_2$ | | |
| | $\tau$ | 0.1 | 0.3 | 0.5 | 0.1 | 0.3 | 0.5 |
| $n$ | 200 | 0.0367 | 0.0367 | 0.0367 | 0.0158 | 0.0154 | 0.0154 |
| | 400 | 0.0138 | 0.0138 | 0.0138 | 0.0078 | 0.0075 | 0.0074 |
| | 800 | 0.0053 | 0.0053 | 0.0053 | 0.0038 | 0.0037 | 0.0036 |
| | 1600 | 0.0025 | 0.0025 | 0.0025 | 0.0019 | 0.0018 | 0.0018 |
| | 3200 | 0.0012 | 0.0012 | 0.0012 | 0.0009 | 0.0009 | 0.0009 |
| | 6400 | 0.0006 | 0.0006 | 0.0006 | 0.0005 | 0.0005 | 0.0004 |

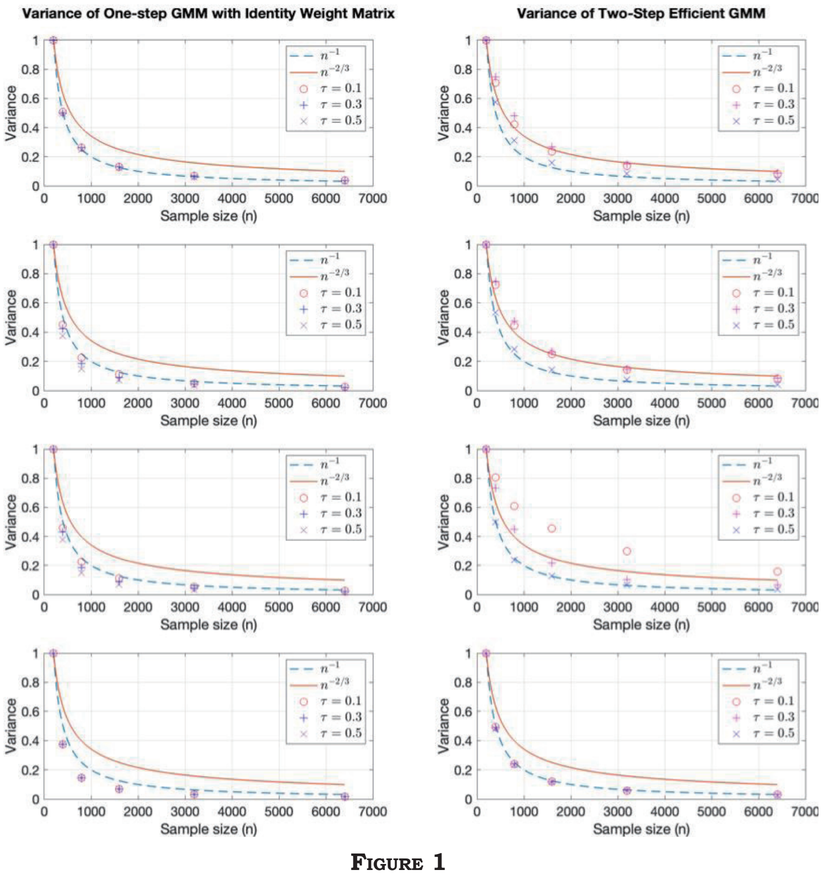


**Figure 1**

Variance Decay Rates of the One-step and Two-step GMM Estimators. First Row ($g_1$), Second Row ($g_2$), Third Row ($g_3$), and Fourth Row ($g_4$)

the variance relative to its value at $n$ = 200 for each $\tau$, $n$, and set of moment conditions. The results are presented in Figure 1. The left column shows the variance of the one-step GMM estimator with the identity weight matrix. Meanwhile, the right column displays the variance of the two-step efficient GMM estimator. Each row corresponds to one of the sets of moment conditions: $g_1$, $g_2$, $g_3$, and $g_4$.

The variance of the one-step GMM estimator with identity matrix decays at approximately an $n^{-1}$ rate even under misspecification ($\tau$ = 0.1, 0.3). By contrast, the variance of the two-step efficient GMM estimator with nonsmooth moments decays significantly slowly under misspecification.

## IV. Quantile Regression with Endogeneity

The IVQR was developed by Chernozhukov and Hansen (2005). They considered a model of quantile treatment effects (QTE) in the presence of endogeneity and derive the moment conditions that are necessary for the identification of QTE without imposing functional form assumptions. This approach provides economic and causal justification for estimation based on these restrictions.

We consider a linear quantile regression model with endogeneity characterized by the structural equation

$$Y = D'\alpha(U) + X'\beta(U), \qquad U \mid X, Z \sim \textit{Uniform}\ (0,1) \tag{1}$$

$$\tau \mapsto D'\alpha(U) + X'\beta(U) \text{ strictly increasing in } \tau,$$

where $Y$ is the scalar outcome variable of interest, $U$ is an unobserved scalar random variable, and $X$ is a vector of included control variables. The covariates $D$ may not be independent of $U$. We assume that a vector of instrumental variables exists, as denoted by $Z$, which is excluded from equation (1) but affects the endogenous variables $D$ with $\dim(Z) \geq \dim(D)$.

Under these assumptions, for $\tau \in (0, 1)$,

$$P\left[Y \leq D'\alpha(\tau) + X'\beta(\tau) \mid X, Z\right] = P\left[U \leq \tau \mid X, Z\right] = \tau\,. \tag{2}$$

In this model, $\alpha(\tau)$ and $\beta(\tau)$ capture the effects of the covariates $D$ and $X$ on the outcome variable for an individual whose unobserved heterogeneity $U$ is fixed at $U = \tau$. By the definition of probability and the law of iterated expectation, (2) implies the following unconditional moment condition:

$$E\left(\left[\tau - 1\left(Y \leq D'\alpha(\tau) + X'\beta(\tau)\right)\right] \Psi\right) = 0\,, \tag{3}$$

where $1(\cdot)$ is an indicator function and $\Psi = \left(X', Z'\right)'$ is a vector of instruments and covariates.

Given these moment conditions, GMM estimation is appropriate.

However, given that the moment function is discontinuous in the parameters of interest, applying conventional minimization techniques is challenging. Consequently, various methods for IVQR estimation have been developed.

Chernozhukov and Hansen (2006, 2008) proposed a method called inverse quantile regression (IQR), which estimates IVQR using the traditional quantile regression approach combined with a grid search algorithm. The IQR method is conducted as follows. The quantile regression objective function is defined as

$$Q_n(\tau, \alpha, \beta, \gamma) := \frac{1}{n}\sum_{i=1}^{n} \rho_\tau\left(Y_i - D_i'\alpha - X_i'\beta - \hat{\Phi}_i'\gamma\right),$$

where $\hat{\Phi}_i' \equiv f(X_i, Z_i)$ is a dim($\alpha$)×1 vector of (transformations of) instruments. The check function $\rho_\tau$ is defined as $\rho_\tau(u) = u(\tau - 1\{u < 0\})$ for $u \in \mathbb{R}$.

For a given value of the structural parameter $\alpha$, the ordinary quantile regression is run using the objective function above. Then, a value of $\alpha$ is obtained, which minimizes the coefficient on the instrumental variable, $\hat{\gamma}(\alpha, \tau)$, as close to 0 as possible. Formally, we have

$$\hat{\alpha}(\tau) \equiv \arg\inf_{\alpha \in \mathrm{A}} \ \hat{\gamma}(\alpha, \tau)' A \hat{\gamma}(\alpha, \tau),$$

$$\text{where } \left(\hat{\beta}(\alpha, \tau), \hat{\gamma}(\alpha, \tau)\right) = \arg\inf_{(\beta,\gamma) \in \mathcal{B}\times\mathcal{G}} \ Q_n(\tau, \alpha, \beta, \gamma)$$

where $A$ is any positive definite matrix and $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{G}$ are compact parameter spaces. The IQR estimator, which is denoted by $\hat{\theta}_{iqr}(\tau)$, is defined as

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

Kaplan and Sun (2017) proposed another method, namely, smoothed IVQR (SIVQR). This approach smoothens the underlying moment condition by applying a kernel to the indicator function in (3). Replacing $1(\cdot)$ with a similar but continuously differentiable function $\tilde{1}(\cdot)$ enables GMM estimation based on smooth moment conditions. When the model is overidentified with dim($Z$) > dim($D$), they transform the original moment condition (3) into

$$E\left([\tau - \tilde{1}(Y \le D'\alpha(\tau) + X'\beta(\tau))]\ \tilde{\Phi}\right) = 0$$

where $\tilde{\Phi} = \left(\hat{D}', X'\right)'$ and $\hat{D}$ is a linear transformation of $X$ and $Z$ that has the same dimension as $D$. This transformation results in an exactly identified model with the transformed instrument vector $\tilde{\Phi}$. IQR and SIVQR use the transformed instruments obtained from the least squares projection of $D$ onto $Z$ and $X$ in practice. Section IV. B and Kaplan and Sun (2017) provide further details.

Machado and Santos Silva (2019) proposed an estimator for conditional quantiles by combining estimates of the location and scale functions, which is referred to as the method of moments-quantile regression (MM-QR). The conditional location-scale model is given by

$$Y = X'\beta + \sigma(X'\gamma)U,$$

where $Y$ is the scalar outcome variable, $X$ includes the endogenous variable $D$ and other exogenous covariates, and $\sigma(\cdot)$ is a known function.

Based on the normalization of the unobserved random variable $U$, Machado and Santos Silva (2019) used the moment conditions $E[ZU] = 0$ and $E[Z(|U|-1)] = 0$ with instruments $Z$ to obtain consistent estimates of $\beta$ and $\gamma$ under very general conditions by applying GMM.

Given the estimates of $\beta$ and $\gamma$, $q(\tau)$ can be estimated using the following moment condition:

$$E\left[\tau - 1\left(\frac{Y - X'\beta}{\sigma(X'\gamma)} \le q(\tau)\right)\right] = 0,$$

where traditional quantile regression can be applied to the estimated residuals. By combining $\hat{\beta}$, $\hat{\gamma}$, and $\hat{q}(\tau)$, the estimates of the desired regression quantile coefficient can be obtained.

The estimators obtained through these methods are not IVQR estimates within the classical GMM framework, which directly uses the moment conditions in equation (3) for GMM estimation. We investigate the convergence rates of the estimators obtained through these various estimation methods compared with the GMM estimator under misspecification.

*A. Simulation Results for IVQR*

The data-generating process for IVQR estimation, as given in Hong and Li (2024), is as follows. For $\alpha_0 = \beta_0 = 1$,

$$y_i = \alpha_0 + \beta_0 D_i + u_i, \quad \begin{pmatrix} u_i \\ D_i \\ W_i \end{pmatrix} \sim N\left( \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & \delta \\ 0 & 1 & 0.5 \\ \delta & 0.5 & 1 \end{pmatrix} \right).$$

Therefore,

$$y_i \mid D_i, W_i \sim N\left( \alpha_0 + \beta_0 D_i + \left( -\frac{2}{3} D_i + \frac{4}{3} W_i \right) \delta,\ 1 - \frac{4}{3}\delta^2 \right).$$

Considering median regression, the population moments for $z_i = (1\ D_i\ W_i)'$ and $\theta = (\alpha, \beta)'$ are given by

$$\pi(\theta) = E\left[ \left( \frac{1}{2} - 1\,(y_i \leq \alpha + \beta D_i) \right) z_i \right]$$

$$= E\left[ \left( \frac{1}{2} - F_{y \mid D,W}\,(y_i \leq \alpha + \beta D_i) \right) z_i \right]$$

$$= E\left[ \left( \frac{1}{2} - \Phi\left( \frac{\alpha - \alpha_0 + (\beta - \beta_0) D_i + \left( \frac{2}{3} D_i - \frac{4}{3} W_i \right) \delta}{\sqrt{1 - \frac{4}{3}\delta^2}} \right) \right) z_i \right].$$

At the true parameter values, the population moments become

$$\pi(\theta_0) = E\left[ \left( \frac{1}{2} - \Phi\left( \frac{\left( \frac{2}{3} D_i - \frac{4}{3} W_i \right) \delta}{\sqrt{1 - \frac{4}{3}\delta^2}} \right) \right) z_i \right]$$

where $\Phi(\cdot)$ is the cumulative standard normal distribution function. This model is correctly specified for median regression when $\delta = 0$. However, when $\delta \neq 0$, the model becomes misspecified.

We generate $n$ = 200, 400, 800, 1600, 3200, 6400 observations and vary the degree of misspecification by setting $\delta$ = 0, 0.1, 0.2, 0.4, 0.6 to estimate $\alpha_0$ and $\beta_0$. The exact computation of the GMM estimator for IVQR models follows a mixed-integer quadratic programming approach, as proposed by Chen and Lee (2018). Two types of one-step GMM estimators are considered:

1. Fixed Weight: Uses the identity matrix as the weight matrix.
2. Estimated Weight: Employs the following weight matrix:

$$\hat{W} = [\, \tau(1 - \tau)n^{-1} \sum z_i z_i' \,]^{-1}$$

The simulation is conducted in MATLAB using Gurobi as the numerical solver. The time and gap are set to zero to ensure full convergence. The number of Monte Carlo repetitions is 1,000.

Table 2 reports the variance of the GMM estimators in median regression as the sample size increases. Results are presented only for the coefficient of the variable of interest, $D_i$. For one-step GMM estimators with fixed weight ($\hat{\beta}_{fixed}$) and estimated weight ($\hat{\beta}_{est}$), the variance decreases at the rate of $n^{-1}$ when the model is correctly specified ($\delta$ = 0).

By contrast, under misspecification ($\delta$ = 0.1, 0.2, 0.4, 0.6), the convergence rate of $\hat{\beta}_{est}$ approaches $n^{-2/3}$ as the degree of misspecification increases. However, regardless of the value of $\delta$, the convergence rate of $\hat{\beta}_{fixed}$ consistently remains at $n^{-1/2}$. These results align with the simulation findings from the previously discussed location model.

This phenomenon is clearly demonstrated in Figure 2. We highlight the differences in variance decay rates by normalizing the variance relative to its value at $n$ = 200 for each sample size and each $\delta$. Under misspecification, the variance of the one-step GMM estimator decays at approximately an $n^{-1}$ rate. However, the convergence rate of the efficient GMM estimator gradually decreases as $\delta$ increases and eventually reaches the $n^{-2/3}$ rate only when $\delta$ is sufficiently large, thus indicating strong misspecification.

Additionally, we calculate the MM-QR estimator proposed by Machado and Santos Silva (2019) using the same DGP, which applies GMM estimation with a directionally differentiable moment condition.

**Table 2**
Variance of the IVQR-GMM Estimator

| | | $\hat{\beta}_{fixed}$ | | | | |
|---|---|---|---|---|---|---|
| | $\delta$ | 0 | 0.1 | 0.2 | 0.4 | 0.6 |
| $n$ | 200 | 0.00884 | 0.00966 | 0.01039 | 0.01472 | 0.02122 |
| | 400 | 0.00417 | 0.00452 | 0.00569 | 0.00803 | 0.00982 |
| | 800 | 0.00221 | 0.00257 | 0.00279 | 0.00415 | 0.00445 |
| | 1600 | 0.00110 | 0.00129 | 0.00150 | 0.00215 | 0.00095 |
| | 3200 | 0.00052 | 0.00061 | 0.00075 | 0.00065 | 0.00035 |
| | 6400 | 0.00027 | 0.00033 | 0.00043 | 0.00018 | 0.00016 |

| | | $\hat{\beta}_{est}$ | | | | |
|---|---|---|---|---|---|---|
| | $\delta$ | 0 | 0.1 | 0.2 | 0.4 | 0.6 |
| $n$ | 200 | 0.00871 | 0.00935 | 0.01065 | 0.01817 | 0.03065 |
| | 400 | 0.00409 | 0.00459 | 0.00580 | 0.01040 | 0.01887 |
| | 800 | 0.00215 | 0.00247 | 0.00346 | 0.00626 | 0.01177 |
| | 1600 | 0.00101 | 0.00131 | 0.00183 | 0.00364 | 0.00731 |
| | 3200 | 0.00053 | 0.00069 | 0.00109 | 0.00203 | 0.00412 |
| | 6400 | 0.00025 | 0.00039 | 0.00061 | 0.00137 | 0.00252 |

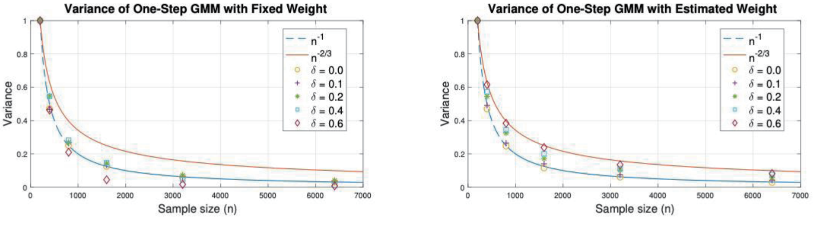


**Figure 2**
Variance Decay Rates of the IVQR One-step GMM Estimators

In STATA, the *ivqreg2* command provides an accessible way to compute MM-QR using the identity function as the scale function. Similar to the previous analysis, the simulation is conducted with $n$ = 200, 400, 800, 1600, 3200, 6400 and $\delta$ = 0, 0.1, 0.2, 0.4, 0.6. Each simulation is repeated 1,000 times. For a fair comparison, we report the estimation

**TABLE 3**
VARIANCE OF THE MMQR ESTIMATOR

| | $\delta$ | 0 | 0.1 | 0.2 | 0.4 | 0.6 |
|---|---|---|---|---|---|---|
| | | $\tilde{\beta}_1$ (One-step) | | | | |
| $n$ | 200 | 0.00583 | 0.00595 | 0.00630 | 0.00792 | 0.01183 |
| | 400 | 0.00279 | 0.00286 | 0.00303 | 0.00380 | 0.00546 |
| | 800 | 0.00139 | 0.00141 | 0.00149 | 0.00183 | 0.00259 |
| | 1600 | 0.00074 | 0.00075 | 0.00079 | 0.00097 | 0.00138 |
| | 3200 | 0.00034 | 0.00034 | 0.00036 | 0.00046 | 0.00067 |
| | 6400 | 0.00018 | 0.00019 | 0.00019 | 0.00024 | 0.00033 |
| | | $\tilde{\beta}_2$ (Two-step) | | | | |
| | $\delta$ | 0 | 0.1 | 0.2 | 0.4 | 0.6 |
| $n$ | 200 | 0.00526 | 0.00530 | 0.00562 | 0.00765 | 0.01233 |
| | 400 | 0.00255 | 0.00261 | 0.00281 | 0.00382 | 0.00603 |
| | 800 | 0.00124 | 0.00128 | 0.00139 | 0.00189 | 0.00289 |
| | 1600 | 0.00067 | 0.00068 | 0.00072 | 0.00097 | 0.00152 |
| | 3200 | 0.00031 | 0.00032 | 0.00035 | 0.00050 | 0.00079 |
| | 6400 | 0.00016 | 0.00016 | 0.00017 | 0.00024 | 0.00040 |

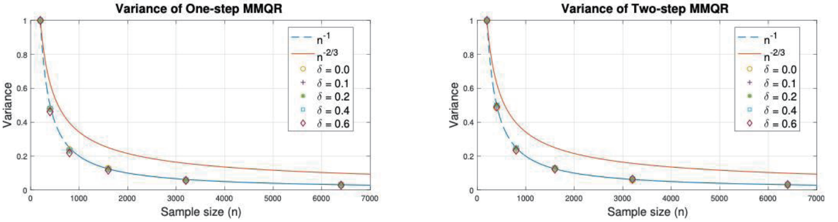


**FIGURE 3**
VARIANCE DECAY RATES OF THE ONE-STEP AND TWO-STEP MMQR ESTIMATORS

results for $\beta_0$ in the median regression.

Table 3 presents the variance of the one-step MMQR estimator $\tilde{\beta}_1$ and the two-step MMQR estimator $\tilde{\beta}_2$ as the sample size increases. The variance of $\tilde{\beta}_1$ and $\tilde{\beta}_2$ decays at approximately an $n^{-1}$ rate as $n$ increases for any value of $\delta$. According to Kang and Lee (2024), when GMM estimation is conducted using a nonsmooth but directionally differentiable moment condition, the convergence rate of the estimator should be $n^{-1/2}$ regardless of whether the model is correctly specified or misspecified. The simulation results align with this theoretical expectation.

Figure 3 illustrates the variance of the MMQR estimator, which is normalized to its value at $n$ = 200. Unlike the results of exact GMM estimation (Chen and Lee, 2018), the MMQR estimator demonstrates that the convergence rate of the variance is approximately $n^{-1}$ regardless of whether fixed or estimated weights are used or has the value of $\delta$.

*B. Transforming Overidentified Moment Conditions into Exactly Identified Ones*

Chernozhukov and Hansen (2006) and Kaplan and Sun (2017) proposed a transformation of overidentified moment conditions into exactly identified moment conditions in quantile regression models with endogeneity. While this transformation may offer computational advantages, it may also obscure potential misspecification in the original overidentified moment condition model.

For illustration, we consider the following simple linear model with no constant and endogeneity:

$$Y = D\beta_0 + e, \quad E[De] \neq 0.$$

For two instruments, namely, $Z_1$ and $Z_2$, $E[Z_1 e] = 0$ but $E[Z_2 e] \neq 0$. In other words, $Z_1$ is a valid instrument, whereas $Z_2$ is not. Both instruments satisfy the relevance condition: $E[Z_1 D] \neq 0$ and $E[Z_2 D] \neq 0$. Then, the moment condition given by

$$E\begin{bmatrix} Z_1(Y - D\beta) \\ Z_2(Y - D\beta) \end{bmatrix}$$

is misspecified because $\beta$ does not simultaneously satisfy the moment condition. The IV estimands are

$$\frac{E[Z_1 Y]}{E[Z_1 D]} = \beta_0, \quad \frac{E[Z_2 Y]}{E[Z_2 D]} = \beta^*$$

and $\beta_0 \neq \beta^*$. We consider a transformation of the original overidentified moment condition into the exactly identified moment condition via a 1×2 matrix $\Pi = (\pi_1, \pi_2)$:

$$E\left[\Pi\begin{pmatrix} Z_1(Y - D\beta) \\ Z_2(Y - D\beta)\end{pmatrix}\right] = E\left[\pi_1 Z_1(Y - D\beta) + \pi_2 Z_2(Y - D\beta)\right]. \tag{4}$$

By solving this moment condition, we find that (4) equals to zero at

$$\bar{\beta} = \frac{\pi_1 E\,[Z_1 D]}{\pi_1 E\,[Z_1 D] + \pi_2 E\,[Z_2 D]}\,\beta_0 + \frac{\pi_2 E\,[Z_2 D]}{\pi_1 E\,[Z_1 D] + \pi_2 E\,[Z_2 D]}\,\beta^*,$$

where $\pi_2 \neq 0$, $\bar{\beta} \neq \beta_0$. Therefore, the exactly identified moment condition holds at a parameter value, which is different from the true value. However , the standard specification tests, such as the J test, cannot be applied to the exactly identified moment condition.

By transforming the instruments using the least squares projection of $D$ onto $Z_1$ and $Z_2$, an overidentified model can be converted into a just-identified model. Given that a $\beta$ that satisfies such moment conditions always exists, moment misspecification, which can occur in overidentified models, cannot arise.

Moreover, the IQR method proposed by Chernozhukov and Hansen (2006) is not a GMM estimator in finite samples[3] because it relies on grid search for estimation. Similarly, the SIVQR method proposed by Kaplan and Sun (2017) is not a GMM estimator with nosmooth moments because it smoothens the indicator function using a kernel. Thus, estimating IVQR using the methods of Chernozhukov and Hansen (2006) and Kaplan and Sun (2017) does not align with the misspecification scenario considered in Hong and Li (2024).

This outcome implies that both estimators obtained using these two methods exhibit standard $\sqrt{n}$ consistency even under misspecification.

*C. Simulation Results for IQR and SIVQR*

We examine the convergence rate of each estimator when the IVQR model is estimated using the methods proposed by Chernozhukov and Hansen (2006) and Kaplan and Sun (2017). In addition, we consider the following DGP from Kang and Lee (2024):

$$y_i = -1 + D_i + \delta(z_{1i} - z_{2i}) + (1 + D_i)\varepsilon_i,$$

[3] However, it is asymptotically a GMM estimator.

$$D_i = \Phi(z_{1i} + z_{2i} + z_{3i} + v_i),$$

where $(Z_{1i}, Z_{2i}, Z_{3i}) \sim N(0, I_3)$ and $(\varepsilon_i, v_i) \sim N(0, I_2)$. $\delta$ is a parameter that controls misspecification, where $\delta = 0$ represents a correctly specified model. Meanwhile, any other value indicates a misspecified model.

If $\delta = 0$, the above model can be rewritten using the Skorohod representation as follows:

$$y_i = \alpha_0(U) + \beta_0(U)D_i, \tag{5}$$

where $U = F_\varepsilon(\varepsilon)$ with $F_\varepsilon$ being the cumulative distribution function of the unobservable $\varepsilon$. Moreover,

$$\alpha_0(\tau) = -1 + F_\varepsilon^{-1}(\tau), \ \ \beta_0(\tau) = 1 + F_\varepsilon^{-1}(\tau).$$

We generate $n$ = 200, 400, 800, 1600, 3200, 6400 observations and consider $\tau$ = 0.25, 0.5, 0.75. The number of Monte Carlo repetitions is 10,000. The simulation is conducted in STATA. Estimation using the IQR method of Chernozhukov and Hansen (2006) is performed with the *ivqregress iqr* command. Meanwhile, the SIVQR method of Kaplan and Sun (2017) is implemented using the *ivqregress smooth* command. Both estimators estimate $\alpha_0(\tau)$ and $\beta_0(\tau)$ in (5).

In Table 4, the variance of IQR and SIVQR estimator decreases at a rate of $n^{-1}$ regardless of the value of $\delta$. As explained in Section IV. B, IQR and SIVQR transform the overidentified model into a just-identified model through linear projection, which prevents the slowdown in the convergence rate of the estimator under moment misspecification.

Figure 4 shows the variance of the IQR estimator in the left column and the variance of the SIVQR estimator in the right column. Each row corresponds to the results for the 0.25, 0.5, and 0.75 quantiles, respectively. We normalize the variance relative to its value at $n$ = 200 for each $\tau$, $n$ and $\delta$. We can confirm that, regardless of whether the model is correctly specified or misspecified, the variance of the estimator decreases inversely proportional to the sample size.

## V. Conclusion

We investigate the convergence rate of the one-step and two-step GMM estimators with nonsmooth moment functions by considering potential misspecification of the moment condition model. Most results are consistent with theory. For directionally differentiable (e.g., check function) moment functions, the variance of GMM estimators decreases at the standard $n^{-1}$ rate regardless of misspecification (Hall and Inoue 2003; Kang and Lee 2024). By contrast, the variance of the GMM estimator with nondirectionally differentiable (e.g., indicator function)

**Table 4**
Variance of the IQR and SIVQR Estimators for $\beta_0(\tau)$

| $\tau = 0.25$ | | $\hat{\beta}_{iqr}$ | | | | $\hat{\beta}_{sivqr}$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\delta$ | 0 | 0.1 | 0.2 | 0.3 | 0 | 0.1 | 0.2 | 0.3 |
| $n$ | 200 | 0.2125 | 0.2149 | 0.2242 | 0.2374 | 0.1758 | 0.1790 | 0.1864 | 0.1982 |
| | 400 | 0.1092 | 0.1100 | 0.1140 | 0.1210 | 0.0924 | 0.0935 | 0.0970 | 0.1032 |
| | 800 | 0.0534 | 0.0543 | 0.0564 | 0.0599 | 0.0464 | 0.0470 | 0.0488 | 0.0518 |
| | 1600 | 0.0266 | 0.0271 | 0.0284 | 0.0301 | 0.0237 | 0.0241 | 0.0250 | 0.0266 |
| | 3200 | 0.0135 | 0.0137 | 0.0142 | 0.0150 | 0.0122 | 0.0123 | 0.0127 | 0.0135 |
| | 6400 | 0.0069 | 0.0069 | 0.0072 | 0.0076 | 0.0062 | 0.0063 | 0.0066 | 0.0070 |
| $\tau = 0.50$ | | $\hat{\beta}_{iqr}$ | | | | $\hat{\beta}_{sivqr}$ | | | |
| | $\delta$ | 0 | 0.1 | 0.2 | 0.3 | 0 | 0.1 | 0.2 | 0.3 |
| $n$ | 200 | 0.1827 | 0.1847 | 0.1904 | 0.2034 | 0.1564 | 0.1580 | 0.1635 | 0.1733 |
| | 400 | 0.0924 | 0.0934 | 0.0972 | 0.1053 | 0.0799 | 0.0816 | 0.0854 | 0.0914 |
| | 800 | 0.0449 | 0.0453 | 0.0469 | 0.0506 | 0.0395 | 0.0398 | 0.0415 | 0.0445 |
| | 1600 | 0.0223 | 0.0228 | 0.0238 | 0.0253 | 0.0201 | 0.0204 | 0.0212 | 0.0226 |
| | 3200 | 0.0114 | 0.0117 | 0.0122 | 0.0131 | 0.0104 | 0.0106 | 0.0111 | 0.0118 |
| | 6400 | 0.0058 | 0.0058 | 0.0061 | 0.0065 | 0.0053 | 0.0054 | 0.0056 | 0.0059 |
| $\tau = 0.75$ | | $\hat{\beta}_{iqr}$ | | | | $\hat{\beta}_{sivqr}$ | | | |
| | $\delta$ | 0 | 0.1 | 0.2 | 0.3 | 0 | 0.1 | 0.2 | 0.3 |
| $n$ | 200 | 0.2172 | 0.2190 | 0.2251 | 0.2394 | 0.1789 | 0.1811 | 0.1875 | 0.1983 |
| | 400 | 0.1054 | 0.1067 | 0.1122 | 0.1185 | 0.0901 | 0.0915 | 0.0956 | 0.1019 |
| | 800 | 0.0535 | 0.0546 | 0.0560 | 0.0589 | 0.0463 | 0.0470 | 0.0486 | 0.0514 |
| | 1600 | 0.0267 | 0.0271 | 0.0281 | 0.0297 | 0.0236 | 0.0238 | 0.0247 | 0.0263 |
| | 3200 | 0.0136 | 0.0138 | 0.0142 | 0.0149 | 0.0122 | 0.0124 | 0.0128 | 0.0136 |
| | 6400 | 0.0067 | 0.0068 | 0.0070 | 0.0074 | 0.0061 | 0.0061 | 0.0063 | 0.0067 |

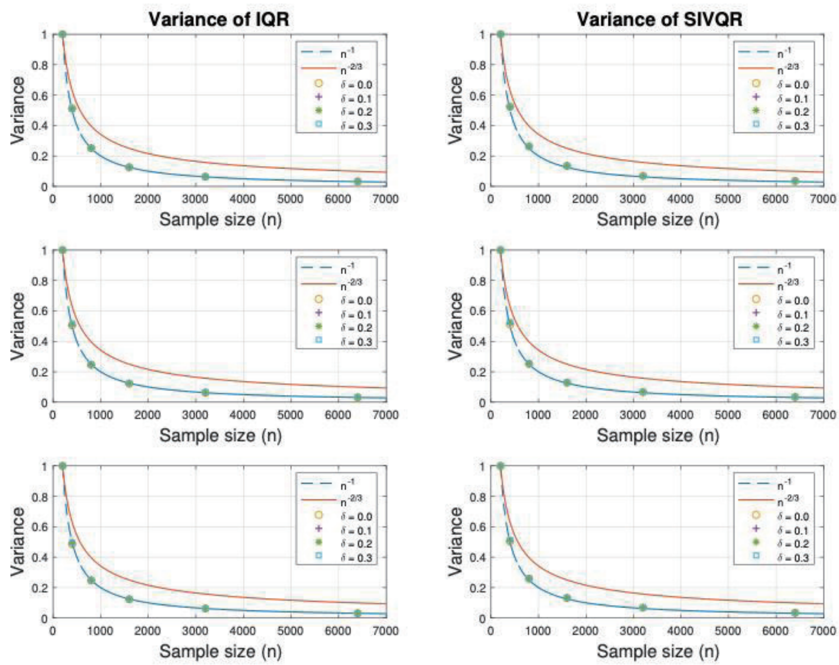


**Figure 4**
Variance Decay Rates of the IQR and SIVQR Estimators for $\beta_0(\tau)$

moment functions decreases at the $n^{-2/3}$ rate under misspecification. Thus, the theoretical findings of Hong and Li (2024) are confirmed.

However, one exception exists. Our simulation results indicate that the variance of the one-step GMM estimator with the identity weight matrix decreases at the standard $n^{-1}$ rate even under severe misspecification. The cause of the discrepancy between the theory and the finite-sample simulation results warrants further investigation.

*(Submitted Jan 15, 2025; Accepted Jan 15, 2025)*

## References

Chen, L., and Lee, S., "Exact computation of GMM estimators for instrumental variable quantile regression models." *Journal of Applied Econometrics 33*(No. 4 2018): 553-567.

Chernozhukov, V., and Hansen, C., “An IV model of quantile treatment effects.” *Econometrica 73*(No. 1 2005): 245-261.

Chernozhukov, V., and Hansen, C., “Instrumental quantile regression inference for structural and treatment effect models.” *Journal of Econometrics 132*(No. 2 2006): 491-525.

Chernozhukov, V., and Hansen, C., “Instrumental variable quantile regression: A robust inference approach.” *Journal of Econometrics 142*(No. 1 2008): 379-398.

Hong, H., and Li, J., “Rate-Adaptive Bootstrap for Possibly Misspecified GMM." *Econometric Theory* (2024): 1-51.

Hall, A. R., and Inoue, A., “The large sample behaviour of the generalized method of moments estimator in misspecified models.” *Journal of Econometrics 114*(No. 2 2003): 361-394.

Hansen, L. P., “Large sample properties of generalized method of moments estimators.” *Econometrica 50*(No. 4 1982): 1029-1054.

Hidalgo, J., Lee, J., and Seo, M. H., "Robust inference for threshold regression models.” *Journal of Econometrics 210*(No. 2 2019): 291-309.

Kang, B., and Lee, S., *Robust Asymptotic and Bootstrap Inference for Nonsmooth GMM*, Available at SSRN 4836402, 2024.

Kaplan, D. M., and Sun, Y., “Smoothed estimating equations for instrumental variables quantile regression.” *Econometric Theory 33*(No. 1 2017): 105-157.

Koo, B., and Seo, M. H., “Structural-break models under mis-specification: Implications for forecasting.” *Journal of Econometrics 188*(No. 1 2015): 166-181.

Machado, J. A.F., and Santos Silva, J.M.C., “Quantiles via moments." *Journal of Econometrics 213*(No. 1 2019): 145-173.

Newey, W. K., and Daniel, M., “Large sample estimation and hypothesis testing.” *Handbook of Econometrics 4*(1994): 2111-2245.