Economics Working Paper Series

2025/006

Misspecification-Robust Asymptotic and Bootstrap Inference for Nonsmooth GMM

Byunghoon Kang and Seojeong Lee

The Department of Economics
Lancaster University Management School
Lancaster LA1 4YX
UK

Misspecification-Robust Asymptotic and Bootstrap Inference for Nonsmooth GMM

Byunghoon Kang* Lancaster University

Seojeong Lee† Seoul National University

March 2, 2025‡

Abstract

This paper develops an asymptotic distribution theory for Generalized Method of Moments (GMM) estimators, including the one-step and iterated estimators, when the moment conditions are nonsmooth and possibly misspecified. We consider nonsmooth moment functions that are directionally differentiable—such as absolute value functions and functions with kinks—but not indicator functions. While GMM estimators remain $\sqrt{n}$-consistent and asymptotically normal for directionally differentiable moments, conventional GMM variance estimators are inconsistent under moment misspecification. We propose a consistent estimator for the asymptotic variance for valid inference. Additionally, we show that the nonparametric bootstrap provides asymptotically valid confidence intervals. Our theory is applied to quantile regression with endogeneity under the location-scale model, offering a robust inference procedure for the GMM estimators in Machado and Santos Silva (2019). Simulation results support our theoretical findings.

1 Introduction

For many important applications of the generalized method of moments (GMM) estimators (Hansen, 1982), the moment function does not satisfy standard smoothness conditions. Existing results, such as those in Pakes and Pollard (1989) and Newey and McFadden (1994), establish asymptotic distribution theory for GMM with nonsmooth criterion functions, assuming correct specification of the moment conditions. However, over-identified GMM models are subject to misspecification of the moment conditions regardless of smoothness of the criterion functions. In practice, researchers often report (marginally) significant over-identifying restrictions test statistic,

*Kang acknowledges the research support from British Academy/Leverhulme Small Research Grants (SRG19-190639). Email: b.kang1@lancaster.ac.uk

†Email: s.jay.lee@snu.ac.kr

‡We thank João Santos Silva, and seminar participants at Exeter, Surrey, Korea University, Seoul National University and Sogang University, and participants in SETA 2023, 2023 Hangzhou International Conference On Frontiers of Data Science, Manchester Econometrics workshop on Advances in Microeconometrics, Groningen Workshop on Causal Inference and Machine Learning, and Bristol Econometric Study Group Conference for helpful comments and suggestions.

suggesting potential moment misspecification. It is well documented that moment misspecification leads to significant bias in conventional GMM variance estimators, which in turn causes size distortions in hypothesis testing. To ensure the validity of the GMM standard errors and confidence intervals, it is crucial to develop inference methods that are robust to potential misspecification.

This paper provides an asymptotic distribution theory for GMM estimators based on overidentified and nonsmooth moment conditions while being robust to potential moment misspecification. We extend existing theories for nonsmooth GMM, such as those by Pakes and Pollard (1989) and Newey and McFadden (1994), to settings where the moment conditions may be misspecified. Specifically, we consider *global misspecification*, where the population moments are equal to fixed nonzero constants. Under global misspecification, the GMM estimator remains consistent for the pseudo-true value, defined as the unique minimizer of the population GMM criterion. Moreover, the asymptotic variance includes additional terms that are typically assumed away under correct specification, rendering conventional heteroskedasticity-robust GMM variance estimators inconsistent.

Robustness to model misspecification does not come for free, as it generally requires stronger smoothness conditions. Standard asymptotic normality results for smooth GMM estimators require only once-differentiability under correct specification, whereas asymptotic results for smooth misspecified GMM (e.g., Hall and Inoue, 2003; Hansen and Lee, 2021) require moment functions to be twice continuously differentiable. Our main result (Theorem 1) assumes *stochastic differentiability* in addition to the *stochastic equicontinuity* conditions typically imposed in the literature on nonsmooth GMM (e.g., Pakes and Pollard, 1989, Theorem 3.3; Newey and McFadden, 1994, Theorem 7.2). While the original stochastic equicontinuity conditions in Pakes and Pollard (1989) and Newey and McFadden (1994) allow for very weak assumptions—potentially permitting discontinuous moment functions—the stochastic differentiability condition requires some form of differentiability, and our theory does not handle discontinuous functions, e.g., indicator functions, and thus restrict the function classes. We view this additional requirement as a necessary cost of addressing general moment misspecification.

There are many interesting applications that satisfy our conditions while failing to meet the stronger requirement of twice continuous differentiability. Examples include GMM estimation of censored regression models with endogeneity (Honoré and Hu, 2004) and GMM estimation of quantile regression with endogeneity, as considered in Machado and Santos Silva (2019), among others. Our results allow moment functions to be locally Lipschitz and differentiable with probability one, encompassing absolute value functions and those with kinks. We also provide sufficient conditions that are straightforward to verify and align with primitive conditions studied in the literature, such as those in Pollard (1985) and van der Vaart and Wellner (1996).

The class of GMM estimators we consider in this paper is the iterated efficient GMM estimator, which includes the one-step GMM estimator with a weight matrix that does not depend on the parameter as a special case. Under moment misspecification, the asymptotic distribution of the two-step GMM depends on that of the previous-step (preliminary) GMM estimator at which the efficient weight matrix is evaluated, making the asymptotic analysis complicated (Hall and Inoue, 2003). The iterated GMM eliminates this dependence, significantly simplifying its asymptotic distribution and making inference more straightforward (Hansen and Lee, 2021). We establish the asymptotic distribution of the iterated GMM estimator with nonsmooth moment functions under possible misspecification.

In practice, the bootstrap is routinely used to compute standard errors or confidence intervals. We establish the consistency of the nonparametric bootstrap confidence interval for nonsmooth GMM estimators, proving that the GMM estimators converge in distribution in probability to the limit distribution. Our result extends Hahn (1996), who established the first-order validity of the bootstrap for nonsmooth GMM under correct specification, and Lee (2014), who established the asymptotic refinement of the nonparametric i.i.d. bootstrap for smooth GMM while allowing for moment misspecification.

In independent work, Hong and Li (2023) show that misspecified GMM estimators with non-directionally differentiable moments are $n^{1/3}$-consistent and have a nonstandard asymptotic distribution. Our results complement theirs in the case of directionally differentiable moment conditions, where the asymptotic distribution remains $\sqrt{n}$-consistent and asymptotically normal. To handle both $n^{1/3}$ and $\sqrt{n}$ rates in a unified framework, their results rely on the cube-root asymptotic theory, similar to Kim and Pollard (1990), and impose assumptions ensuring the finite-dimensional distribution and stochastic equicontinuity of the scaled empirical process for the sample criterion functions. Our paper, in contrast, relies on different assumptions on stochastic equicontinuity and stochastic differentiability that are directly comparable to those in Newey and McFadden (1994) and Pakes and Pollard (1989). Furthermore, the results on the iterated GMM estimator in Section 3.3 are unique to this paper.

This paper does not consider local misspecification, where the population moment condition is modeled as a drifting sequence within an $n^{-1/2}$-neighborhood of zero when evaluated at the true parameter. However, this approach has recently gained considerable attention in the literature, e.g., Conley et al. (2012), Andrews et al. (2017), Armstrong and Kolesar (2021), and Bonhomme and Weidner (2022), among many others. In the locally misspecified GMM setup, the difficulty arises from the first-order bias that must be accounted for in the asymptotic distribution. Several important papers develop valid inference in the local misspecification framework; see Armstrong and Kolesar (2021), Bonhomme and Weidner (2022), Chernozhukov et al. (2023), and references therein.

In the quasi-maximum likelihood framework, Cho and White (2018) consider a directionally differentiable quasi-likelihood function and derive the limit distribution of the standard test statistics. We note that misspecification of the likelihood function differs from the moment misspecification considered in this paper, as the latter can occur only in over-identified models.

The remainder of the paper is organized as follows. Section 2 defines the model setup and GMM estimators and presents two examples. Section 3 establishes the asymptotic theory for the one-step GMM and iterated GMM. Section 4 discusses the variance estimator and the bootstrap.

Section 5 presents results for the methods in Machado and Santos Silva (2019) to illustrate the application of our general results. Section 6 provides simulation evidence, and Section 7 considers an illustrative empirical example of fish demand. All proofs are in Appendix A, and details on the implementation of covariance matrix estimation are provided in Appendix B.

2 Model Setup and Estimators

We have observations $\{X_i\}$ for $i = 1, ..., n$, and $\theta \in \Theta \subset \mathbb{R}^p$ is a p-dimensional parameter of interest. The researcher specifies a vector of moment conditions and assumes that there is a unique parameter value that satisfies the vector of moment conditions:

$$\mathbb{E}[g_n(\theta_0)] = 0, \tag{1}$$

where $g_n(\theta)$ is an m-dimensional vector of data and parameters, which is allowed to be a nonsmooth function of the parameter θ, with $m > p$. The hypothesis (1) is referred to as the correct specification of the model. The conventional nonsmooth GMM theory is established under the assumption of (1).

We consider the class of minimum distance (MD) estimators, including GMM and classical minimum distance (CMD) estimators as two important special cases. For GMM, $g_n(\theta) = n^{-1}\sum_{i=1}^{n} g_i(\theta)$, where $g_i(\theta) = g(X_i, \theta)$ is an m-dimensional moment function. For CMD, $g_n(\theta) = s_n - s(\theta)$, where $s(\theta)$ typically consists of structural parameters or "predictions"under the model, and s_n is a vector of sample statistics or reduced-form estimates.

Along with the nonsmoothness of $g_n(\theta)$, we allow for moment misspecification, which is defined as

$$\mathbb{E}[g_n(\theta)] \neq 0, \ \forall \theta \in \Theta, \tag{2}$$

meaning that no parameter value satisfies the assumed moment condition. For CMD, (2) implies that the "moment matching" or "mapping" $\mathbb{E}[s_n] = s(\theta)$ does not necessarily hold for all values of θ (or in the limit, $\lim_{n\to\infty} \mathbb{E}[s_n - s(\theta)] \neq 0$). Moment misspecification is technically distinct from likelihood misspecification (White, 1982), where the moment condition is given by the first-order conditions (FOCs). In that context, the moment condition holds at the maximizer regardless of likelihood misspecification because the number of parameters equals the number of FOCs. In contrast, moment misspecification occurs only in over-identified models. Despite this distinction, the two concepts share a fundamental similarity: in both cases, the probability limit of the estimator is defined as the minimizer of a well-specified statistical distance measure, referred to as the pseudo-true value. A formal definition will be provided in the next section.

Define the GMM criterion function:

$$J_n(\theta, \phi) = g_n(\theta)' W_n(\phi) g_n(\theta) \tag{3}$$

where the parameter ϕ is the initial value used to form the weight matrix,

$$W_n(\theta) = \left(\frac{1}{n} \sum_{i=1}^{n} v_i(\theta) v_i(\theta)' \right)^{-1}$$

for some known function $v_i(\theta) = v(X_i, \theta)$. When $v_i(\theta) = g_i(\theta)$ or $v_i(\theta) = g_i(\theta) - g_n(\theta)$, the resulting estimator is the efficient GMM.

The one-step estimator minimizes the criterion function using the weight matrix $W_n(\phi) = W_n$ that does not depend on any unknown parameter:

$$\widehat{\theta}_1 = \arg\min_{\theta} g_n(\theta)' W_n g_n(\theta). \tag{4}$$

We can construct $W_n = \left(n^{-1} \sum_{i=1}^{n} v_i v_i'\right)^{-1}$, where $v_i \equiv v(X_i)$. Using $\widehat{\theta}_1$ as an initial value, the two-step estimator is given by

$$\widehat{\theta}_2 = \arg\min_{\theta \in \Theta} J_n(\theta, \widehat{\theta}_1),$$

and the s-step estimator for $s \geq 2$ is

$$\widehat{\theta}_s = \arg\min_{\theta \in \Theta} J_n(\theta, \widehat{\theta}_{s-1}). \tag{5}$$

Conventionally, the efficient GMM is obtained by setting $s = 2$ (two-step efficient GMM) with $v_i(\theta) = g_i(\theta)$ or $v_i(\theta) = g_i(\theta) - g_n(\theta)$. Note that any s-step estimator for $s \geq 2$ with the same form of the weight matrix is efficient under correct specification. In practice, however, the efficient s-step point estimates can change considerably across steps.

The iterated GMM estimator is an alternative to the s-step estimators, overcoming their arbitrariness. Hansen, Heaton, and Yaron (1996) investigate the finite sample properties of the iterated GMM, while Hansen and Lee (2021) formally establish its asymptotic properties, allowing for moment misspecification. The iterated estimator is defined as the limit of the sequence:

$$\widehat{\theta} = \lim_{s \to \infty} \widehat{\theta}_s. \tag{6}$$

Alternatively, the iterated estimator can be viewed as a fixed point. Define the mapping:

$$\overline{\theta}_n(\phi) = \arg\min_{\theta \in \Theta} J_n(\theta, \phi). \tag{7}$$

Then, the limit (6) satisfies the fixed-point equation:

$$\overline{\theta}_n(\widehat{\theta}) = \widehat{\theta}. \tag{8}$$

Below, we provide examples of nonsmooth moment functions covered in this paper.

Example 1 (Dynamic censored regression)**.** Honoré and Hu (2004) consider the estimation of a censored regression model with lagged dependent variables in panel data. Consider the following simple dynamic censored panel regression model:

$$y_{it} = \max(0, y_{it-1}\theta + \alpha_i + \varepsilon_{it})$$

where y_{it} is the variable of interest, α_i is the individual fixed effect, and $\{\varepsilon_{it}\}_{t=1}^{T}$ is a sequence of i.i.d. random variables conditional on (y_{i0}, α_i). To estimate the parameter θ, Honoré and Hu (2004) propose the moment conditions:

$$\mathbb{E}[g_i(\theta)] = \mathbb{E}\begin{pmatrix} g_{i2}(\theta) \\ g_{i3}(\theta) \\ \vdots \\ g_{iT}(\theta) \end{pmatrix} \tag{9}$$

where $g_{it}(\theta) = \max\{0, y_{it} - y_{it-1}\theta\} - y_{it-1} = 1\{y_{it} > y_{it-1}\theta\}(y_{it} - y_{it-1}\theta) - y_{it-1}$ for $t = 2, \ldots, T$, with $T \geq 3$. Under correct specification, the GMM estimator is consistent for the true parameter satisfying the moment conditions: $\mathbb{E}[g_i(\theta_0)] = 0$. However, moment conditions can be misspecified, for example, due to incorrect lag specifications or the presence of heterogeneous effects.

Theorem 1 in Honoré and Hu (2004) establishes the consistency and asymptotic normality of the GMM estimator based on the moment conditions (9). Their proof follows from Theorem 3.3 of Pakes and Pollard (1989), which assumes correct specification of the moment conditions. Furthermore, Theorem 3.3 of Pakes and Pollard (1989) considers the identity weight matrix in the criterion function. However, the limiting distribution in our Theorem 1 differs from theirs under moment misspecification, even when $W_n = I$, as our asymptotic variance includes additional terms. To ensure inference is robust to potential model misspecification, a non-trivial extension of Pakes and Pollard (1989) with a general random weight matrix W_n is required, which will be presented in the next section.

Example 2 (Quantile regression with endogeneity)**.** Machado and Santos Silva (2019) consider a quantile regression model with endogeneity. They focus on the conditional location-scale model of quantile regression:

$$Y_i = X_i'\beta + \sigma(X_i'\gamma)U_i \tag{10}$$

where $Y_i \in \mathbb{R}$ is the outcome variable, $X_i \in \mathbb{R}^k$ is a vector of covariates (which can be potentially endogenous), and $\sigma(X'\gamma) > 0$ is a (known) scale function controlling how X_i affects the dispersion of the distribution of Y_i beyond its location. The location-scale model has been widely studied in the quantile regression literature, including works by Koenker and Bassett (1982), Gutenbrunner and Jurečková (1992), Koenker and Zhao (1994), He (1997), Zhao (2000), and Koenker and Xiao (2002), among many others.

To identify the parameter $\theta = (\beta', \gamma')' \in \mathbb{R}^{2k}$, Machado and Santos Silva (2019) use the moment conditions $\mathbb{E}[U_i|Z_i] = 0$ and $\mathbb{E}[|U_i||Z_i] = 1$, based on the normalization of the unobserved random variable U_i with instruments $Z_i \in \mathbb{R}^m$, where $m \geq k$ (potentially including exogenous regressors from X_i). They propose the following estimation procedure (GMM-Quantile Regression, hereinafter GMM-QR):

$$\widehat{\theta} = (\widehat{\beta}', \widehat{\gamma}')' = \arg\min_{\theta} g_n(\theta)' W_n g_n(\theta),$$

where the moment function $g_i(\theta) \equiv g(X_i, Y_i, Z_i, \theta)$ is given by

$$g_i(\theta) = \begin{bmatrix} Z_i U_i \\ Z_i (|U_i| - 1) \end{bmatrix}, \quad U_i = \frac{Y_i - X_i'\beta}{\sigma(X_i'\gamma)},$$

and W_n is a positive definite weight matrix that does not depend on any unknown parameter. If $\theta = (\beta', \gamma')'$ were known, the following moment condition

$$\mathbb{E}\left[\tau - 1\left(\frac{Y_i - X_i'\beta}{\sigma(X_i'\gamma)} \leq q(\tau)\right)\right] = 0$$

identifies $q(\tau)$, the marginal quantile of U_i, such that $P(U_i \leq q(\tau)) = P(U_i \leq q(\tau)|Z_i) = \tau$ for the quantile index $\tau \in (0, 1)$. The structural quantile function is defined as

$$q(X_i, \tau) = X_i'\beta + \sigma(X_i'\gamma)q(\tau) \tag{11}$$

which satisfies $P(Y_i \leq q(X_i, \tau)) = P(Y_i \leq q(X_i, \tau)|Z_i) = \tau$ under the assumption $Z_i \perp\!\!\!\perp U_i$. The causal effect of X_i on the τth quantile of the conditional distribution of Y_i given X_i, referred to as the *regression quantile coefficient*, is defined as $\alpha(\tau) = (\partial/\partial X_i)q(X_i, \tau)$, which is typically the main object of interest. This model, with a general scale function $\sigma(\cdot)$, allows for nonlinear quantile effects.

The approach of Machado and Santos Silva (2019) can be useful for estimating the structural quantile functions defined by Chernozhukov and Hansen (2006, 2008), as GMM estimation in standard Instrumental Variable Quantile Regression (IVQR) models can be computationally challenging due to the need for non-smooth and non-convex optimization of the GMM criterion function.[1] Additionally, the location-scale model structure ensures that structural quantiles do not cross (e.g., He, 1997).

However, the simplicity and tractability of the location-scale model suggest that the moment conditions used in Machado and Santos Silva (2019) may be subject to misspecification. Kaplan (2022) finds that point estimates using GMM-QR, 2SLS, IVQR, and smoothed IVQR can differ significantly (although they yield qualitatively similar interpretations) and suggests that this may be

[1] See Chen and Lee (2018), Zhu (2019), and Kaido and Wüthrich (2021) for recent developments in computational methods for IVQR-GMM estimation. See also Chernozhukov, Hansen, and Wüthrich (2018) for a review of the IVQR literature. Alternative computational methods for IVQR models include the quasi-Bayesian approach of Chernozhukov and Hong (2003) and the Inverse Quantile Regression (IQR) procedures of Chernozhukov and Hansen (2006, 2008), among others.

due to misspecification of the location-scale model. Therefore, it is important to conduct inference that is robust to potential model misspecification.

If the location-scale model is misspecified, the pseudo-true value of the location and scale parameters $\theta_0 = (\beta_0', \gamma_0')' \in \mathbb{R}^{2k}$ is defined as

$$\theta_0 = \arg\min_{\theta} g(\theta)' W g(\theta),$$

where $g(\theta) = \mathbb{E}[g_n(\theta)]$, and $W_n \xrightarrow{p} W > 0$. The GMM-QR estimand is statistically well-defined and can be interpreted as "the best location-scale parameter" in the sense that it minimizes the distance between the data and the moment conditions from the location-scale model, coinciding with the true value under correct model specification. More importantly, the target parameter here is not the location/scale parameters θ_0, but rather the structural quantile function defined by Chernozhukov and Hansen (2006, 2008) and the marginal effect as in (11).[2] In Section 5, we provide the asymptotic distribution theory of the GMM estimators in Machado and Santos Silva (2019) that is robust to misspecification of the location-scale structure.

3 Asymptotic Normality

In this section, we provide the asymptotic distribution of the estimator $\widehat{\theta}$ while allowing for misspecification. To do so, we first list and discuss the conditions required to ensure asymptotic normality.

The following assumptions are presented in a unified framework to accommodate different GMM estimators. For the one-step estimator, we use the weighting matrix $\widehat{W}_n = W_n$, where $W_n \xrightarrow{p} W > 0$, and the pseudo-true value θ_0 is defined as the minimizer of the population one-step GMM criterion, which will be specified in (13).

For the iterated GMM estimator, we use $\widehat{W}_n = W_n(\widehat{\theta})$, where $W_n(\theta) = \left(\frac{1}{n}\sum_{i=1}^n v_i(\theta)v_i(\theta)'\right)^{-1}$, and $W(\theta) = (n^{-1}\sum_{i=1}^n \mathbb{E}[v_i(\theta)v_i(\theta)'])^{-1}$ for some known function $v_i(\theta) = v(X_i, \theta)$. We define $W = W(\theta_0)$, where θ_0 is the minimizer of the population iterated-GMM criterion in (16).

Assumption 1.

1. *θ_0 is a unique minimizer of the population GMM criterion and in the interior of the compact parameter space Θ.*

2. *Let $g(\theta) = E[g_n(\theta)]$. $g(\theta)$ is twice differentiable at θ_0 with derivatives $G = G(\theta_0)$ and $F = F(\theta_0)$, where $G(\theta) = \frac{\partial}{\partial\theta'} g(\theta)$ and $F(\theta) = \frac{\partial}{\partial\theta'} vec(G(\theta)')$.*

3. *$\inf_{\phi\in\Theta} \lambda_{min}(W(\phi)) \geq C > 0$ for some constant C where $\lambda_{min}(\cdot)$ denotes the smallest eigenvalue.*

[2] Wüthrich (2020) investigates the IVQR estimand when some underlying IVQR assumptions (e.g., rank similarity) are violated under binary treatments and binary instruments in the just-identified setting. In general, characterizing both IVQR and GMM-QR estimands in over-identified models under moment misspecification is challenging.

4. $g_n(\widehat{\theta})'\widehat{W}_n g_n(\widehat{\theta}) \leq \inf_{\theta\in\Theta} g_n(\theta)'\widehat{W}_n g_n(\theta) + o_p(n^{-1})$.

5. $\widehat{\theta} \xrightarrow{p} \theta_0$.

6. *For any* $\delta_n \to 0$ *there exists* G_n *such that*

$$\sup_{||\theta-\theta_0||\leq\delta_n} \frac{\sqrt{n}||g_n(\theta) - g_n(\theta_0) - (g(\theta) - g(\theta_0)) - (G_n - G)(\theta - \theta_0)||}{||\theta - \theta_0||(1 + \sqrt{n}||\theta - \theta_0||)} \xrightarrow{p} 0.$$

7.

$$\sqrt{n}\begin{pmatrix} g_n(\theta_0) - g(\theta_0) \\ (\widehat{W}_n - W)g(\theta_0) \\ (G_n - G)'Wg(\theta_0) \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma & \Lambda & \Gamma \\ \Lambda' & \Psi & \Upsilon \\ \Gamma' & \Upsilon' & \Xi \end{pmatrix}\right).$$

Except for Assumption 1.6, the conditions are standard in the literature. Specifically, they correspond to the conditions in Theorem 7.2 of Newey and McFadden (1994). Assumption 1.1 extends their Conditions (i) and (iii), allowing for moment misspecification. The correct specification assumption of the moment condition in Theorem 7.2 of Newey and McFadden (1994) is removed, and the pseudo-true value is defined as the unique minimizer of the corresponding GMM criterion.

Assumption 1.2 is a slight strengthening of their Condition (ii). Note that the differentiability of the moment function is generally much stronger than the differentiability of its expected value.

Assumptions 1.3-1.5 correspond to the hypotheses of Theorem 7.2 of Newey and McFadden (1994). For one-step GMM, Assumption 1.3 reduces to $W > 0$.

Assumption 1.7 aligns with the condition in Theorem 2 of Hall and Inoue (2003) to establish the asymptotic distribution of the GMM estimator under misspecification. Notably, when $g(\theta_0) = 0$ (correct specification), Assumption 1.7 simplifies to $\sqrt{n}g_n(\theta_0) \xrightarrow{d} N(0, \Sigma)$, which corresponds to Condition (iv) of Theorem 7.2 of Newey and McFadden (1994). Assumptions 1.5 and 1.7 can be replaced by primitive conditions.

3.1 Stochastic Differentiability Condition

Assumption 1.6 (the stochastic differentiability condition) deserves further discussion. Under the very mild condition $G_n - G = o_p(1)$, Assumption 1.6 implies the usual stochastic equicontinuity condition:

$$\sup_{||\theta-\theta_0||\leq\delta_n} \frac{\sqrt{n}||g_n(\theta) - g_n(\theta_0) - (g(\theta) - g(\theta_0))||}{1 + \sqrt{n}||\theta - \theta_0||} \xrightarrow{p} 0$$

for any $\delta_n \to 0$. This is analogous to condition (v) of Theorem 7.2 of Newey and McFadden (1994) and condition (iii) of Theorem 3.3 of Pakes and Pollard (1989), except that we require centering due to misspecification because $g(\theta_0) \neq 0$. For the GMM model, where $g_n(\theta) = n^{-1}\sum_{i=1}^n g(X_i, \theta)$, this stochastic equicontinuity condition holds under mild regularity conditions (Andrews (1994), Newey and McFadden (1994, Theorem 7.3)).[3]

Assumption 1.6 is analogous to the stochastic differentiability condition considered in Pollard (1985) and van der Vaart and Wellner (1996). Pollard (1985) defines that the function $g(\cdot,\theta):\mathbb{R}\to\mathbb{R}$ is *stochastically differentiable* at θ_0 if it admits the following decomposition with a $k\times 1$ vector-valued function $\Delta(x)$ that depends only on x, and a remainder term $r(\cdot,\theta)$:

$$g(x,\theta) = g(x,\theta_0) + \Delta(x)'(\theta-\theta_0) + ||\theta-\theta_0||r(x,\theta),$$

for which

$$\sup_{U_n} \frac{n^{-1/2}\sum_{i=1}^{n}||r(X_i,\theta)-\mathbb{E}[r(X_i,\theta)]||}{1+\sqrt{n}||\theta-\theta_0||} \xrightarrow{p} 0$$

for each sequence of balls U_n that shrinks to θ_0 as $n\to\infty$.

Now, it is worth discussing primitive conditions under which Assumption 1.6 holds for GMM where $g_n(\theta)=n^{-1}\sum_{i=1}^{n} g(X_i,\theta)$. If the moment function $g(x,\theta)$ admits a linear approximation near θ_0, i.e., there exists $\Delta(x,\theta_0)$ (not necessarily a pointwise partial derivative) and $r(x,\theta)\equiv[g(x,\theta)-g(x,\theta_0)-\Delta(x,\theta_0)(\theta-\theta_0)]/||\theta-\theta_0||$, then Assumption 1.6 holds with $G_n=n^{-1}\sum_{i=1}^{n}\Delta(X_i,\theta_0)$, and $G=\mathbb{E}[G_n]=\mathbb{E}[\Delta(X_i,\theta_0)]$ when

$$\sup_{||\theta-\theta_0||\le\delta_n} \frac{\sqrt{n}||n^{-1}\sum_{i=1}^{n} r(X_i,\theta)-\mathbb{E}[r(X_i,\theta)]||}{1+\sqrt{n}||\theta-\theta_0||} \xrightarrow{p} 0. \tag{12}$$

Equation (12) is a version of the stochastic differentiability condition defined in Pollard (1985).

Furthermore, the condition (12) is essentially a stochastic equicontinuity condition for the remainder term $r(\cdot,\theta)$, as in Pollard (1984, 1985) and van der Vaart and Wellner (1996, Lemma 3.2.19). Therefore, the stochastic differentiability conditions hold when $\{r(\cdot,\theta)=\frac{g(\cdot,\theta)-g(\cdot,\theta_0)-\Delta(\cdot,\theta_0)(\theta-\theta_0)}{||\theta-\theta_0||}: ||\theta-\theta_0||<\delta\}$ is a Donsker class for some $\delta>0$. This is a very weak condition on the function class, and many papers in the empirical process literature discuss that it holds in various relevant cases; see, for example, van der Vaart and Wellner (1996) and Chernozhukov and Hong (2003). Pollard (1985) provides detailed examples by verifying the stochastic differentiability conditions using empirical process techniques such as the bracketing method and combinatorial method.

If twice continuous differentiability holds, then Assumption 1.6 automatically holds with $G_n=n^{-1}\sum_{i=1}^{n}(\partial/\partial\theta')g(X_i,\theta_0)$ and $G=\mathbb{E}[G_n]$. The existence of G_n in Assumption 1.6 requires some form of differentiability; however, it is not necessarily a pointwise partial derivative. In the definition of directional differentiability (in the sense of Gâteaux), one would still require the existence of a continuous map $\Delta(\cdot,\theta)$ such that the remainder term is small. In the stochastic differentiability condition, it suffices that the remainder term be small in an average sense.

In CMD, where the sample moment condition is given by $g_n(\theta)=s_n-s(\theta)$, which is linearly additively separable, this condition is automatically satisfied because $G_n=G$, and the asymptotic variance of GMM includes only variation in the weight matrix $\widehat{W}_n$.

[3] In Lemma 1 in the Appendix, we provide a primitive condition for the stochastic equicontinuity condition without assuming correct specification $g(\theta_0)=0$.

We now provide examples where we can verify Assumption 1.6. Examples 1 and 2 in Section 2 satisfy the conditions in Assumption 1.6 but not the stronger continuous differentiability conditions. Example 1 in Section 2 considers the dynamic censored panel regression model in Honoré and Hu (2004). For $t = 2, \dots, T$ with $T \geq 3$, we first consider each moment function

$$g_{it}(\cdot, \theta) = \max\{0, y_{it} - y_{it-1}\theta\} - y_{it-1}, \quad \theta \in \mathbb{R},$$

and then apply the same idea to the stacked moment conditions in (9). We define

$$r_{it}(\cdot, \theta) = \frac{1\{y_{it} > y_{it-1}\theta\}(y_{it} - y_{it-1}\theta) - 1\{y_{it} > y_{it-1}\theta_0\}(y_{it} - y_{it-1}\theta_0) + (\theta - \theta_0)y_{it-1}1\{y_{it} > y_{it-1}\theta\}}{|\theta - \theta_0|}.$$

Under mild moment conditions ($\mathbb{E}\sup_t |y_{it}|^2 < \infty$), $|r_{it}(\cdot, \theta)|$ is uniformly bounded when θ is in a bounded support, and the envelope function $\sup_\theta |r_{it}(\cdot, \theta)|$ has a finite second moment. Using the combinatorial method (Vapnik and Chervonenkis (1971)), similar to the example in Pollard (1985) on the MLE of the location parameter of the double exponential function, it is straightforward to verify that the remainder functions form a polynomial class. It follows that $r_{it}(\cdot, \theta)$ satisfies the stochastic differentiability condition (12).

We will formally verify the conditions for Example 2 in Section 5.

3.2 One-step GMM

Now we derive the asymptotic distribution of the one-step GMM estimator. We define the pseudo-true parameter θ_0 as the solution to the population analog of the one-step GMM estimator $\widehat{\theta} = \widehat{\theta}_1$ in (4),

$$\theta_0 = \arg\min_\theta g(\theta)' W g(\theta). \tag{13}$$

Theorem 1. *Suppose that Assumption 1 holds for the one-step estimator $\widehat{\theta}$ with $\widehat{W}_n = W_n$ in* (4) *and θ_0 defined in* (13). *If $H \equiv G'WG + (g(\theta_0)'W \otimes I_p)F$ is nonsingular, then*

$$\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{d} N(0, H^{-1}\Omega H^{-1})$$

where $\Omega = G'W\Sigma WG + G'W\Lambda G + G'\Lambda' WG + G'\Psi G + G'W\Gamma + \Gamma' WG + G'\Upsilon + \Upsilon' G + \Xi$.

Theorem 1 extends the results in Pakes and Pollard (1989, Theorem 3.3) and Newey and McFadden (1994, Theorems 3.2 and 7.2) to a setup where the moment conditions may be misspecified. Under correct specification, where $g(\theta_0) = 0$, the asymptotic variance simplifies to

$$(G'WG)^{-1}G'W\Sigma WG(G'WG)^{-1},$$

which is the standard formula in Hansen (1982) for smooth functions and in Pakes and Pollard (1989) and Newey and McFadden (1994) for potentially nonsmooth functions. Theorem 3.3 in Pakes and Pollard (1989) and Theorem 7.2 in Newey and McFadden (1994) explicitly assume $g(\theta_0) = 0$. Under general misspecification, where $g(\theta) \neq 0$ for all $\theta \in \Theta$, the asymptotic variance of GMM differs from the conventional GMM variance formula by incorporating additional variation in the sample Jacobian and the sample weight matrix.

The asymptotic distribution in Theorem 1 coincides with those in Hall and Inoue (2003) and Lee (2014) but under strictly weaker conditions than those papers, which assume twice continuous differentiability of $g(x, \theta)$.

3.3 Efficient GMM

Consider the iterated GMM estimator, defined in (8). We derive the asymptotic distribution of the estimator allowing for misspecification. Define the population GMM criterion function

$$J(\theta, \phi) = g(\theta)' W(\phi) g(\theta). \tag{14}$$

Also define the population analog of the mapping (7):

$$\theta(\phi) = \arg\min_{\theta \in \Theta} J(\theta, \phi). \tag{15}$$

Under correct specification, Hansen and Lee (2021) show that $\theta(\phi) = \theta_0$, $\forall \phi \in \Theta$ where θ_0 is the true value that satisfies the moment condition. Under misspecification, however, $\theta(\phi)$ varies with ϕ in general. This is also pointed out by Hall and Inoue (2003), who showed that the one-step and two-step GMM estimators have different limits under misspecification.

The iterated GMM pseudo-true value θ_0 does not depend on ϕ, even under misspecification. It is defined as the fixed point that solves:

$$\theta_0 = \arg\min_{\theta \in \Theta} J(\theta, \theta_0). \tag{16}$$

In other words, $\theta(\theta_0) = \theta_0$. Hansen and Lee (2021) establish the existence of the fixed point (16) by showing that for some $0 \leq c < 1$ and for any ϕ_1, $\phi_2 \in \Theta$,

$$\|\theta(\phi_1) - \theta(\phi_2)\| \leq c \|\phi_1 - \phi_2\| \tag{17}$$

and thus a unique θ_0 that satisfies (16) exists by the Banach fixed point theorem. Sufficient conditions are provided in Assumption 1 of Hansen and Lee (2021), which include our Assumptions 1.2-1.3, along with other conditions ensuring that the degree of misspecification is not too large. Based on their result, we assume that the iterated GMM pseudo-true value θ_0 satisfies our Assumption 1.1.

For the iterated GMM estimator (8), Hansen and Lee (2021) show the existence of this limit by assuming twice differentiability of the moment function. Below, we establish the existence and uniqueness of the iterated GMM estimator with possibly nonsmooth moment functions. To do this, we assume uniform convergence for the sample moment condition and the inverse of the weight matrix.

Assumption 2.

1. $\sup_{\theta\in\Theta} \|g_n(\theta) - g(\theta)\| \xrightarrow{p} 0$
2. $\sup_{\theta\in\Theta} \left\|W_n(\theta)^{-1} - W(\theta)^{-1}\right\| \xrightarrow{p} 0$

We can find primitive conditions for uniform convergence. For example, Assumption 2 holds if (i) X_i are i.i.d., (ii) $\mathbb{E}\|g(X_i,\theta)\| < \infty$ and $\mathbb{E}\|v(X_i,\theta)\|^2 < \infty$, $\forall\theta \in \Theta$, and (iii) a Lipschitz condition (W-LIP in Andrews, 1992) holds for $g(X_i,\theta)$ and $v(X_i,\theta)v(X_i,\theta)'$, respectively. For clustered observations, Hansen and Lee (2019) provide primitive conditions for the uniform law of large numbers in their Theorems 5 and 6. Since the main result of this paper applies to both i.i.d. and clustered samples, we maintain Assumption 2 instead of listing primitive conditions.

The following theorem establishes the existence and uniqueness of the iterated GMM estimator and its consistency.

Theorem 2. *Suppose that Assumption 1.1 holds for $\theta(\phi)$ defined in* (15) *for all $\phi \in \Theta$, and Assumption 1.3 and Assumption 2 hold. Then for θ_0 defined in* (16)*, $\widehat{\theta} \xrightarrow{p} \theta_0$ as $n \to \infty$.*

Next, we establish the asymptotic distribution of the iterated GMM estimator. For differentiable moment functions, the standard asymptotic distribution of the iterated GMM estimator is obtained by applying the first-order Taylor expansion in the sample first-order condition. However, this approach is not possible for non-smooth moment conditions.

As we cannot directly apply asymptotic distribution results in the literature, e.g., Huber (1967) and Newey and McFadden (1994), to our setup where the weight matrix $W(\theta)$ depends on the parameter, we require non-trivial extensions of the general results for the GMM estimator. For examples, the nonsingular second derivative of the population criterion function plays an important role in the asymptotic distribution of Theorem 7.1 of Newey and McFadden (1994) and Theorem 2 of Pollard (1985). We observe that the population criterion function $Q_0(\theta) = -J(\theta,\theta)/2 = -g(\theta)'W(\theta)g(\theta)/2$ has a second derivative at θ_0, which is different than the matrix H below in Theorem 3. The latter includes the derivative of the weight matrix $W(\theta)$ under misspecification.

Let $S = S(\theta_0)$ where $S(\theta) = \frac{\partial}{\partial\theta'} \operatorname{vec}\left(W(\theta)^{-1}\right)$.

Theorem 3. *Suppose that Assumption 1 holds for the iterated GMM estimator $\widehat{\theta}$ with $\widehat{W}_n = W_n(\widehat{\theta})$ in* (6) *and θ_0 defined in* (16)*. If $H \equiv G'WG + (g(\theta_0)'W \otimes I_p)F - (g(\theta_0)'W \otimes G'W)S$ is nonsingular, then,*

$$\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{d} N(0, H^{-1}\Omega H^{-1})$$

where $\Omega = G'W\Sigma WG + G'W\Lambda G + G'\Lambda'WG + G'\Psi G + G'W\Gamma + \Gamma'WG + G'\Upsilon + \Upsilon'G + \Xi$.

4 The Asymptotic Variance and the Bootstrap

For inference robust to moment misspecification, the covariance matrix of the GMM estimators needs to be estimated. We can directly estimate the variance matrix by plugging in the estimates of H and Ω from Theorems 1 and 3 using standard numerical derivatives. See Appendix B for implementation details.

In some applications, however, variance estimation can be complicated. For example, in quantile regression with endogeneity, where Ω depends on conditional density functions, nonparametric density estimation is required. A popular alternative is to use bootstrap methods to construct confidence intervals, similar to the standard quantile regression literature (e.g., Buchinsky, 1995; Hahn, 1995).

Existing results for the first-order validity of the bootstrap for nonsmooth GMM assume correct specification, such as Proposition 1 of Hahn (1996). Below, we establish the first-order validity of the bootstrap under moment misspecification. Therefore, one can use critical values from the bootstrap distribution of the GMM estimators to conduct tests and construct confidence intervals.

Let $\{X_i^*\}_{i=1}^n$ be drawn randomly with replacement from the original data $\{X_i\}_{i=1}^n$, and let $g_n^*(\theta)$ be the same sample moment condition as $g_n(\theta)$ but based on the bootstrap data, so that $g_n^*(\theta) = n^{-1}\sum_{i=1}^n g(X_i^*, \theta)$. Hereinafter, the superscript * denotes a probability or moment computed under the bootstrap distribution conditional on the original data $\{X_i\}_{i=1}^n$.

We define the bootstrap GMM criterion function:

$$J_n^*(\theta, \phi) = g_n^*(\theta)' W_n^*(\phi) g_n^*(\theta), \tag{18}$$

where the weight matrix $W_n^*(\phi)$ is based on the bootstrap data with the initial value ϕ.

The bootstrap estimator for the one-step GMM is

$$\widehat{\theta}_1^* = \arg\min_{\theta} g_n^*(\theta)' W_n^* g_n^*(\theta). \tag{19}$$

where $W_n^*(\phi) = W_n^*$ does not depend on any unknown parameter. Using $\widehat{\theta}_1^*$ as an initial value, the two-step estimator is $\widehat{\theta}_2^* = \arg\min_{\theta\in\Theta} J_n^*(\theta, \widehat{\theta}_1^*)$, and the s-step estimator is

$$\widehat{\theta}_s^* = \arg\min_{\theta\in\Theta} J_n^*(\theta, \widehat{\theta}_{s-1}^*). \tag{20}$$

The bootstrap estimator for the iterated GMM is the limit of this sequence

$$\widehat{\theta}^* = \lim_{s\to\infty} \widehat{\theta}_s^*. \tag{21}$$

Define the mapping

$$\overline{\theta}_n^*(\phi) = \arg\min_{\theta\in\Theta} J_n^*(\theta, \phi). \tag{22}$$

Then, the limit (21) is a fixed point of the equation

$$\overline{\theta}_n^*(\widehat{\theta}^*) = \widehat{\theta}^*. \tag{23}$$

We show in the following theorem that the conventional nonparametric bootstrap consistently estimates the asymptotic distribution of $\sqrt{n}(\widehat{\theta} - \theta_0)$, and thus provides valid asymptotic coverage probabilities of the bootstrap percentile confidence intervals.[4] A word on notation: $\xrightarrow{p^*}$ and $\xrightarrow{d^*}$ denote the convergence in probability, in probability, and the convergence in distribution in probability, respectively. In addition, we write $\xi_n^* = o_{p^*}(1)$ if $\xi_n^* \xrightarrow{p^*} 0$ and $\xi_n^* = O_{p^*}(1)$ if ξ_n^* is bounded in probability, in probability.

Theorem 4. *Suppose that X_i is i.i.d. and Assumption 1 holds. In addition, assume that the following conditions hold for the bootstrap estimator $\widehat{\theta}^* = \widehat{\theta}_1^*$ in (19);*

1. $g_n^*(\widehat{\theta}^*)' W_n^* g_n^*(\widehat{\theta}^*) \leq \inf_{\theta \in \Theta} g_n^*(\theta)' W_n^* g_n^*(\theta) + o_p^*(n^{-1})$.
2. $\widehat{\theta}^* \xrightarrow{p^*} \theta_0$.
3. *For any $\delta_n \to 0$ there exists G_n^* such that*

$$\sup_{||\theta-\theta_0|| \leq \delta_n} \frac{\sqrt{n}||g_n^*(\theta) - g_n^*(\theta_0) - (g_n(\theta) - g_n(\theta_0)) - (G_n^* - G_n)(\theta - \theta_0)||}{||\theta - \theta_0||(1 + \sqrt{n}||\theta - \theta_0||)} \xrightarrow{p^*} 0.$$

Then,

$$\sqrt{n}(\widehat{\theta}^* - \widehat{\theta}) \xrightarrow{d^*} N(0, H^{-1}\Omega H^{-1})$$

where Ω and H are defined in Theorem 1.

Condition 3 is a bootstrap stochastic differentiability assumption analogous to Assumption 1.6. From Giné and Zinn (1990), Condition 3 holds under the same Assumption 1.6 when G_n consists of sample averages. For example, as discussed in Section 3.1, when there exists $\Delta(x, \theta_0)$ such that $g(x,\theta) = g(x,\theta_0) + \Delta(x,\theta_0)(\theta - \theta_0) + r(x,\theta)||\theta - \theta_0||$, Assumption 1.6 holds if the remainder function $r(\cdot, \theta)$ satisfies the stochastic equicontinuity condition. Then, by Giné and Zinn (1990), the stochastic equicontinuity of $r(\cdot, \theta)$ implies the bootstrap stochastic equicontinuity:

$$\sup_{||\theta-\theta_0|| \leq \delta_n} \frac{\sqrt{n}\left|n^{-1}\sum_{i=1}^n r(X_i^*, \theta) - \sum_{i=1}^n r(X_i, \theta)\right|}{1 + \sqrt{n}||\theta - \theta_0||} \xrightarrow{p^*} 0.$$

which implies Condition 3 with $G_n^* = n^{-1}\sum_{i=1}^n \Delta(X_i^*, \theta_0)$.

[4] For brevity, we only provide results for the one-step GMM, but analogous results for the iterated GMM can be shown similarly as in Theorem 4 using the results in the proof of Theorem 3.

5 Quantile Regression with Endogeneity

We consider the following conditional location-scale model of quantile regression, as introduced in Koenker and Bassett (1982):

$$Y_i = X_i'\beta + \sigma(X_i'\gamma)U_i, \tag{24}$$

where $Y_i \in \mathbb{R}$ is the outcome variable, $X_i \in \mathbb{R}^k$ is a vector of covariates (which can be potentially endogenous), and $\sigma(X'\gamma) > 0$ is a known function. We explore the properties of the GMM-QR estimator proposed in Machado and Santos Silva (2019):

$$\widehat{\theta} = (\widehat{\beta}', \widehat{\gamma}')' = \arg\min_{\theta} g_n(\theta)' W_n g_n(\theta), \tag{25}$$

based on the moment function $g_n(\theta) = n^{-1}\sum_{i=1}^n g_i(\theta)$, where $g_i(\theta) = g(X_i, Y_i, Z_i, \theta) = (Z_iU_i, Z_i(|U_i| - 1))$ with instruments $Z_i \in \mathbb{R}^m$ $(m \geq k)$. The weight matrix W_n is a $2m \times 2m$ positive definite matrix of the form $(n^{-1}\sum_{i=1}^n v_iv_i')^{-1}$. Common choices for v_iv_i' include the identity matrix and Z_iZ_i', where Z_i is the instrument vector. The default choice for W_n in the Stata command `ivqreg2` is the identity matrix.

Since the moment conditions used in Machado and Santos Silva (2019) are based on the normalization of the unobserved random variable U_i, they may be subject to misspecification, i.e.,

$$\mathbb{E}[g_n(\theta)] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\begin{pmatrix} Z_i\left(\frac{Y_i - X_i'\beta}{\sigma(X_i'\gamma)}\right) \\ Z_i\left(\left|\frac{Y_i - X_i'\beta}{\sigma(X_i'\gamma)}\right| - 1\right)\end{pmatrix} \neq 0, \ \forall \theta \in \Theta. \tag{26}$$

Under the misspecified model (26), we define the pseudo-true value of the location and scale parameters $\theta_0 = (\beta_0', \gamma_0')' \in \mathbb{R}^{2k}$ as:

$$\theta_0 = \arg\min_{\theta} g(\theta)' W g(\theta), \tag{27}$$

where $g(\theta) = \mathbb{E}[g_n(\theta)]$, and $W_n \xrightarrow{p} W > 0$.[5]

Given (β_0, γ_0), we define $q_0(\tau)$, the marginal quantile of U_i, such that $P(U_i \leq q_0(\tau)) = P(U_i \leq q_0(\tau)|Z_i) = \tau$, which satisfies the following moment condition:

$$\mathbb{E}\left[\tau - 1\left(\frac{Y_i - X_i'\beta_0}{\sigma(X_i'\gamma_0)} \leq q_0(\tau)\right)\right] = 0, \tag{28}$$

where $\tau \in (0, 1)$ is the quantile index. As in Chernozhukov and Hansen (2006), we are interested in the conditional quantiles that satisfy $P(Y_i \leq q_0(X_i, \tau)|Z_i) = \tau$. Given the location-scale model,

[5] Although Machado and Santos Silva (2019) only consider the one-step estimator with $W_n = I_{2m}$, the iterated GMM estimator can also be obtained by iterating the s-step estimator with the efficient weight matrix $W_n(\phi)$ until convergence. The iterated GMM pseudo-true value θ_0 is similarly defined as in (16).

we define the structural quantile function as:

$$q_0(X_i, \tau) = X_i'\beta_0 + \sigma(X_i'\gamma_0)q_0(\tau),$$

and define the regression quantile coefficient as $\alpha_0(X_i, \tau) \equiv \frac{\partial q_0(X_i,\tau)}{\partial X_i} = \beta_0 + \sigma'(X_i'\gamma_0)\gamma_0 q_0(\tau)$, where $\sigma'(x) = \partial\sigma(x)/\partial x$.

Given the GMM-QR estimator $\widehat{\theta} = (\widehat{\beta}', \widehat{\gamma}')'$ in (25), $\widehat{q}(\tau)$ is estimated by the sample moment function of (28):

$$\frac{1}{n}\sum_{i=1}^{n}\left[\tau - 1\left(\frac{Y_i - X_i'\widehat{\beta}}{\sigma(X_i'\widehat{\gamma})} \leq \widehat{q}(\tau)\right)\right] = 0, \tag{29}$$

or by the τth quantile of the standardized residuals $\widehat{U}_i = (Y_i - X_i'\widehat{\beta})/\sigma(X_i'\widehat{\gamma})$. The structural quantile function is estimated as:

$$\widehat{q}(X_i, \tau) = X_i'\widehat{\beta} + \sigma(X_i'\widehat{\gamma})\widehat{q}(\tau). \tag{30}$$

The following theorem shows the joint asymptotic distribution of the one-step GMM-QR estimator $\widehat{\theta}$ under moment misspecification.

Theorem 5. *Consider the model* (24) *with a sample of i.i.d observations* $\{Y_i, X_i, Z_i\}_{i=1}^n$. *Suppose that the following assumptions hold.*

1. *θ_0 is a unique value in the interior of the compact parameter space $\Theta \subset \mathbb{R}^{2k}$.*

2. *The random variables U_i are independent of Z_i, and have a continuous density function $f_U(u)$, and the cumulative density function $F_U(u)$. $f_U(u)$ is bounded away from 0. For some $\epsilon > 0$, $[\lim_{\tau\searrow\epsilon} q(\tau), \lim_{\tau\nearrow 1-\epsilon} q(\tau)]$ is bounded $\tau \in \mathcal{T} = (\epsilon, 1-\epsilon)$.*

3. *$E[|U_i|^{2+\nu}], E[||X_i||^{2+\nu}], E[||v_i||^{4+\nu}], E[||Z_i||^{2+\nu}], E[||Z_iX_i'||^{2+\nu}], E[\sigma'(X_i'\gamma)|^{2+\nu}], E[1/|\sigma(X_i'\gamma)|^{2+\nu}] < \infty$ for some $\nu > 0$.*

Then,

$$\begin{pmatrix} \sqrt{n}(\widehat{\beta} - \beta_0) \\ \sqrt{n}(\widehat{\gamma} - \gamma_0) \\ \sqrt{n}(\widehat{q}(\tau) - q_0(\tau)) \end{pmatrix} \xrightarrow{d} N(0, \Omega), \tag{31}$$

where $\Omega = E(\psi_i \psi_i')$,

$$\begin{aligned}
\psi_i =& \begin{pmatrix} H^{-1} m_i \\ \frac{1}{f_U(q_0(\tau))}[\tau - 1(U_i \leq q_0(\tau))] \end{pmatrix}, \\
H =& G'WG + (g(\theta_0)'W \otimes I_{2k})F, \\
m_i =& G'Wg_i(\theta_0) + G_i(\theta_0)'Wg(\theta_0) - G'Wv_iv_i'Wg(\theta_0), \\
g_i(\theta) =& \begin{pmatrix} Z_i \frac{Y_i - X_i'\beta}{\sigma(X_i'\gamma)} \\ Z_i \left(\left| \frac{Y_i - X_i'\beta}{\sigma(X_i'\gamma)} \right| - 1 \right) \end{pmatrix} \\
G_i(\theta) =& \begin{pmatrix} -\frac{1}{\sigma(X_i'\gamma)} Z_i X_i' & -\frac{\sigma'(X_i'\gamma)(Y_i - X_i'\beta)}{\sigma(X_i'\gamma)^2} Z_i X_i' \\ -\frac{1}{\sigma(X_i'\gamma)} sgn\left(\frac{Y_i - X_i'\beta}{\sigma(X_i'\gamma)}\right) Z_i X_i' & -\frac{\sigma'(X_i'\gamma)(Y_i - X_i'\beta)}{\sigma(X_i'\gamma)^2} sgn\left(\frac{Y_i - X_i'\beta}{\sigma(X_i'\gamma)}\right) Z_i X_i' \end{pmatrix},
\end{aligned}$$

$sgn(x) = 1\{x \geq 0\} - 1\{x \leq 0\}$ *is a sign function,* $G = G(\theta_0)$ *with the population Jacobian* $G(\theta) = E[G_i(\theta)]$, $W = E[v_i v_i']^{-1}$, *and* $F = F(\theta_0)$, $F(\theta) = \frac{\partial}{\partial \theta'} vec(G(\theta)')$, *provided that* H *is non-singular.*

Theorem 5 extends the result in Machado and Santos Silva (2019) to allow for potential moment misspecification. Under correct specification, where $g(\theta_0) = 0$, the asymptotic variance $\Omega = \mathbb{E}(\psi_i \psi_i')$ simplifies to:

$$\psi_i = \begin{pmatrix} (G'WG)^{-1}G'Wg_i(\theta_0) \\ \frac{1}{f_U(q_0(\tau))}[\tau - 1(U_i \leq q_0(\tau))] \end{pmatrix}.$$

Furthermore, under the just-identified model ($k = m$), the asymptotic variance does not depend on the weight matrix and reduces to the formula in Theorem 5 of Machado and Santos Silva (2019).

Inference about the regression quantile coefficient can be performed using Theorem 5 and the delta method. For example, in the linear case when $\sigma(\cdot)$ is the identity function, the regression quantile coefficient is $\widehat{\alpha}(\tau) = \widehat{\beta} + \widehat{\gamma}\widehat{q}(\tau)$, and we have:

$$\sqrt{n}(\widehat{\alpha}(\tau) - \alpha_0(\tau)) \xrightarrow{d} N(0, A\Omega A'), \tag{32}$$

where $A = (I_{k \times k} \quad q_0(\tau)I_{k \times k} \quad \gamma_0)$ is a $k \times (2k+1)$ matrix. The misspecification-robust variance estimator of $\widehat{\alpha}(\tau)$ can then be obtained from (32):

$$\begin{aligned}
&\widehat{V}_{mr}(\widehat{\alpha}(\tau)) = \widehat{A}\widehat{\Omega}\widehat{A}', \\
&\widehat{A} = (I_{k \times k} \quad \widehat{q}(\tau)I_{k \times k} \quad \widehat{\gamma}), \\
&\widehat{\Omega} = \frac{1}{n}\sum_{i=1}^{n} \hat{\psi}_i \hat{\psi}_i', \quad \hat{\psi}_i = \begin{pmatrix} \widehat{H}^{-1}\hat{m}_i \\ \frac{1}{\hat{f}_U(\widehat{q}(\tau))}[\tau - 1(\widehat{U}_i \leq \widehat{q}(\tau))] \end{pmatrix},
\end{aligned}$$

where:

$$\widehat{H} = \widehat{G}'W_n\widehat{G} + (g_n(\widehat{\theta})'W_n \otimes I_{2k})\widehat{F},$$
$$\hat{m}_i = \widehat{G}'W_n g_i(\widehat{\theta}) + G_i(\widehat{\theta})'W_n g_n(\widehat{\theta}) - \widehat{G}'W_n v_i v_i' W_n g_n(\widehat{\theta}).$$

The derivative matrices $\widehat{G}, \widehat{F}$ can be estimated using standard numerical derivative methods, as described in Appendix B. We use the standard kernel density estimator for $f_U(\cdot)$:

$$\hat{f}_U(\widehat{q}(\tau)) = \frac{1}{h_n} K\left(\frac{\widehat{q}(\tau)}{h_n}\right),$$

where $K(\cdot)$ is a kernel function and h_n is a bandwidth satisfying $h_n \to 0, \sqrt{n}h_n \to \infty$. We use the Hall-Sheather (1988) bandwidth for h_n, which is the standard choice for kernel density estimation (e.g., Koenker 1994). The standard error is obtained by taking the diagonal elements of $\sqrt{\widehat{V}_{mr}(\widehat{\alpha}(\tau))/n}$.

For the iterated GMM estimator, the standard error is constructed in the same way using Theorem 3, with:

$$\widehat{H} = \widehat{G}'W_n(\widehat{\theta})\widehat{G} + (g_n(\widehat{\theta})'W_n(\widehat{\theta}) \otimes I_{2k})\widehat{F} - (g_n(\widehat{\theta})'W_n(\widehat{\theta}) \otimes \widehat{G}'W_n(\widehat{\theta}))\widehat{S},$$
$$\hat{m}_i = \widehat{G}'W_n(\widehat{\theta})g_i(\widehat{\theta}) + G_i(\widehat{\theta})'W_n(\widehat{\theta})g_n(\widehat{\theta}) - \widehat{G}'W_n(\widehat{\theta})v_i(\widehat{\theta})v_i(\widehat{\theta})'W_n(\widehat{\theta})g_n(\widehat{\theta}),$$

where $\widehat{G}, \widehat{F}$, and $\widehat{S}$ can be estimated using numerical derivative methods.

Similar to the standard quantile regression literature (e.g., Hahn (1995) and Buchinsky (1995)), we can avoid nonparametric density estimation by using bootstrap methods, as discussed in Section 4.

6 Simulation

In this section, we investigate the finite sample performance of the GMM-QR estimator and the misspecification-robust asymptotic and bootstrap standard errors under both correct specification and misspecification. We consider the following simple location-scale model, as introduced in Section 5:

$$y_i = \beta_0 + \beta_1 D_i + (\gamma_0 + \gamma_1 D_i)u_i,$$

where D_i is a scalar endogenous variable. The instrumental variables are $Z_i = (z_{1i}, z_{2i}, z_{3i})'$ and the moment function is $g_i(\beta, \gamma) = \left(Z_i\left(\frac{y_i - \beta_0 - \beta_1 D_i}{\gamma_0 + \gamma_1 D_i}\right), Z_i\left(\left|\frac{y_i - \beta_0 - \beta_1 D_i}{\gamma_0 + \gamma_1 D_i}\right| - 1\right)\right)$.

The data-generating process (DGP) allowing misspecification is

$$
\begin{aligned}
&y_i = \beta_0 + \beta_1 D_i + \delta(z_{1i} - z_{2i}) + (\gamma_0 + \gamma_1 D_i)u_i, \\
&D_i = \Phi(z_{1i} + z_{2i} + z_{3i} + v_i), \\
&Z_i \sim N(0, I_3), \begin{pmatrix} u_i \\ v_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right).
\end{aligned}
$$

We set the location and scale parameters as $(\beta_0, \beta_1) = (-1, 1)$ and $(\gamma_0, \gamma_1) = (1, 1)$. The number of observations is set to $n = 500, 1000$, and we consider the quantiles $\tau = 0.5, 0.7, 0.9$.

Here, δ controls the degree of misspecification, and we vary $\delta \in \{0, 0.1, 0.2, 0.3\}$. When $\delta = 0$, the model is correctly specified, and we can write the structural quantile function in the form: $q(D_i, \tau) = \eta_0(\tau) + \alpha_0(\tau)D_i$, where $\eta_0(\tau) = \beta_0 + \gamma_0 F_u^{-1}(\tau)$ and $\alpha_0(\tau) = \beta_1 + \gamma_1 F_u^{-1}(\tau)$, with $F_u^{-1}(\tau)$ being the inverse CDF of the unobservable u_i evaluated at τ. The parameter δ controls the degree of misspecification by allowing for the violation of the exclusion restriction on the instruments. For $\delta \neq 0$,

$$
\mathbb{E}\left[Z_i \left(\frac{y_i - \beta_0 - \beta_1 D_i}{\gamma_0 + \gamma_1 D_i} \right) \right] = \delta \mathbb{E}\left[Z_i \frac{z_{1i} - z_{2i}}{\gamma_0 + \gamma_1 D_i} \right] \neq 0,
$$

so the moment condition fails to hold.

The mean and the standard deviation of GMM-QR estimator for regression quantile coefficient $\alpha_0(\tau)$ are computed in Tables 1 and 2. We consider the one-step GMM ($\widehat{\alpha}_1(\tau)$) and iterated GMM ($\widehat{\alpha}(\tau)$) estimators, and report the means of the conventional (heteroskedasticity-robust) standard errors (se $\widehat{\alpha}(\tau)$) and the misspecification robust standard error (se_{mr} $\widehat{\alpha}(\tau)$).

Tables 1 and 2 show that the proposed misspecification-robust standard errors are consistent under misspecification ($\delta \neq 0$), and remain accurate even under correct specification ($\delta = 0$) for almost all cases across τ. The means of robust standard errors (se_{mr} $\widehat{\alpha}(\tau)$) are very close to the standard deviations (sd $\widehat{\alpha}(\tau)$) for all values of δ. The conventional standard error is not consistent and severely downward biased under misspecification, and this bias increases with δ. Furthermore, the iterated GMM estimator tends to have smaller variance than the one-step estimator. Hansen and Lee (2021) provides a heuristic argument of the variance reduction due to the contraction property of the iteration mapping.

The simulation results show that the proposed misspecification-robust standard errors approximate the standard deviations of the GMM-QR estimators well regardless of model misspecification.

	δ (degree of misspecification)	0	0.1	0.2	0.3
$\tau = 0.5$	$\widehat{\alpha}_1(\tau)$	1.0144	0.9995	1.0159	0.9976
	sd $\widehat{\alpha}_1(\tau)$	0.2183	0.2362	0.2496	0.2870
	se $\widehat{\alpha}_1(\tau)$	0.2102	0.2120	0.2156	0.2203
	se$_{mr}$ $\widehat{\alpha}_1(\tau)$	0.2202	0.2070	0.2276	0.2743
	$\widehat{\alpha}(\tau)$	1.0164	0.9998	1.0088	0.9943
	sd $\widehat{\alpha}(\tau)$	0.2215	0.2382	0.2499	0.2777
	se $\widehat{\alpha}(\tau)$	0.2077	0.2094	0.2127	0.2170
	se$_{mr}$ $\widehat{\alpha}(\tau)$	0.2271	0.2120	0.2342	0.3213
$\tau = 0.7$	$\widehat{\alpha}_1(\tau)$	1.5389	1.5061	1.4896	1.3989
	sd $\widehat{\alpha}_1(\tau)$	0.2341	0.2493	0.2716	0.3114
	se $\widehat{\alpha}_1(\tau)$	0.2256	0.2278	0.2316	0.2380
	se$_{mr}$ $\widehat{\alpha}_1(\tau)$	0.2395	0.2243	0.2510	0.3107
	$\widehat{\alpha}(\tau)$	1.5416	1.5182	1.5298	1.5001
	sd $\widehat{\alpha}(\tau)$	0.2362	0.2491	0.2665	0.2927
	se $\widehat{\alpha}(\tau)$	0.2225	0.2245	0.2277	0.2327
	se$_{mr}$ $\widehat{\alpha}(\tau)$	0.2438	0.2286	0.2493	0.3288
$\tau = 0.9$	$\widehat{\alpha}_1(\tau)$	2.2766	2.2449	2.1301	2.0058
	sd $\widehat{\alpha}_1(\tau)$	0.3005	0.3199	0.3632	0.4034
	se $\widehat{\alpha}_1(\tau)$	0.2954	0.2972	0.3019	0.3122
	se$_{mr}$ $\widehat{\alpha}_1(\tau)$	0.3171	0.3129	0.3623	0.4523
	$\widehat{\alpha}(\tau)$	2.2753	2.2699	2.2501	2.2624
	sd $\widehat{\alpha}(\tau)$	0.3043	0.3203	0.3370	0.3569
	se $\widehat{\alpha}(\tau)$	0.2900	0.2917	0.2954	0.3029
	se$_{mr}$ $\widehat{\alpha}(\tau)$	0.3141	0.3039	0.3299	0.3902

Table 1: Monte Carlo Results for Quantile Regression: $n = 500$

6.1 Finite Sample Coverage

We investigate the finite sample performance of the confidence intervals (CIs). We report the coverage properties of the CIs based on the one-step and iterated GMM estimators, using conventional (non-robust) standard errors and misspecification-robust standard errors. We also report the coverage of bootstrap-based CIs.

Based on the same simulation setup described above, Table 3 reports the nominal 95% coverage properties of the following CIs under different degrees of misspecification ($\delta = 0, 0.2, 0.4, 0.6$) and quantiles ($\tau = 0.5, 0.7, 0.9$). First, we report the coverage of CI_{CONV} based on the conventional heteroskedasticity-robust standard errors ($se(\widehat{\alpha}(\tau))$):

$$CI_{CONV} = [\widehat{\alpha}(\tau) \pm 1.96 \times se(\widehat{\alpha}(\tau))].$$

CI_{MR} is based on the misspecification-robust standard errors ($se_{MR}(\widehat{\alpha}(\tau))$):

$$CI_{MR} = [\widehat{\alpha}(\tau) \pm 1.96 \times se_{MR}(\widehat{\alpha}(\tau))].$$

	δ (degree of misspecification)	0	0.1	0.2	0.3
$\tau = 0.5$	$\widehat{\alpha}_1(\tau)$	0.9946	1.0036	0.9952	1.0077
	sd $\widehat{\alpha}_1(\tau)$	0.1559	0.1632	0.1659	0.1947
	se $\widehat{\alpha}_1(\tau)$	0.1497	0.1505	0.1526	0.1569
	se$_{mr}$ $\widehat{\alpha}_1(\tau)$	0.1526	0.1416	0.1482	0.1694
	$\widehat{\alpha}(\tau)$	0.9934	1.0029	0.9923	1.0075
	sd $\widehat{\alpha}(\tau)$	0.1566	0.1623	0.1667	0.1910
	se $\widehat{\alpha}(\tau)$	0.1487	0.1496	0.1516	0.1557
	se$_{mr}$ $\widehat{\alpha}(\tau)$	0.1548	0.1433	0.1500	0.1793
$\tau = 0.7$	$\widehat{\alpha}_1(\tau)$	1.5141	1.5156	1.4650	1.4000
	sd $\widehat{\alpha}_1(\tau)$	0.1582	0.1682	0.1917	0.2131
	se $\widehat{\alpha}_1(\tau)$	0.1605	0.1614	0.1642	0.1688
	se$_{mr}$ $\widehat{\alpha}_1(\tau)$	0.1642	0.1528	0.1613	0.1871
	$\widehat{\alpha}(\tau)$	1.5135	1.5297	1.5142	1.5057
	sd $\widehat{\alpha}(\tau)$	0.1587	0.1691	0.1861	0.1960
	se $\widehat{\alpha}(\tau)$	0.1594	0.1602	0.1627	0.1666
	se$_{mr}$ $\widehat{\alpha}(\tau)$	0.1654	0.1552	0.1597	0.1876
$\tau = 0.9$	$\widehat{\alpha}_1(\tau)$	2.2796	2.2380	2.1394	1.9763
	sd $\widehat{\alpha}_1(\tau)$	0.2139	0.2202	0.2499	0.2930
	se $\widehat{\alpha}_1(\tau)$	0.2086	0.2100	0.2129	0.2205
	se$_{mr}$ $\widehat{\alpha}_1(\tau)$	0.2145	0.2115	0.2334	0.2746
	$\widehat{\alpha}(\tau)$	2.2798	2.2696	2.2670	2.2260
	sd $\widehat{\alpha}(\tau)$	0.2145	0.2148	0.2377	0.2580
	se $\widehat{\alpha}(\tau)$	0.2066	0.2080	0.2106	0.2164
	se$_{mr}$ $\widehat{\alpha}(\tau)$	0.2128	0.2068	0.2141	0.2383

Table 2: Monte Carlo Results for Quantile Regression: $n = 1000$

Finally, we report the coverage of the percentile bootstrap CI:

$$CI_{BOOT-PC} = [\widehat{\alpha}^*_{0.025}(\tau), \widehat{\alpha}^*_{0.975}(\tau)]$$

where $\widehat{\alpha}^*_{0.025}(\tau)$ and $\widehat{\alpha}^*_{0.975}(\tau)$ are the 0.025 and 0.975 quantiles of the bootstrap distribution of $\widehat{\alpha}^*(\tau)$.[6]

We find that the coverage of CI_{CONV} falls significantly below 95% under misspecification ($\delta \neq 0$), and the coverage decreases as the degree of misspecification increases. Using misspecification-robust standard errors for the one-step and iterated GMM estimators improves coverage when the degree of misspecification is large ($\delta = 0.4, 0.6$). Yet, CI_{MR} exhibits undercoverage for $\delta \neq 0$. This may be due to the nonparametric density estimation in the variance-covariance matrix, as discussed in Sections 4 and 5.

Finally, without any direct density estimation, the bootstrap percentile CIs ($CI_{BOOT-PC}$) per-

[6]Simulation results are based on 10,000 bootstrap replications using the warp-speed methods proposed in Giacomini, Politis, and White (2013).

		δ (degree of misspecification)	0	0.2	0.4	0.6
$\tau = 0.5$	One-step	CI_{CONV}	0.940	0.914	0.820	0.663
		CI_{MR}	0.946	0.899	0.881	0.863
		$CI_{BOOT-PC}$	0.950	0.957	0.961	0.978
	Iterated	CI_{CONV}	0.935	0.912	0.834	0.676
		CI_{MR}	0.947	0.906	0.914	0.946
		$CI_{BOOT-PC}$	0.952	0.955	0.956	0.973
$\tau = 0.7$	One-step	CI_{CONV}	0.937	0.912	0.828	0.684
		CI_{MR}	0.945	0.906	0.904	0.910
		$CI_{BOOT-PC}$	0.951	0.954	0.959	0.975
	Iterated	CI_{CONV}	0.932	0.912	0.842	0.698
		CI_{MR}	0.947	0.909	0.917	0.947
		$CI_{BOOT-PC}$	0.953	0.954	0.960	0.969
$\tau = 0.9$	One-step	CI_{CONV}	0.941	0.911	0.838	0.740
		CI_{MR}	0.949	0.942	0.956	0.964
		$CI_{BOOT-PC}$	0.958	0.955	0.958	0.976
	Iterated	CI_{CONV}	0.936	0.917	0.861	0.745
		CI_{MR}	0.944	0.925	0.928	0.956
		$CI_{BOOT-PC}$	0.956	0.956	0.956	0.969

Table 3: Coverage Probabilities of 95% Confidence Intervals (CI). Sample sizes $n = 500$ CI_{CONV}: CI based on the conventional GMM standard errors. CI_{MR}: CI based on the Misspecification-robust GMM standard errors. $CI_{BOOT-PC}$: The percentile bootstrap CI.

form well across different values of τ and δ, with coverage close to 95% in most cases.[7]

7 An Illustrative Empirical Application: Demand for Fish

We illustrate our methods in an empirical application of estimating the demand for fish. We use the dataset from Graddy (1995), which records transactions of whiting in the New York fish market. This dataset has also been used in the quantile regression literature, including Chernozhukov and Hansen (2008), Chernozhukov, Hansen, and Jansson (2009), and Chen and Lee (2018).

We consider the location-scale model of the demand equation:

$$Q_i = \beta_0 + \beta_1 P_i + \beta_2' X_i + (\gamma_0 + \gamma_1 P_i + \gamma_2' X_i) U_i, \tag{33}$$

where Q_i is the logarithm of the total amount of whiting sold each day, and P_i is the logarithm of the average daily fish price, which is endogenous. The vector X_i includes exogenous explanatory

[7] We also investigated results for CIs based on bootstrap standard errors; however, they are not reported here for brevity, as the results were overly conservative—coverage was nearly 99% in most cases. This finding aligns with the theoretical results in Hahn and Liao (2021).

		$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$
GMM-QR	$\widehat{\alpha}(\tau)$	-1.2390	-1.1026	-1.0405
(one-step)	$se(\widehat{\alpha}(\tau))$	(0.6059)	(0.3744)	(0.3925)
	$se_{MR}(\widehat{\alpha}(\tau))$	[0.6205]	[0.7086]	[0.7949]
	95% Bootstrap CI	[-2.5290, -0.0384]	[-2.0987, -0.2680]	[-2.0830, -0.2069]
GMM-QR	$\widehat{\alpha}(\tau)$	-1.3918	-1.0766	-0.9461
(iterated)	$se(\widehat{\alpha}(\tau))$	(0.6150)	(0.3311)	(0.3181)
	$se_{MR}(\widehat{\alpha}(\tau))$	[1.5624]	[0.9496]	[0.7230]
	95% Bootstrap CI	[-2.4305, -0.0982]	[-1.9112, -0.2829]	[-1.7269, -0.2654]
QR	$\widehat{\alpha}(\tau)$	-0.5295	-0.5449	-0.5571
	$se(\widehat{\alpha}(\tau))$	(0.1857)	(0.1626)	(0.1845)
IVQR	$\widehat{\alpha}(\tau)$	-1.0880	-0.8876	-0.9755
	$se(\widehat{\alpha}(\tau))$	(0.4773)	(0.5056)	(0.3027)
SIVQR	$\widehat{\alpha}(\tau)$	-1.2860	-0.7610	-1.0176
	$se(\widehat{\alpha}(\tau))$	(1.3004)	(0.4753)	(0.6032)
Inverse-QR	$\widehat{\alpha}(\tau)$	-1.3680	-0.8860	-1.2685
	$se(\widehat{\alpha}(\tau))$	(0.5704)	(0.4673)	(0.3911)

Table 4: GMM-QR estimation of demand elasticity.
Baseline specification without day fixed effects
OLS: $\underset{(0.1650)}{-0.5408}$, 2SLS: $\underset{(0.3841)}{-1.0141}$

variables, such as indicators for the days of the week (Monday, Tuesday, Wednesday, and Thursday). The structural quantile function can be estimated using the GMM-QR estimator, with two indicator variables for weather conditions at sea (*Stormy* and *Mixed*) as instruments.[8]

We are interested in the regression quantile coefficient:

$$\widehat{\alpha}(\tau) = \widehat{\beta}_1 + \widehat{\gamma}_1 \widehat{q}(\tau), \quad \tau \in (0,1), \tag{34}$$

where $\widehat{q}(\tau)$ is the τth quantile of the standardized residuals, $\widehat{U}_i = \frac{Q_i - \widehat{\beta}_0 - \widehat{\beta}_1 P_i - \widehat{\beta}_2' X_i}{\widehat{\gamma}_0 + \widehat{\gamma}_1 P_i + \widehat{\gamma}_2' X_i}$. The regression quantile coefficient $\widehat{\alpha}(\tau)$ estimates the price elasticity of demand, which varies across different quantile levels.

Tables 4 and 5 report $\widehat{\alpha}(\tau)$ based on the one-step and iterated GMM-QR estimators with conventional standard errors ($se(\widehat{\alpha}(\tau))$) and misspecification-robust standard errors ($se_{MR}(\widehat{\alpha}(\tau))$). We also report bootstrap standard errors and 95% percentile bootstrap confidence intervals, using 10,000 bootstrap replications.

Table 4 presents results for the baseline specification without day fixed effects, while Table 5 shows results for the specification with day fixed effects. For comparison, we report estimation

[8] *Stormy* is a dummy variable indicating wave height greater than 4.5 feet and wind speed greater than 18 knots, while *Mixed* indicates wave height greater than 3.8 feet and wind speed greater than 13 knots. See Chernozhukov and Hansen (2008), Chernozhukov, Hansen, and Jansson (2009) for a detailed description of the data.

		$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$
GMM-QR	$\widehat{\alpha}(\tau)$	-0.9806	-1.0415	-1.0917
(one-step)	$se(\widehat{\alpha}(\tau))$	(0.4123)	(0.3256)	(0.3809)
	$se_{MR}(\widehat{\alpha}(\tau))$	[0.3492]	[0.3445]	[0.4660]
	95% Bootstrap CI	[-2.5825, -0.0563]	[-2.0871, -0.3030]	[-2.1256, -0.2179]
GMM-QR	$\widehat{\alpha}(\tau)$	-1.0190	-0.9303	-0.8620
(iterated)	$se(\widehat{\alpha}(\tau))$	(0.4139)	(0.3088)	(0.2970)
	$se_{MR}(\widehat{\alpha}(\tau))$	[0.6373]	[0.4703]	[0.3904]
	95% Bootstrap CI	[-2.7555, -0.1278]	[-1.8934, -0.2544]	[-1.6820, -0.1762]
QR	$\widehat{\alpha}(\tau)$	-0.5812	-0.5674	-0.5558
	$se(\widehat{\alpha}(\tau))$	(0.1706)	(0.1364)	(0.1592)
IVQR	$\widehat{\alpha}(\tau)$	-0.6915	-0.7152	-1.0904
	$se(\widehat{\alpha}(\tau))$	(0.3253)	(0.4828)	(0.2465)
SIVQR	$\widehat{\alpha}(\tau)$	-0.7920	-0.6487	-0.8980
	$se(\widehat{\alpha}(\tau))$	(0.7626)	(0.4071)	(0.5694)
Inverse-QR	$\widehat{\alpha}(\tau)$	-1.3635	-0.5950	-1.1790
	$se(\widehat{\alpha}(\tau))$	(0.5304)	(0.4398)	(0.3653)

Table 5: GMM-QR estimation of demand elasticity.
Specification with day fixed effects
OLS: $\underset{(0.1521)}{-0.5625}$, 2SLS: $\underset{(0.3577)}{-0.9301}$

results from ordinary least squares (OLS) and two-stage least squares (2SLS), which do not vary with quantile levels. We then present quantile regression (QR) estimates based on the location-scale model without addressing endogeneity. Finally, we report IVQR estimation results, including the exact computation of the GMM estimator for IVQR models using mixed integer quadratic programming, as proposed by Chen and Lee (2018), and the smoothed IVQR (SIVQR) estimator of Kaplan and Sun (2017). Additionally, we report the inverse QR estimation (Inverse-QR) results by Chernozhukov, Hansen, and Jansson (2009).

Chernozhukov, Hansen, and Jansson (2009, Table 1) and Chen and Lee (2018, Table 6) report the same results for IVQR and Inverse-QR.

For the baseline specification in Table 4, we find that the point estimates for all methods accounting for endogeneity are similar at quantile levels $\tau \in \{0.25, 0.75\}$, although results differ slightly between GMM-QR and other methods at $\tau = 0.5$. We also find that the QR method, which does not account for endogeneity, can yield significantly different results compared to other methods. Notably, the misspecification-robust standard errors are generally larger than the conventional standard errors. The demand elasticity coefficients are not statistically significant at the conventional 5% level when using robust standard errors, except for the one-step GMM-QR estimator at $\tau = 0.25$. However, the percentile bootstrap CIs do not include zero for both the one-step and iterated estimators at all values of τ, and they are narrower than the normal-based CIs using the misspecification-robust standard errors.

When the day fixed effects are included in the model, Table 5 indicates that the misspecification-robust standard errors are not necessarily larger than the conventional standard errors. We find that both the one-step and iterated GMM-QR coefficient estimates, $\widehat{\alpha}(\tau)$, are close to -1 and statistically significant at all quantile levels using the misspecification-robust standard errors, except for the iterated estimator at $\tau = 0.25$. The percentile bootstrap CIs again do not include zero for all values of τ and both estimators.

We also find that the demand elasticity coefficient can vary significantly across different methods at $\tau \in \{0.25, 0.5\}$, as discussed in Chen and Lee (2018). Table 5 indicates that both IVQR and Inverse-QR lead to statistically significant results at $\tau \in \{0.25, 0.75\}$. However, we do not reject $\alpha(\tau) = 0$ at $\tau = 0.5$ for IVQR and Inverse-QR, nor at any quantile level $\tau \in \{0.25, 0.5, 0.75\}$ for SIVQR.

We note that using conventional GMM standard errors for these methods in the IVQR model is not valid under misspecification, and this warrants further investigation.[9]

8 Conclusion

This paper develops an asymptotic theory for GMM estimators under nonsmooth and possibly misspecified moment conditions. While the estimators remain $\sqrt{n}$-consistent and asymptotically normal with directionally differentiable moment conditions, conventional variance estimators are inconsistent under misspecification. We provide a consistent variance estimator and establish the validity of nonparametric bootstrap inference. Our results apply to many economic examples, and we illustrate our theoretical findings in the quantile regression model of Machado and Santos Silva (2019).

Our result has an important practical implication: Hahn (1996) established the validity of the GMM bootstrap percentile interval based on under correct specification. We show that the bootstrap interval remains valid under misspecification. Thus, in popular applications such as quantile regressions with IV, the GMM bootstrap percentile intervals enable misspecification-robust inference, whereas other existing methods are not valid under moment misspecification.

While it is beyond the scope of this paper, several potential directions exist for extending the results. First, developing an asymptotic distribution theory for a general semiparametric class of M-estimators under misspecification would be an interesting avenue of research. Notably, we do not consider fully nonparametric or semiparametric conditional moment restriction models. Identification, estimation, and inference for general conditional moment restriction models with possibly nonsmooth moment functions have been studied by Chen, Linton, and van Keilegom (2003), Chen and Pouzo (2009, 2012), and Chen and Liao (2015). Ai and Chen (2007) develop estimation and inference methods for general conditional moment restriction models with potential

[9] Because IVQR moment conditions involve indicator functions, the GMM estimator is $\sqrt{n}$-consistent under correct specification but $n^{1/3}$-consistent under misspecification. Hong and Li (2023) propose a rate-adaptive bootstrap procedure to consistently estimate the asymptotic distribution regardless of model specification.

misspecification, assuming smooth moment functions.

Second, investigating the pseudo-true value in general econometric models and its relationship with the true value is of interest. See Andrews et al. (2024) and references therein for recent developments on this topic. We leave these directions for future research.

Appendix A: Proofs

The following Lemma 1 provides a primitive condition for the stochastic equicontinuity condition that allows for $g(x,\theta)$ be Lipschitz at θ_0 and differentiable at θ_0 with probability one, rather than continuously differentiable. Lemma 1 extends Theorem 7.3 of Newey and McFadden (1994), which establishes the same result under correct specification $E[g_n(\theta_0)] = 0$.

Lemma 1. *Suppose that* $g(\theta) = E[g_n(\theta)] \neq 0,\ \forall\theta \in \Theta$ *where* $g_n(\theta) = n^{-1}\sum_{i=1}^n g(X_i,\theta)$. *Suppose there exists* $\Delta(x,\theta_0)$ *and* $\varepsilon > 0$ *such that with probability one* $r(x,\theta) \equiv ||g(x,\theta) - g(x,\theta_0) - \Delta(x,\theta_0)(\theta-\theta_0)||/||\theta-\theta_0|| \to 0$ *as* $\theta \to \theta_0$, $E[\sup_{||\theta-\theta_0||<\varepsilon} r(x,\theta)] < \infty$, $n^{-1}\sum_{i=1}^n \Delta(X_i,\theta_0) \xrightarrow{p} E[\Delta(X_i,\theta_0)]$. *Then, for any* $\delta_n \to 0$, $\sup_{||\theta-\theta_0||\le\delta_n} \frac{\sqrt{n}||g_n(\theta)-g_n(\theta_0)-(g(\theta)-g(\theta_0))||}{1+\sqrt{n}||\theta-\theta_0||} \xrightarrow{p} 0$ *and* $g(\theta)$ *is differentiable at* θ_0 *with derivative* $G = E[G_n]$ *where* $G_n = n^{-1}\sum_{i=1}^n \Delta(X_i,\theta_0)$.

Proof of Lemma 1:

For any $\varepsilon > 0$, let $r(x,\varepsilon) = \sup_{||\theta-\theta_0||<\varepsilon} r(x,\theta)$. Note that $r(x,\varepsilon) \to 0$ as $\varepsilon \to 0$ with probability one, and thus $E[(r(x,\varepsilon)] \to 0$ as $\varepsilon \to 0$ by the dominated convergence theorem. For all θ such that $||\theta-\theta_0|| \le \delta_n$,

$$
\begin{aligned}
&\sup_{||\theta-\theta_0||\le\delta_n} \frac{\sqrt{n}||g_n(\theta)-g_n(\theta_0)-(g(\theta)-g(\theta_0))||}{1+\sqrt{n}||\theta-\theta_0||} \\
&\le \frac{\sqrt{n}||\frac{1}{n}\sum_{i=1}^n(\Delta(X_i,\theta_0)-E[\Delta(X_i,\theta_0)])\times(\theta-\theta_0)|| + \sqrt{n}(\frac{1}{n}\sum_{i=1}^n r(X_i,\delta_n)+E[r(X_i,\delta_n)])||\theta-\theta_0||}{1+\sqrt{n}||\theta-\theta_0||} \\
&\le \frac{1}{n}\sum_{i=1}^n(\Delta(X_i,\theta_0)-E[\Delta(X_i,\theta_0)]) + O_p(E[r(X_i,\delta_n)]) \xrightarrow{p} 0
\end{aligned}
$$

by the definition of $r(x,\varepsilon)$ and the Markov Inequality.

We also note that $||g(\theta)-g(\theta_0)-G(\theta-\theta_0)|| = ||n^{-1}\sum_{i=1}^n\{E[g(X_i,\theta)]-E[g(X_i,\theta_0)]-E[\Delta(X_i,\theta_0)](\theta-\theta_0)\}||$. For $\theta\to\theta_0$ and $\varepsilon = ||\theta-\theta_0||$,

$$||g(\theta)-g(\theta_0)-G(\theta-\theta_0)|| \le E[r(X_i,\varepsilon)]||\theta-\theta_0|| \to 0$$

which shows $g(\theta)$ is differentiable at θ_0 with derivative G. This completes the proof. □

Lemma 2 shows $\sqrt{n}$-consistency and the asymptotic distribution of the bootstrap estimator $\widehat{\theta}^*$ which maximizes the bootstrap sample criterion function. Lemma 2 provides the bootstrap extension of the Theorem 7.1 of Newey and McFadden (1994).

Let $Q_n^*(\theta)$ be the same criterion function as $Q_n(\theta)$ but based on the bootstrap data. Define the remainder terms as

$$R_n(\theta) = \frac{Q_n(\theta) - Q_n(\theta_0) - D_n'(\theta - \theta_0) - (Q(\theta) - Q(\theta_0))}{\|\theta - \theta_0\|}, \tag{35}$$

$$R_n^*(\theta) = \frac{Q_n^*(\theta) - Q_n^*(\theta_0) - D_n^{*\prime}(\theta - \theta_0) - (Q_n(\theta) - Q_n(\theta_0))}{\|\theta - \theta_0\|}. \tag{36}$$

Lemma 2. *Suppose that $Q(\theta)$ is maximized on Θ at $\theta_0 \in int(\Theta)$, $Q(\theta)$ is twice differentiable at θ_0 with nonsingular second derivative H, and that $\widehat{\theta} \xrightarrow{p} \theta_0$. Suppose that*

(i) $Q_n(\widehat{\theta}) \geq \sup_{\theta \in \Theta} Q_n(\theta) - o_p(n^{-1})$;

(ii) $\sqrt{n} D_n \xrightarrow{d} N(0, \Omega)$;

(iii) *for any* $\delta_n \to 0$, $\sup_{\|\theta - \theta_0\| \leq \delta_n} |\sqrt{n} R_n(\theta)/[1 + \sqrt{n}\|\theta - \theta_0\|]| \xrightarrow{p} 0$.

Further, suppose the following holds for the bootstrap estimator $\widehat{\theta}^*$ such that $\widehat{\theta}^* \xrightarrow{p^*} \theta_0$. Suppose that

(i)* $Q_n^*(\widehat{\theta}^*) \geq \sup_{\theta \in \Theta} Q_n^*(\theta) - o_p^*(n^{-1})$;

(ii)* $\sqrt{n} D_n^* \xrightarrow{d^*} N(0, \Omega)$;

(iii)* *for any* $\delta_n \to 0$, $\sup_{\|\theta - \theta_0\| \leq \delta_n} |\sqrt{n} R_n^*(\theta)/[1 + \sqrt{n}\|\theta - \theta_0\|]| \xrightarrow{p^*} 0$.

Then, $\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{d} N(0, H^{-1}\Omega H^{-1})$, *and* $\sqrt{n}(\widehat{\theta}^* - \widehat{\theta}) \xrightarrow{d^*} N(0, H^{-1}\Omega H^{-1})$.

Proof of Lemma 2: By the hypotheses of the Lemma, the asymptotic distribution results for $\widehat{\theta}$ directly follow from Theorem 7.1 of Newey and McFadden (1994). Therefore, it suffices to show that the bootstrap estimator converges in distribution to the same asymptotic distribution in probability.

We first prove $\sqrt{n}$-consistency of the bootstrap estimator $\widehat{\theta}^*$. By the same argument for $\widehat{\theta}$, the negative definiteness[10] of H and $\widehat{\theta}^* \xrightarrow{p^*} \theta_0$ imply that

$$Q(\widehat{\theta}^*) \leq Q(\theta_0) - C^* \|\widehat{\theta}^* - \theta_0\|^2 \tag{37}$$

for some $C^* > 0$. Choose U_n^* so that $\widehat{\theta}^* \in U_n^*$ w.p.a.1, so that by (iii^*)

$$|R_n^*(\widehat{\theta}^*)| \leq (1 + \sqrt{n}\|\widehat{\theta}^* - \theta_0\|) o_p^*(n^{-1/2}). \tag{38}$$

[10]Note that this H corresponds to $-H$ in the main text.

Then by (i^*), (35), (36), (37), and (38),

$$
\begin{aligned}
0 \leq & Q_n^*(\widehat{\theta}^*) - Q_n^*(\theta_0) + o_p^*(n^{-1}) \\
= & Q_n(\widehat{\theta}^*) - Q_n(\theta_0) + D_n^{*'}(\widehat{\theta}^* - \theta_0) + \|\widehat{\theta}^* - \theta_0\| R_n^*(\widehat{\theta}^*) + o_p^*(n^{-1}) \\
= & Q(\widehat{\theta}^*) - Q(\theta_0) + (D_n + D_n^*)'(\widehat{\theta}^* - \theta_0) + \|\widehat{\theta}^* - \theta_0\|(R_n(\widehat{\theta}^*) + R_n^*(\widehat{\theta}^*)) + o_p^*(n^{-1}) \\
\leq & - C^*\|\widehat{\theta}^* - \theta_0\|^2 + O_p^*(n^{-1/2})\|\widehat{\theta}^* - \theta_0\| + \|\widehat{\theta}^* - \theta_0\| \left(1 + \sqrt{n}\|\widehat{\theta}^* - \theta_0\|\right) o_p^*(n^{-1/2}) + o_p^*(n^{-1}) \\
\leq & - (C^* + o_p^*(1))\|\widehat{\theta}^* - \theta_0\|^2 + O_p^*(n^{-1/2})\|\widehat{\theta}^* - \theta_0\| + o_p^*(n^{-1}).
\end{aligned}
$$

Since $C^* + o_p^*(1)$ is bounded away from zero with probability approaching 1 in probability, it follows that $\|\widehat{\theta}^* - \theta_0\| + O_p^*(n^{-1/2}) \leq O_p^*(n^{-1/2})$. Therefore, by the triangle inequality

$$
\|\widehat{\theta}^* - \theta_0\| \leq \left| \|\widehat{\theta}^* - \theta_0\| + O_p^*(n^{-1/2}) \right| + \left| O_p^*(n^{-1/2}) \right| \leq O_p^*(n^{-1/2}). \tag{39}
$$

Thus, $\sqrt{n}$-consistency is proved.

Next, let $\widetilde{\theta}^* = \widetilde{\theta} - H^{-1}D_n^*$. Then $\widetilde{\theta}^*$ is $\sqrt{n}$-consistent for θ_0 because

$$
\|\widetilde{\theta}^* - \theta_0\| \leq \|\widetilde{\theta} - \theta_0\| + \|H^{-1}\|\|D_n^*\| \leq O_p^*(n^{-1/2}) \tag{40}
$$

by (ii^*). Now by (36), (35), twice differentiability of $Q(\theta)$, and $\sqrt{n}$-consistency of $\widehat{\theta}^*$,

$$
\begin{aligned}
& 2\left(Q_n^*(\widehat{\theta}^*) - Q_n^*(\theta_0)\right) \\
= & 2\left(Q_n(\widehat{\theta}^*) - Q_n(\theta_0)\right) + 2D_n^{*'}(\widehat{\theta}^* - \theta_0) + 2R_n^*(\widehat{\theta}^*)\|\widehat{\theta}^* - \theta_0\| \\
= & 2\left(Q(\widehat{\theta}^*) - Q(\theta_0)\right) + 2(D_n + D_n^*)'(\widehat{\theta}^* - \theta_0) + 2(R_n(\widehat{\theta}^*) + R_n^*(\widehat{\theta}^*))\|\widehat{\theta}^* - \theta_0\| \\
= & (\widehat{\theta}^* - \theta_0)'H(\widehat{\theta}^* - \theta_0) + 2(D_n + D_n^*)'(\widehat{\theta}^* - \theta_0) + 2(R_n(\widehat{\theta}^*) + R_n^*(\widehat{\theta}^*))\|\widehat{\theta}^* - \theta_0\| + o\left(\|\widehat{\theta}^* - \theta_0\|^2\right) \\
= & (\widehat{\theta}^* - \theta_0)'H(\widehat{\theta}^* - \theta_0) + 2(D_n + D_n^*)'(\widehat{\theta}^* - \theta_0) + o_p^*(n^{-1}).
\end{aligned}
$$

Since

$$
D_n^{*'}(\widehat{\theta}^* - \theta_0) = -(\widetilde{\theta}^* - \widetilde{\theta})'H(\widehat{\theta}^* - \theta_0) = -(\widetilde{\theta}^* - \theta_0)'H(\widehat{\theta}^* - \theta_0) - D_n'(\widehat{\theta}^* - \theta_0),
$$

it follows that

$$
2\left(Q_n^*(\widehat{\theta}^*) - Q_n^*(\theta_0)\right) = (\widehat{\theta}^* - \theta_0)'H(\widehat{\theta}^* - \theta_0) - 2(\widetilde{\theta}^* - \theta_0)'H(\widehat{\theta}^* - \theta_0) + o_p^*(n^{-1}).
$$

Similarly, it can be shown that

$$
2\left(Q_n^*(\widetilde{\theta}^*) - Q_n^*(\theta_0)\right) = -(\widetilde{\theta}^* - \theta_0)'H(\widetilde{\theta}^* - \theta_0) + o_p^*(n^{-1}).
$$

By (i^*) and $\widehat{\theta}^* \xrightarrow{p^*} \theta_0$,

$$\begin{aligned} o_p^*(n^{-1}) \leq & Q_n^*(\widehat{\theta}^*) - Q_n(\widetilde{\theta}^*) \\ =& 2\left(Q_n^*(\widehat{\theta}^*) - Q_n^*(\theta_0)\right) - 2\left(Q_n^*(\widetilde{\theta}^*) - Q_n^*(\theta_0)\right) \\ =& (\widehat{\theta}^* - \widetilde{\theta}^*)' H (\widehat{\theta}^* - \widetilde{\theta}^*) + o_p^*(n^{-1}) \\ \leq & - C^* \|\widehat{\theta}^* - \widetilde{\theta}^*\| + o_p^*(n^{-1}) \end{aligned}$$

for some $C^* > 0$ (note that this is a generic positive constant). Thus, $\|\widehat{\theta}^* - \widetilde{\theta}^*\| = o_p^*(n^{-1/2})$. Since $\|\widehat{\theta} - \widetilde{\theta}\| = o_p(n^{-1/2})$, by the triangle inequality

$$\begin{aligned} \|\sqrt{n}(\widehat{\theta}^* - \widehat{\theta}) - (-H^{-1}\sqrt{n}D_n^*)\| =& \|\sqrt{n}(\widehat{\theta}^* - \widehat{\theta}) - \sqrt{n}(\widetilde{\theta}^* - \widetilde{\theta})\| \\ \leq & \sqrt{n}\|\widehat{\theta}^* - \widetilde{\theta}^*\| + \sqrt{n}\|\widehat{\theta} - \widetilde{\theta}\| = o_p^*(1). \end{aligned}$$

The conclusion follows by the Slutsky theorem. □

Proof of Theorem 1:

The proof proceeds by checking the conditions of Theorem 7.1 of Newey and McFadden (1994), which establish the asymptotic distribution of the maximizer of the nonsmooth sample criterion function. We refer the theorem to as TNM in this proof. We restate the conditions here for convenience: (0) for the sample criterion function $Q_n(\theta)$, $Q_n(\widehat{\theta}) \geq \sup_{\theta \in \Theta} Q_n(\theta) - o_p(n^{-1})$ and $\widehat{\theta} \xrightarrow{p} \theta_0$; (i) the population criterion function $Q(\theta)$ is maximized on the parameter space Θ at θ_0; (ii) θ_0 is an interior point of Θ; (iii) $Q(\theta)$ is twice differentiable at θ_0 with nonsingular second derivative H; (iv) $\sqrt{n}D_n \xrightarrow{d} N(0, \Omega)$ where D_n is a random vector that appears in the remainder term $R_n(\theta)$, analogous to the derivative of the sample criterion function, if exists; (v) for any $\delta_n \to 0$, $\sup_{||\theta - \theta_0|| \leq \delta_n} |R_n(\theta)/[1 + \sqrt{n}\|\theta - \theta_0\|]| \xrightarrow{p} 0$ where

$$R_n(\theta) = \sqrt{n}[Q_n(\theta) - Q_n(\theta_0) - (Q(\theta) - Q(\theta_0)) - D_n'(\theta - \theta_0)]//||\theta - \theta_0||.$$

First, we define

$$Q(\theta) = -g(\theta)'Wg(\theta)/2, \quad Q_n(\theta) = -g_n(\theta)'W_n g_n(\theta)/2 + \Delta_n(\theta)$$

where $\Delta_n(\theta) = \varepsilon_n(\theta)'W_n\varepsilon_n(\theta)/2 + (g_n(\theta_0) - g(\theta_0))'W_n\varepsilon_n(\theta) + g(\theta_0)'(W_n - W)\varepsilon_n(\theta)$, and $\varepsilon_n(\theta) = [g_n(\theta) - g_n(\theta_0) - (g(\theta) - g(\theta_0))]/[1 + \sqrt{n}||\theta - \theta_0||]$.

Note that for $\delta_n \to 0$, and $U = \{\theta : ||\theta - \theta_0|| \leq \delta_n\}$,

$$\begin{aligned}
\sup_U \sqrt{n}||\varepsilon_n(\theta)|| &= \sup_U \frac{\sqrt{n}||g_n(\theta) - g_n(\theta_0) - (g(\theta) - g(\theta_0))||}{1 + \sqrt{n}||\theta - \theta_0||} \\
&\leq \sup_U \frac{\sqrt{n}||g_n(\theta) - g_n(\theta_0) - (g(\theta) - g(\theta_0)) - (G_n - G)(\theta - \theta_0)||}{(1 + \sqrt{n}||\theta - \theta_0||)} + \sup_U \frac{\sqrt{n}||(G_n - G)(\theta - \theta_0)||}{(1 + \sqrt{n}||\theta - \theta_0||)} \\
&\leq \sup_U \frac{\sqrt{n}||g_n(\theta) - g_n(\theta_0) - (g(\theta) - g(\theta_0)) - (G_n - G)(\theta - \theta_0)||}{||\theta - \theta_0||(1 + \sqrt{n}||\theta - \theta_0||)} + ||G_n - G|| = o_p(1).
\end{aligned}$$

by Assumptions 1.6 and 1.7.

For any $\delta_n \to 0$,

$$\begin{aligned}
\sup_{||\theta-\theta_0||\leq\delta_n} |\Delta_n(\theta)| &= \sup_{||\theta-\theta_0||\leq\delta_n} |\varepsilon_n(\theta)' W_n \varepsilon_n(\theta)/2 + (g_n(\theta_0) - g(\theta_0))' W_n \varepsilon_n(\theta) + g(\theta_0)'(W_n - W)\varepsilon_n(\theta)| \\
&\leq O_p(1) \sup_{||\theta-\theta_0||\leq\delta_n} ||\varepsilon_n(\theta)||(||\varepsilon_n(\theta)|| + ||g_n(\theta_0) - g(\theta_0)|| + ||W_n - W||) = o_p(n^{-1})
\end{aligned}$$

by $\sup_{||\theta-\theta_0||\leq\delta_n} \sqrt{n}||\varepsilon_n(\theta)|| = o_p(1)$, $||g_n(\theta_0) - g(\theta_0)|| = O_p(n^{-1/2})$, and $||W_n - W|| = O_p(n^{-1/2})$ under Assumption 1.7. Then, we have $Q_n(\widehat{\theta}) \geq \sup_{||\theta-\theta_0||\leq\delta_n} Q_n(\theta) - o_p(1/n)$ by Assumption 1.4. Together with Assumption 1.5, Condition (0) of TNM is satisfied.

Conditions (i) and (ii) of TNM are satisfied by our Assumption 1.1.

By the second-order Taylor expansion, we obtain

$$g(\theta) = g(\theta_0) + G(\theta - \theta_0) + \frac{1}{2}\sum_{j=1}^{m}(\theta_j - \theta_j)\frac{\partial G}{\partial \theta_j}(\theta - \theta_0) + o(||\theta - \theta_0||^2).$$

where θ_j denotes the jth element of θ. It follows then,

$$\begin{aligned}
Q(\theta) =& -\frac{1}{2}\left[g(\theta_0) + G(\theta - \theta_0) + \frac{1}{2}\sum_{j=1}^{m}(\theta_j - \theta_j)\frac{\partial G}{\partial \theta_j}(\theta - \theta_0)\right]' W \\
&\times \left[g(\theta_0) + G(\theta - \theta_0) + \frac{1}{2}\sum_{j=1}^{m}(\theta_j - \theta_j)\frac{\partial G}{\partial \theta_j}(\theta - \theta_0)\right] + o(||\theta - \theta_0||^2) \\
=& -g(\theta_0)' W g(\theta_0)/2 - g(\theta_0)' W G(\theta - \theta_0) - (\theta - \theta_0)' G' W G(\theta - \theta_0)/2 \\
&- g(\theta_0)' W \sum_{j=1}^{m}(\theta_j - \theta_{j0})\frac{\partial G}{\partial \theta_j}(\theta - \theta_0)/2 + o(||\theta - \theta_0||^2) \\
=& Q(\theta_0) + (\theta - \theta_0)'\big(-G' W G - \big(g(\theta_0)' W \otimes I_p\big) F\big)(\theta - \theta_0)/2 + o(||\theta - \theta_0||^2) \\
=& Q(\theta_0) - (\theta - \theta_0)' H (\theta - \theta_0)/2 + o(||\theta - \theta_0||^2)
\end{aligned}$$

so that $Q(\theta)$ is twice differentiable at θ_0, where in the third equality we use $G' W g(\theta_0) = 0$ from the population FOC. Thus, Condition (iii) of TNM is satisfied.

For Condition (iv) of TNM, we define $D_n = -G_n' W_n g_n(\theta_0)$. Under $g(\theta_0) \neq 0$, we have

$$\sqrt{n} D_n = -\sqrt{n} G_n' W_n g_n(\theta_0) = -\sqrt{n} G_n' W_n [g_n(\theta_0) - g(\theta_0)] - \sqrt{n} G_n' W_n g(\theta_0)$$

$$= -\sqrt{n} G_n' W_n [g_n(\theta_0) - g(\theta_0)] - \sqrt{n} G_n' (W_n - W) g(\theta_0) - \sqrt{n}(G_n - G)' W g(\theta_0) \xrightarrow{d} N(0, \Omega)$$

by Assumption 1.7, the Slutzky theorem, and the population FOC, $G' W g(\theta_0) = 0$.

It remains to check Condition (v) of TNM. To do so, we first derive the terms included in the remainder $R_n(\theta)$. Since $g_n(\theta) = g_n(\theta_0) + g(\theta) - g(\theta_0) + \varepsilon_n(\theta)(1 + \sqrt{n}\|\theta - \theta_0\|)$,

$$\begin{aligned} g_n(\theta)' W_n g_n(\theta) =& (1 + \sqrt{n}||\theta - \theta_0||)^2 \varepsilon_n(\theta)' W_n \varepsilon_n(\theta) + g(\theta)' W_n g(\theta) \\ &+ (g_n(\theta_0) - g(\theta_0))' W_n (g_n(\theta_0) - g(\theta_0)) + 2(g_n(\theta_0) - g(\theta_0))' W_n g(\theta) \\ &+ 2[(g_n(\theta_0) - g(\theta_0)) + g(\theta)]' W_n \varepsilon_n(\theta)(1 + \sqrt{n}||\theta - \theta_0||). \end{aligned}$$

Since $\varepsilon_n(\theta_0) = 0$,

$$\begin{aligned} & g_n(\theta)' W_n g_n(\theta) - g_n(\theta_0)' W_n g_n(\theta_0) - 2\Delta_n(\theta) \\ =& (n\|\theta - \theta_0\|^2 + 2\sqrt{n}\|\theta - \theta_0\|) \varepsilon_n(\theta)' W_n \varepsilon_n(\theta) + 2(g_n(\theta_0) - g(\theta_0))' W_n \varepsilon_n(\theta) \sqrt{n}||\theta - \theta_0|| \\ &+ 2(g(\theta) - g(\theta_0))' W_n \varepsilon_n(\theta)(1 + \sqrt{n}||\theta - \theta_0||) + 2g(\theta_0)'(W_n - W)\varepsilon_n(\theta)\sqrt{n}||\theta - \theta_0|| \\ &+ 2g(\theta_0)' W \varepsilon_n(\theta)(1 + \sqrt{n}||\theta - \theta_0||) + 2(g_n(\theta_0) - g(\theta_0))' W_n (g(\theta) - g(\theta_0)) \\ &+ g(\theta)' W_n g(\theta) - g(\theta_0)' W_n g(\theta_0). \end{aligned}$$

It is also useful to note that

$$\begin{aligned} & g(\theta)' W_n g(\theta) - g(\theta_0)' W_n g(\theta_0) - g(\theta)' W g(\theta) + g(\theta_0)' W g(\theta_0) \\ =& (g(\theta) - g(\theta_0))'(W_n - W)(g(\theta) - g(\theta_0)) + 2(g(\theta) - g(\theta_0))'(W_n - W) g(\theta_0) \end{aligned}$$

Using these algebraic results, we have

$$\begin{aligned} & Q_n(\theta) - Q_n(\theta_0) - (Q(\theta) - Q(\theta_0)) - D_n'(\theta - \theta_0) \\ =& -g_n(\theta)' W_n g_n(\theta)/2 + \Delta_n(\theta) + g_n(\theta_0)' W_n g_n(\theta_0)/2 - \Delta_n(\theta_0) + g(\theta)' W g(\theta)/2 - g(\theta_0)' W g(\theta_0)/2 \\ &+ [G_n' W_n (g_n(\theta_0) - g(\theta_0)) + G_n'(W_n - W) g(\theta_0) + (G_n - G)' W g(\theta_0)]'(\theta - \theta_0) \\ =& \sum_{j=1}^{10} r_{jn}(\theta), \end{aligned}$$

where

$$\begin{aligned}
r_{1n}(\theta) &= -(n||\theta-\theta_0||^2 + 2\sqrt{n}||\theta-\theta_0||)\varepsilon_n(\theta)'W_n\varepsilon_n(\theta)/2 \\
r_{2n}(\theta) &= -[g_n(\theta_0)-g(\theta_0)]'W_n\varepsilon_n(\theta)\sqrt{n}||\theta-\theta_0|| \\
r_{3n}(\theta) &= -[g(\theta)-g(\theta_0)]'W_n\varepsilon_n(\theta)(1+\sqrt{n}||\theta-\theta_0||) \\
r_{4n}(\theta) &= -g(\theta_0)'(W_n-W)\varepsilon_n(\theta)\sqrt{n}||\theta-\theta_0|| \\
r_{5n}(\theta) &= -g(\theta_0)'W[\varepsilon_n(\theta)(1+\sqrt{n}||\theta-\theta_0||)-(G_n-G)(\theta-\theta_0)] \\
r_{6n}(\theta) &= -(g(\theta)-g(\theta_0))'[W_n-W](g(\theta)-g(\theta_0))/2 \\
r_{7n}(\theta) &= -[g(\theta)-g(\theta_0)-G(\theta-\theta_0)]'(W_n-W)g(\theta_0) \\
r_{8n}(\theta) &= [(G_n-G)(\theta-\theta_0)]'W_n(g_n(\theta_0)-g(\theta_0)) \\
r_{9n}(\theta) &= -[g(\theta)-g(\theta_0)-G(\theta-\theta_0)]'W_n(g_n(\theta_0)-g(\theta_0)) \\
r_{10n}(\theta) &= [(G_n-G)(\theta-\theta_0)]'(W_n-W)g(\theta_0)
\end{aligned}$$

using $\Delta_n(\theta_0)=0$, and $G'Wg(\theta_0)=0$. Condition (v) of TNM is satisfied if

$$\sup_{||\theta-\theta_0||\le\delta_n}\frac{|R_n(\theta)|}{1+\sqrt{n}||\theta-\theta_0||} = \sup_{||\theta-\theta_0||\le\delta_n}\frac{\sqrt{n}\sum_{j=1}^{10}|r_{jn}(\theta)|}{||\theta-\theta_0||(1+\sqrt{n}||\theta-\theta_0||)} = o_p(1). \tag{41}$$

We now show (41). For $r_{1n}(\theta)$,

$$\sup_U \frac{\sqrt{n}|r_{1n}(\theta)|}{||\theta-\theta_0||(1+\sqrt{n}||\theta-\theta_0||)} \le Cn\sup_U ||\varepsilon_n(\theta)||^2||W_n|| = o_p(1)$$

for some constant $C>0$ because $\sup_U \sqrt{n}||\varepsilon_n(\theta)|| = o_p(1)$ by Assumption 1.6 and 1.7. For $r_{2n}(\theta)$,

$$\sup_U \frac{\sqrt{n}|r_{2n}(\theta)|}{||\theta-\theta_0||(1+\sqrt{n}||\theta-\theta_0||)} \le \sqrt{n}||g_n(\theta_0)-g(\theta_0)||\cdot||W_n||\sup_U\sqrt{n}\|\varepsilon_n(\theta)\| = o_p(1)$$

by $||g_n(\theta_0)-g(\theta_0)|| = O_p(n^{-1/2})$. For $r_{3n}(\theta)$,

$$\sup_U \frac{\sqrt{n}|r_{3n}(\theta)|}{||\theta-\theta_0||(1+\sqrt{n}||\theta-\theta_0||)} \le \sup_U\frac{||g(\theta)-g(\theta_0)||}{||\theta-\theta_0||}\sup_U\sqrt{n}\|\varepsilon_n(\theta)\|\cdot||W_n|| = o_p(1)$$

by $||g(\theta)-g(\theta_0)|| = O(||\theta-\theta_0||)$. For $r_{4n}(\theta)$,

$$\sup_U \frac{\sqrt{n}|r_{4n}(\theta)|}{||\theta-\theta_0||(1+\sqrt{n}||\theta-\theta_0||)} \le ||g(\theta_0)||\sqrt{n}||W_n-W||\sup_U\sqrt{n}\|\varepsilon_n(\theta)\| = o_p(1)$$

by $||W_n - W|| = O_p(n^{-1/2})$. For $r_{5n}(\theta)$,

$$\begin{aligned}
&\sup_U \frac{\sqrt{n}|r_{5n}(\theta)|}{||\theta-\theta_0||(1+\sqrt{n}||\theta-\theta_0||)} \\
=&\sup_U \sqrt{n}\frac{|g(\theta_0)'W[g_n(\theta)-g_n(\theta_0)-(g(\theta)-g(\theta_0))-(G_n-G)(\theta-\theta_0)]|}{||\theta-\theta_0||(1+\sqrt{n}||\theta-\theta_0||)} \\
\leq&||g(\theta_0)||\cdot||W||\sup_U \sqrt{n}\frac{||[g_n(\theta)-g_n(\theta_0)-(g(\theta)-g(\theta_0))-(G_n-G)(\theta-\theta_0)]||}{||\theta-\theta_0||(1+\sqrt{n}||\theta-\theta_0||)} = o_p(1)
\end{aligned}$$

where the last inequality holds by Assumption 1.6. For $r_{6n}(\theta)$,

$$\sup_U \frac{\sqrt{n}|r_{6n}(\theta)|}{||\theta-\theta_0||(1+\sqrt{n}||\theta-\theta_0||)} \leq \sup_U ||g(\theta)-g(\theta_0)||\sqrt{n}||W_n-W|| = o_p(1),$$

by $||g(\theta)-g(\theta_0)|| = O(||\theta-\theta_0||)$. For $r_{7n}(\theta)$,

$$\begin{aligned}
\sup_U \frac{\sqrt{n}|r_{7n}(\theta)|}{||\theta-\theta_0||(1+\sqrt{n}||\theta-\theta_0||)} \leq&\sup_U \frac{||g(\theta)-g(\theta_0)-G(\theta-\theta_0)||}{||\theta-\theta_0||}\sqrt{n}||W_n-W||\cdot||g(\theta_0)|| \\
=&\sup_U O(||\theta-\theta_0||)O_p(1) = o_p(1).
\end{aligned}$$

For $r_{8n}(\theta)$,

$$\sup_U \frac{\sqrt{n}|r_{8n}(\theta)|}{||\theta-\theta_0||(1+\sqrt{n}||\theta-\theta_0||)} \leq \sup_U \sqrt{n}||G_n-G||\cdot||W_n||\cdot||g_n(\theta_0)-g(\theta_0)|| = o_p(1).$$

For $r_{9n}(\theta)$,

$$\begin{aligned}
\sup_U \frac{\sqrt{n}|r_{9n}(\theta)|}{||\theta-\theta_0||(1+\sqrt{n}||\theta-\theta_0||)} \leq&\sup_U \frac{||g(\theta)-g(\theta_0)-G(\theta-\theta_0)||}{||\theta-\theta_0||}||W_n||\sqrt{n}||g_n(\theta_0)-g(\theta_0)|| \\
=&\sup_U O(||\theta-\theta_0||)O_p(1) = o_p(1).
\end{aligned}$$

Finally, for $r_{10n}(\theta)$,

$$\sup_U \frac{\sqrt{n}|r_{10n}(\theta)|}{||\theta-\theta_0||(1+\sqrt{n}||\theta-\theta_0||)} \leq \sup_U \sqrt{n}||G_n-G||\cdot||W_n-W||\cdot||g(\theta_0)|| = o_p(1).$$

Thus, the conclusion follows by TNM and we have

$$\sqrt{n}(\widehat{\theta}-\theta_0) = H^{-1}\sqrt{n}D_n + o_p(1) \xrightarrow{d} N(0, H^{-1}\Omega H^{-1}).$$

This completes the proof. □

Proof of Theorem 2: The proof proceeds by showing the following steps. For the GMM estimator $\overline{\theta}_n(\phi)$ defined in (7), (i) $\sup_{\phi\in\Theta}\left\|\overline{\theta}_n(\phi)-\theta(\phi)\right\| \xrightarrow{p} 0$; (ii) With probability tending to one, the map $\overline{\theta}_n(\phi)$ is a contraction and the fixed point $\widehat{\theta}$ exists and is unique; (iii) $\|\widehat{\theta}-\theta_0\| \xrightarrow{p} 0$.

The proof of (i) is identical to that of Theorem 3.1 of Hansen and Lee (2021) because the uniform convergence of $g_n(\theta)$ and $W_n(\theta)$ hold by our Assumption 2, the uniform invertibility of $W(\phi)$ holds by our Assumption 1.3, and the existence of unique minimizer of the population s-step GMM criterion holds by Assumption 1.1.

Since we do not assume differentiability of the moment function, the proof of (ii) proceeds differently from that of Theorem 3.2 of Hansen and Lee (2021). We show the following: For some $0 \le c < 1$ and for any ϕ_1, $\phi_2 \in \Theta$, with probability tending to one,

$$\left\|\overline{\theta}_n(\phi_1) - \overline{\theta}_n(\phi_2)\right\| \le c \left\|\phi_1 - \phi_2\right\|. \tag{42}$$

Since $\phi_1 = \phi_2$ is trivial, suppose that $\phi_1 \neq \phi_2$. Suppose that (17) holds with $0 \le c_0 < 1$. For any ϕ_1, $\phi_2 \in \Theta$, (i) implies that for a large enough n,

$$\sup_{\phi\in\Theta} \left\|\overline{\theta}_n(\phi) - \theta(\phi)\right\| \le \frac{1-c_0}{3}\|\phi_1 - \phi_2\|. \tag{43}$$

Now by the triangle inequality, (17), and (43),

$$\begin{aligned}
\left\|\overline{\theta}_n(\phi_1) - \overline{\theta}_n(\phi_2)\right\| &= \left\|\theta(\phi_1) - \theta(\phi_2) + \overline{\theta}_n(\phi_1) - \overline{\theta}_n(\phi_2) - \theta(\phi_1) + \theta(\phi_2)\right\| \\
&\le \|\theta(\phi_1) - \theta(\phi_2)\| + \left\|\overline{\theta}_n(\phi_1) - \theta(\phi_1) - (\overline{\theta}_n(\phi_2) - \theta(\phi_2))\right\| \\
&\le c_0 \|\phi_1 - \phi_2\| + 2 \sup_{\phi\in\Theta} \left\|\overline{\theta}_n(\phi) - \theta(\phi)\right\| \\
&\le \frac{c_0+2}{3}\|\phi_1 - \phi_2\|
\end{aligned}$$

with probability tending to one. Since the map $\overline{\theta}_n(\phi)$ is a contraction with probability tending to one, the fixed point exists and is unique by the Banach fixed point theorem.

Lastly, the proof of (iii) is identical to that of Theorem 3.3 of Hansen and Lee (2021). □

Proof of Theorem 3:

Since the iterated GMM criterion function takes a different form with those considered in Theorem 7.1 of Newey and McFadden (1994), we cannot proceed by directly checking the conditions of Theorem 7.1. as we did in the proof of Theorem 1. Instead, the proof proceeds by taking similar steps with that of Theorem 7.1 of Newey and McFadden (1994) but extending their proof to allow for the iterated estimator under moment misspecification.

First, we want to show that $\sqrt{n}||\widehat{\theta} - \theta_0|| = O_p(1)$. We define the sample and the population criterion function as follows:

$$\begin{aligned}
Q_n(\theta,\phi) &= - J_n(\theta,\phi)/2 + \Delta_n(\theta,\phi) = -g_n(\theta)'W_n(\phi)g_n(\theta)/2 + \Delta_n(\theta,\phi) \\
Q(\theta,\phi) &= - g(\theta)'W(\phi)g(\theta)/2,
\end{aligned}$$

where $\Delta_n(\theta,\phi) = \varepsilon_n(\theta)'W_n(\phi)\varepsilon_n(\theta)/2 + (g_n(\theta_0) - g(\theta_0))'W_n(\phi)\varepsilon_n(\theta) + g(\theta_0)'(W_n(\phi) - W)\varepsilon_n(\theta)$, and $\varepsilon_n(\theta) = [g_n(\theta) - g_n(\theta_0) - (g(\theta) - g(\theta_0))]/[1 + \sqrt{n}||\theta - \theta_0||]$.

Also define

$$R_n(\theta, \widehat{\theta}) = [Q_n(\theta, \widehat{\theta}) - Q_n(\theta_0, \widehat{\theta}) - (Q(\theta, \theta_0) - Q(\theta_0, \theta_0)) - D_n'(\theta - \theta_0)]/||\theta - \theta_0||$$

where $D_n = -G_n' W_n(\widehat{\theta}) g_n(\theta_0)$. Under $g(\theta_0) \neq 0$, we have

$$\begin{aligned}\sqrt{n}D_n =& -\sqrt{n}G_n'W_n(\widehat{\theta})[g_n(\theta_0) - g(\theta_0)] - \sqrt{n}G_n'(W_n(\widehat{\theta}) - W)g(\theta_0) - \sqrt{n}(G_n - G)'Wg(\theta_0) \\ &\xrightarrow{d} N(0, \Omega)\end{aligned}$$

by Assumption 1.7, the Slutsky theorem, and using $G'Wg(\theta_0) = 0$ by the population FOC.

Unlike Theorem 7.1 of Newey and McFadden (1994) where the population criterion function depend only on θ, our population criterion function depends on θ and ϕ, and ϕ converges to the fixed point as we iterate. The second-order Taylor expansion of $Q(\theta, \phi = \theta_{s-1})$ with respect to θ evaluated at $\theta = \theta_s$ for any $s \geq 2$,

$$\begin{aligned}Q(\theta, \theta_{s-1}) =& Q(\theta_s, \theta_{s-1}) + \left.\frac{\partial Q(\theta, \phi)}{\partial \theta'}\right|_{\theta=\theta_s, \phi=\theta_{s-1}} (\theta - \theta_s) \\ &+ \frac{1}{2}(\theta - \theta_s)' \left.\frac{\partial^2 Q(\theta, \phi)}{\partial \theta \partial \theta'}\right|_{\theta=\theta_s, \phi=\theta_{s-1}} (\theta - \theta_s) + o(||\theta - \theta_s||^2).\end{aligned}$$

Let $s \to \infty$, then we have

$$\begin{aligned}Q(\theta, \theta_0) =& Q(\theta_0, \theta_0) + \left.\frac{\partial Q(\theta, \phi)}{\partial \theta'}\right|_{\phi=\theta=\theta_0} (\theta - \theta_0) \\ &+ \frac{1}{2}(\theta - \theta_0)' \left.\frac{\partial m(\theta)}{\partial \theta}\right|_{\theta=\theta_0} (\theta - \theta_0) + o(||\theta - \theta_0||^2)\end{aligned} \tag{44}$$

where $m(\theta) \equiv \left.\frac{\partial Q(\theta,\phi)}{\partial \theta}\right|_{\phi=\theta} = -G(\theta)'W(\theta)g(\theta)$. Using the alternative representations of $m(\theta)$,

$$G(\theta)'W(\theta)g(\theta) = (g(\theta)' \otimes G(\theta)') \operatorname{vec} W(\theta) = (g(\theta)'W(\theta) \otimes I_p) \operatorname{vec} G(\theta)',$$

and the identity

$$\frac{\partial}{\partial \theta'} \operatorname{vec} W(\theta) = -(W(\theta) \otimes W(\theta)) \frac{\partial}{\partial \theta'} \operatorname{vec}(W(\theta)^{-1}),$$

we have

$$\frac{\partial m(\theta)}{\partial \theta'} = \frac{\partial^2 Q(\theta, \phi)}{\partial \theta \partial \theta'} + \left.\frac{\partial^2 Q(\theta, \phi)}{\partial \phi \partial \theta'}\right|_{\phi=\theta} \tag{45}$$

$$= -\big(G(\theta)'W(\theta)G(\theta) + (g(\theta)'W(\theta) \otimes I_p)F(\theta) - (g(\theta)'W(\theta) \otimes G(\theta)'W(\theta))S(\theta)\big) \tag{46}$$

$$\equiv -H(\theta), \tag{47}$$

where $F(\theta) = \frac{\partial}{\partial \theta'} \text{vec}(G(\theta)')$ and $S(\theta) = \frac{\partial}{\partial \theta'} \text{vec}(W(\theta)^{-1})$. Since $G'Wg(\theta_0) = 0$, (44) becomes

$$Q(\theta, \theta_0) = Q(\theta_0, \theta_0) - (\theta - \theta_0)' H(\theta - \theta_0)/2 + o(||\theta - \theta_0||^2) \tag{48}$$

where $H = H(\theta_0)$. By θ_0 being the maximum and H being nonsingular, $H > 0$.[11] Therefore, by $\widehat{\theta} \xrightarrow{p} \theta_0$, with probability approaching one,

$$Q(\widehat{\theta}, \theta_0) \leq Q(\theta_0, \theta_0) - C||\widehat{\theta} - \theta_0||^2$$

for some $C > 0$.

Choose U_n so that $\widehat{\theta} \in U_n$ with probability approaching 1, so that

$$\sqrt{n}|R_n(\theta, \widehat{\theta})| \leq (1 + \sqrt{n}||\widehat{\theta} - \theta_0||)o_p(1), \tag{49}$$

which holds by the same arguments to show (41) by replacing W_n with $W_n(\widehat{\theta})$. It can be also shown that $\sup_{||\theta - \theta_0|| \leq \delta_n} |\Delta_n(\theta, \widehat{\theta})| = o_p(n^{-1})$ for any $\delta_n \to 0$, similarly as in the proof of Theorem 1 by replacing W_n with $W_n(\widehat{\theta})$ under the Assumption 1.7, and thus $Q_n(\widehat{\theta}, \widehat{\theta}) \geq \sup_{||\theta - \theta_0|| \leq \delta_n} Q_n(\theta, \widehat{\theta}) - o_p(1/n)$ by the Assumption 1.4.

Then we have

$$\begin{aligned}
0 \leq & Q_n(\widehat{\theta}, \widehat{\theta}) - Q_n(\theta_0, \widehat{\theta}) + o_p(n^{-1}) = Q(\widehat{\theta}, \theta_0) - Q(\theta_0, \theta_0) + D_n'(\widehat{\theta} - \theta_0) + R_n(\widehat{\theta}, \widehat{\theta})||\widehat{\theta} - \theta_0|| + o_p(n^{-1}) \\
\leq & - C||\widehat{\theta} - \theta_0||^2 + ||D_n|| \, ||\widehat{\theta} - \theta_0|| + ||\widehat{\theta} - \theta_0||(1 + \sqrt{n}||\widehat{\theta} - \theta_0||)o_p(n^{-1/2}) + o_p(n^{-1}) \\
\leq & - [C + o_p(1)]||\widehat{\theta} - \theta_0||^2 + O_p(n^{-1/2})||\widehat{\theta} - \theta_0|| + o_p(n^{-1}).
\end{aligned}$$

Thus, we have $||\widehat{\theta} - \theta_0||^2 \leq O_p(n^{-1/2})||\widehat{\theta} - \theta_0|| + o_p(n^{-1})$ because $C + o_p(1)$ is bounded away from zero with probability approaching one. Then, by the same arguments in the proof of Theorem 7.1 in Newey and McFadden (1994), we have $||\widehat{\theta} - \theta_0|| = O_p(n^{-1/2})$.

Next, let $\widetilde{\theta} = \theta_0 + H^{-1}D_n$, and then $\widetilde{\theta}$ is $\sqrt{n}$ consistent by construction. It follows that

$$\begin{aligned}
2[Q_n(\widehat{\theta}, \widehat{\theta}) - Q_n(\theta_0, \widehat{\theta})] &= - (\widehat{\theta} - \theta_0)' H(\widehat{\theta} - \theta_0) + 2D_n'(\widehat{\theta} - \theta_0) + o_p(n^{-1}) \\
&= - (\widehat{\theta} - \theta_0)' H(\widehat{\theta} - \theta_0) + 2(\widetilde{\theta} - \theta_0)' H(\widehat{\theta} - \theta_0) + o_p(n^{-1})
\end{aligned} \tag{50}$$

by (48) and (49) . Similarly,

$$\begin{aligned}
2[Q_n(\widetilde{\theta}, \widehat{\theta}) - Q_n(\theta_0, \widehat{\theta})] &= -(\widetilde{\theta} - \theta_0)' H(\widetilde{\theta} - \theta_0) + 2(\widetilde{\theta} - \theta_0)' H(\widetilde{\theta} - \theta_0) + o_p(n^{-1}) \\
&= (\widetilde{\theta} - \theta_0)' H(\widetilde{\theta} - \theta_0) + o_p(n^{-1}).
\end{aligned} \tag{51}$$

Then, since $\widetilde{\theta}$ is contained within Θ with probability approaching one, by Assumption 1.4,

$$2[Q_n(\widehat{\theta}, \widehat{\theta}) - Q_n(\theta_0, \widehat{\theta})] - 2[Q_n(\widetilde{\theta}, \widehat{\theta}) - Q_n(\theta_0, \widehat{\theta})] \geq o_p(n^{-1}),$$

[11]Note that our definition of H corresponds to $-H$ in Theorem 7.1 of Newey and McFadden (1994).

so by using (50) and (51),

$$o_p(n^{-1}) \leq -(\widehat{\theta}-\theta_0)'H(\widehat{\theta}-\theta_0) + 2(\widetilde{\theta}-\theta_0)'H(\widehat{\theta}-\theta_0) - (\widetilde{\theta}-\theta_0)'H(\widetilde{\theta}-\theta_0)$$
$$= -(\widehat{\theta}-\widetilde{\theta})'H(\widehat{\theta}-\widetilde{\theta}) \leq -C||\widehat{\theta}-\widetilde{\theta}||^2.$$

Therefore, $||\sqrt{n}(\widehat{\theta}-\theta_0)-(H^{-1}\sqrt{n}D_n)|| = \sqrt{n}||\widehat{\theta}-\widetilde{\theta}|| \xrightarrow{p} 0$, so the conclusion follows by the Slutsky theorem and $H^{-1}\sqrt{n}D_n \xrightarrow{d} N(0, H^{-1}\Omega H^{-1})$. □

Proof of Theorem 4:

First we define,

$$Q_n^*(\theta) = -g_n^*(\theta)'W_n^*g_n^*(\theta)/2 + \Delta_n^*(\theta), \quad Q_n(\theta) = -g_n(\theta)'W_ng_n(\theta)/2 + \Delta_n(\theta)$$

where $\Delta_n^*(\theta) = \varepsilon_n^*(\theta)'W_n^*\varepsilon_n(\theta)/2 + (g_n^*(\theta_0)-g(\theta_0))'W_n^*\varepsilon_n^*(\theta) + g(\theta_0)'(W_n^*-W)\varepsilon_n^*(\theta)$, $\varepsilon_n^*(\theta) = [g_n^*(\theta) - g_n^*(\theta_0) - (g(\theta)-g(\theta_0))]/[1+\sqrt{n}||\theta-\theta_0||]$. $Q(\theta), \Delta_n(\theta), \varepsilon_n(\theta)$ are defined in the proof of Theorem 1.

The rest of the proof proceeds by verifying the conditions of Lemma 2 using the results in the proof of Theorem 1. Note that for $\delta_n \to 0$, and $U = \{\theta : ||\theta-\theta_0|| \leq \delta_n\}$,

$$\begin{aligned}
&\sup_U \sqrt{n}||\varepsilon_n^*(\theta)|| \\
&= \sup_U \frac{\sqrt{n}||g_n^*(\theta) - g_n^*(\theta_0) - (g(\theta)-g(\theta_0))||}{1+\sqrt{n}||\theta-\theta_0||} \\
&\leq \sup_U \frac{\sqrt{n}||g_n^*(\theta) - g_n^*(\theta_0) - (g_n(\theta)-g_n(\theta_0))||}{(1+\sqrt{n}||\theta-\theta_0||)} + \sup_U \frac{\sqrt{n}||g_n(\theta) - g_n(\theta_0) - (g(\theta)-g(\theta_0))||}{(1+\sqrt{n}||\theta-\theta_0||)} \\
&\leq \sup_U \frac{\sqrt{n}||g_n^*(\theta) - g_n^*(\theta_0) - (g_n(\theta)-g_n(\theta_0)) - (G_n^*-G_n)(\theta-\theta_0)||}{||\theta-\theta_0||(1+\sqrt{n}||\theta-\theta_0||)} + ||G_n^*-G_n|| + o_p(1) \\
&= o_{p^*}(1)
\end{aligned}$$

by the triangle inequality, Conditions 3 and 4 of the Theorem, and $\sup_U \sqrt{n}||\varepsilon_n(\theta)|| = o_p(1)$. Also let

$$\begin{aligned}
D_n^* &= -G_n^{*\prime}W_n^*g_n^*(\theta_0) + G_n'W_ng_n(\theta_0) \\
&= -G_n^{*\prime}W_n^*(g_n^*(\theta_0)-g_n(\theta_0)) - G_n^{*\prime}(W_n^*-W_n)g_n(\theta_0) - (G_n^*-G_n)'W_ng_n(\theta_0), \qquad (52) \\
D_n &= -G_n'W_ng_n(\theta_0) + G'Wg(\theta_0) \\
&= -G_n'W_n(g_n(\theta_0)-g(\theta_0)) - G_n'(W_n-W)g(\theta_0) - (G_n-G)'Wg(\theta_0). \qquad (53)
\end{aligned}$$

Conditions (i)-(iii) of Lemma 2 hold under the same assumptions in Theorem 1. For Conditions (i^*)-(iii^*), first note that the bootstrap version of Assumption 1.7 holds due to Giné and Zinn (1990):

$$\sqrt{n}\begin{pmatrix} g_n^*(\theta_0)-g_n(\theta_0) \\ (W_n^*-W_n)g_n(\theta_0) \\ (G_n^*-G_n)'W_ng_n(\theta_0) \end{pmatrix} \xrightarrow{d^*} N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma & \Lambda & \Gamma \\ \Lambda' & \Psi & \Upsilon \\ \Gamma' & \Upsilon' & \Xi \end{pmatrix}\right). \qquad (54)$$

Therefore, Condition (i^*) holds because $\sup_{||\theta-\theta_0||\le\delta_n}|\Delta_n^*(\theta)| = o_{p^*}(1/n)$ by using $||g_n^*(\theta_0)-g(\theta_0)|| \le ||g_n^*(\theta_0)-g_n(\theta_0)|| + ||g_n(\theta)-g(\theta_0)|| = O_{p^*}(n^{-1/2}), ||W_n^*-W|| \le ||W_n^*-W_n|| + ||W_n-W|| = O_{p^*}(n^{-1/2})$, and thus $Q_n^*(\widehat{\theta}^*) \ge \sup_{||\theta-\theta_0||\le\delta_n} Q_n^*(\theta) - o_{p^*}(1/n)$. Condition (ii^*) follows by (54). Thus, we remain to verify Condition (iii^*) of Lemma 2.

By (36),

$$
\begin{aligned}
&R_n^*(\theta)\|\theta-\theta_0\| \\
=&Q_n^*(\theta)-Q_n^*(\theta_0)-(Q_n(\theta)-Q_n(\theta_0))-D_n^{*\prime}(\theta-\theta_0) \\
=&Q_n^*(\theta)-Q_n^*(\theta_0)-(Q(\theta)-Q(\theta_0)-(D_n^*+D_n)'(\theta-\theta_0) \\
&-\big(Q_n(\theta)-Q_n(\theta_0)-(Q(\theta)-Q(\theta_0))-D_n'(\theta-\theta_0)\big) \\
=&Q_n^*(\theta)-Q_n^*(\theta_0)-(Q(\theta)-Q(\theta_0)) \\
&+[G_n^{*\prime}W_n^*(g_n^*(\theta_0)-g(\theta_0))+G_n^{*\prime}(W_n^*-W)g(\theta_0)+(G_n^*-G)'Wg(\theta_0)]'(\theta-\theta_0)-R_n(\theta)(\theta-\theta_0) \\
=&\sum_{j=1}^{10} r_{jn}^*(\theta)-R_n(\theta)(\theta-\theta_0),
\end{aligned}
$$

where $r_{jn}^*(\theta)$ is similarly defined as in the proof of Theorem 1 by replacing $g_n(\theta), G_n, W_n, \varepsilon_n(\theta)$ with $g_n^*(\theta), G_n^*(\theta), W_n^*(\theta), \varepsilon_n^*(\theta)$, respectively. Using the same argument as in the proof of (41), we can show that

$$
\sup_{||\theta-\theta_0||\le\delta_n} \frac{\sqrt{n}\sum_{j=1}^{10}|r_{jn}^*(\theta)|}{||\theta-\theta_0||(1+\sqrt{n}||\theta-\theta_0||)} = o_{p^*}(1). \tag{55}
$$

For example,

$$
\begin{aligned}
&\sup_U \frac{\sqrt{n}|r_{5n}^*(\theta)|}{||\theta-\theta_0||(1+\sqrt{n}||\theta-\theta_0||)} \\
=&\sup_U \sqrt{n}\frac{|g(\theta_0)'W[g_n^*(\theta)-g_n^*(\theta_0)-(g(\theta)-g(\theta_0))-(G_n^*-G)(\theta-\theta_0)]|}{||\theta-\theta_0||(1+\sqrt{n}||\theta-\theta_0||)} \\
\le&||g(\theta_0)||\cdot||W|| \ \sup_U \sqrt{n}\frac{||[g_n^*(\theta)-g_n^*(\theta_0)-(g(\theta)-g(\theta_0))-(G_n^*-G)(\theta-\theta_0)]||}{||\theta-\theta_0||(1+\sqrt{n}||\theta-\theta_0||)} \\
\le&||g(\theta_0)||\cdot||W|| \ \big[\sup_U \sqrt{n}\frac{||[g_n^*(\theta)-g_n^*(\theta_0)-(g_n(\theta)-g_n(\theta_0))-(G_n^*-G_n)(\theta-\theta_0)]||}{||\theta-\theta_0||(1+\sqrt{n}||\theta-\theta_0||)} \\
&+\sup_U \sqrt{n}\frac{||[g_n(\theta)-g_n(\theta_0)-(g(\theta)-g(\theta_0))-(G_n-G)(\theta-\theta_0)]||}{||\theta-\theta_0||(1+\sqrt{n}||\theta-\theta_0||)}\big] \\
=&o_{p^*}(1),
\end{aligned}
$$

where the last equality holds by Assumption 1.6 (stochastic differentiability) and Condition 3 of the Theorem (the bootstrap stochastic differentiability).

Therefore, for $\delta_n \to 0$, and $U = \{\theta : ||\theta-\theta_0|| \le \delta_n\}$, we have

$$
\sup_U \left|\frac{\sqrt{n}R_n^*(\theta)}{1+\sqrt{n}\|\theta-\theta_0\|}\right| \le \sup_U \left|\frac{\sqrt{n}\sum_{j=1}^{10}|r_{jn}^*(\theta)|}{||\theta-\theta_0||(1+\sqrt{n}||\theta-\theta_0||)}\right| + \sup_U \left|\frac{\sqrt{n}R_n(\theta)}{1+\sqrt{n}\|\theta-\theta_0\|}\right| = o_{p^*}(1).
$$

The conclusion follows by Lemma 2 and we have

$$\sqrt{n}(\widehat{\theta}^* - \widehat{\theta}) \xrightarrow{d^*} N(0, H^{-1}\Omega H^{-1}).$$

This completes the proof. □

Proof of Theorem 5:

We first verify the assumptions of Theorem 1. By the standard consistency arguments (Newey and McFadden, 1994, Theorem 2.6) and the Central Limit Theorem, Assumption 1.5 and Assumption 1.7 are satisfied with $G_n = \sum_{i=1}^n G_i(\theta_0)$, where

$$G_i(\theta) = \begin{pmatrix} -\frac{1}{\sigma(X_i'\gamma)} Z_i X_i' & -\frac{\sigma'(X_i'\gamma)(Y_i - X_i'\beta)}{\sigma(X_i'\gamma)^2} Z_i X_i' \\ -\frac{1}{\sigma(X_i'\gamma)} sgn\left(\frac{Y_i - X_i'\beta}{\sigma(X_i'\gamma)}\right) Z_i X_i' & -\frac{\sigma'(X_i'\gamma)(Y_i - X_i'\beta)}{\sigma(X_i'\gamma)^2} sgn\left(\frac{Y_i - X_i'\beta}{\sigma(X_i'\gamma)}\right) Z_i X_i' \end{pmatrix},$$

$sgn(x) = 1\{x \geq 0\} - 1\{x \leq 0\}$ is a sign function.

Let

$$\mathcal{F} = \{r_i(\cdot, \theta) = \frac{g_i(\theta) - g_i(\theta_0) - G_i(\theta_0)(\theta - \theta_0)}{||\theta - \theta_0||} : ||\theta - \theta_0|| < \delta\}$$

be a class of functions indexed by $\theta = (\beta', \gamma')' \in \Theta$ for some $\delta > 0$, where $g_i(\theta)$ and $G_i(\theta)$ are defined in the Theorem 5. As the moment function $g_i(\theta)$ involves the absolute function, we can verify that $|r_i(\cdot, \theta)|$ are uniformly bounded, $\sup_\theta |r_i(\cdot, \theta)|$ has a finite second moment under the conditions 1 and 3 in Theorem 5. Further, we can show $\mathcal{F}$ is a Vapnik-Chervonenkis (VC) class by using the same arguments as in Pollard (1985, Example 8), which implies Assumption 1.6.

Thus, all assumptions of Theorem 1 hold, and thus we have

$$\sqrt{n}(\widehat{\theta} - \theta_0) = -H^{-1}\sqrt{n}D_n + o_p(1) = -H^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n m_i + o_p(1) \tag{56}$$

where

$$m_i = G'Wg_i(\theta_0) + G_i(\theta_0)'Wg(\theta_0) - G'Wv_iv_i'Wg(\theta_0).$$

Under the Assumptions in Theorem 5, using the same arguments in the proof of Machado and Santos Silva (2019, Theorem 5), we have the following linear representation conditional on $\sqrt{n}$ consistent estimators of $\theta_0 = (\beta_0', \gamma_0')'$,

$$\sqrt{n}(\widehat{q}(\tau) - q_0(\tau)) = -\frac{1}{f_U(q_0(\tau))}\frac{1}{\sqrt{n}}\sum_{i=1}^n \tau - 1(U_i \leq q_0(\tau)) + o_p(1) \tag{57}$$

where $f_U(\cdot)$ is a density function of the random variable U_i. Note that (56) and (57) hold jointly, and thus we have the limiting distribution as in (31). Furthermore, asymptotic distribution of the regression quantile coefficient estimator $\widehat{\alpha}(\tau)$ in (32) is obtained by the standard delta-method. This completes the proof. $\square$

Appendix B: Covariance Matrix Estimation

We now consider estimation of the covariance matrix of the GMM estimators. First, we consider the one-step GMM estimator $\widehat{\theta}$. Let $\widehat{\Omega}$ be an estimator of Ω, where

$$\Omega = \lim_{n\to\infty} Var\left(\sqrt{n}\{G_n' W_n[g_n(\theta_0) - g(\theta_0)] + G_n'(W_n - W)g(\theta_0) + (G_n - G)'Wg(\theta_0)\right).$$

If $g_n(\theta) = n^{-1}\sum_{i=1}^n g(X_i,\theta)$ and $g(X_i,\theta)$ be Lipschitz at θ_0 and differentiable with probability one (see Lemma 1), i.e., there exists $\Delta(x,\theta_0)$ and $\varepsilon > 0$ such that with probability one $r(x,\theta) \equiv ||g(x,\theta) - g(x,\theta_0) - \Delta(x,\theta_0)(\theta-\theta_0)||/||\theta-\theta_0|| \to 0$ as $\theta \to \theta_0$, then the following estimator can be used

$$\widehat{\Omega} = \frac{1}{n}\sum_{i=1}^n \hat{m}_i \hat{m}_i',$$
$$\hat{m}_i = \widehat{G}' W_n g_i(\widehat{\theta}) + G_i(\widehat{\theta})' W_n g_n(\widehat{\theta}) - \widehat{G}' W_n W(X_i) W_n g_n(\widehat{\theta}) \tag{58}$$

where $\widehat{G} = n^{-1}\sum_{i=1}^n G_i(\widehat{\theta})$, $G_i(\theta) = \Delta(X_i,\theta)$. For example, we can construct the asymptotic covariance matrix for GMM-QR estimator using (58) in Section 5.

Numerical derivative methods can be used to estimate G, H and are consistent under weak conditions (see, for example, Newey and McFadden (1994), Hong, Mahajan, and Nekipelov (2015)). A numerical derivative estimator $\widehat{G}$ for G has jth column,

$$\widehat{G}_j = g_n(\widehat{\theta} + e_j\varepsilon_n) - g_n(\widehat{\theta} - e_j\varepsilon_n)/2\varepsilon_n$$

where e_j is the ith unit vector, ε_n is a small positive constant that depends on the sample size. Similarly, H can be estimated by a second-order numerical derivative of the criterion function $J_n(\theta) = g_n(\theta)' W_n g_n(\theta)$,

$$\widehat{H}_{i,j} = \left[J_n(\widehat{\theta} + e_i\varepsilon_n + e_j\varepsilon_n) - J_n(\widehat{\theta} - e_i\varepsilon_n + e_j\varepsilon_n) - J_n(\widehat{\theta} + e_i\varepsilon_n - e_j\varepsilon_n) + J_n(\widehat{\theta} - e_i\varepsilon_n - e_j\varepsilon_n)\right]/4\varepsilon_n^2.$$

Alternatively, H can be estimated by the explicit formula

$$\widehat{H} = \widehat{G}' W_n \widehat{G} + (g_n(\widehat{\theta})' W_n \otimes I_p)\widehat{F}$$

where $\widehat{F}$ is a numerical derivative estimator for F, which is the $mp \times p$ matrix with i, jth block column matrix

$$\widehat{F}_{i,j} = \left[g_n(\widehat{\theta} + e_i\varepsilon_n + e_j\varepsilon_n) - g_n(\widehat{\theta} - e_i\varepsilon_n + e_j\varepsilon_n) - g_n(\widehat{\theta} + e_i\varepsilon_n - e_j\varepsilon_n) + g_n(\widehat{\theta} - e_i\varepsilon_n - e_j\varepsilon_n)\right]/4\varepsilon_n^2.$$

Then, we construct the variance estimator as

$$\widehat{V}_{mr}(\widehat{\theta}) = \widehat{H}^{-1}\widehat{\Omega}\widehat{H}^{-1\prime}$$

and the standard error is obtained by taking the diagonal elements of $\sqrt{\widehat{V}_{mr}(\widehat{\theta})/n}$.[12]

Next, we consider the iterated GMM estimator. Similar to the one-step GMM case where G_n takes the form of sample averages, we construct the following misspecification-robust variance estimator

$$\begin{aligned}
&\widehat{V}_{mr}(\widehat{\theta}) = \widehat{H}^{-1}\widehat{\Omega}\widehat{H}^{-1\prime}, \\
&\widehat{H} = \widehat{G}'W_n(\widehat{\theta})\widehat{G} + (g_n(\widehat{\theta})'W_n(\widehat{\theta}) \otimes I_p)\widehat{F} - (g_n(\widehat{\theta})'W_n(\widehat{\theta}) \otimes \widehat{G}'W_n(\widehat{\theta}))\widehat{S}, \\
&\widehat{\Omega} = \frac{1}{n}\sum_{i=1}^{n} \hat{m}_i \hat{m}_i', \\
&\hat{m}_i = \widehat{G}'W_n(\widehat{\theta})g_i(\widehat{\theta}) + G_i(\widehat{\theta})'W_n(\widehat{\theta})g_n(\widehat{\theta}) - \widehat{G}'W_n(\widehat{\theta})W(X_i,\widehat{\theta})W_n(\widehat{\theta})g_n(\widehat{\theta}),
\end{aligned}$$

where $\widehat{G} = n^{-1}\sum_{i=1}^{n} G_i(\widehat{\theta})$, $G_i(\theta) = \Delta(X_i, \theta)$, $\widehat{F}$ and $\widehat{S}$ can be estimated by the numerical derivative methods similar to the one-step GMM case. H can be also estimated by the numerical methods based on the criterion functions $J_n(\theta, \phi) = g_n(\theta)'W_n(\phi)g_n(\theta)$, similar to Hansen and Lee (2021).

Finally, the asymptotic standard errors of the iterated GMM estimator is obtained by taking the diagonal elements of $\sqrt{\widehat{V}_{mr}(\widehat{\theta})/n}$.

References

Ai, C., and X, Chen. (2007). Estimation of possibly misspecified semiparametric conditional moment restriction models with different conditioning variables. *Journal of Econometrics*, 141(1), 5-43.

Andrews, D.W.K. (1994). Empirical Process Methods in Econometrics. *Handbook of Econometrics*, 4, Elsevier North Holland, 2247-2294

Andrews I., Gentzkow M., and Shapiro J.M. (2017). Measuring the Sensitivity of Parameter Estimates to Estimation Moments. *Quarterly Journal of Economics*, 132, 1553–1592.

Andrews I., Barnhard, H., Carlson, J. (2024). True and Pseudo-True Parameters, working paper.

Armstrong, T and Kolesár, M (2021). Sensitivity Analysis using Approximate Moment Condition Models. *Quantitative Economics*, 12, 77–108

Bonhomme, S. and M. Weidner (2022). Minimizing Sensitivity to Model Misspecification. *Quantitative Economics*, 13(3), 907–954

[12] We use the numerical derivatives for $\widehat{G}, \widehat{F}$ in our simulation and empirical results.

Buchinsky, M. (1995). Estimating the Asymptotic Covariance Matrix for Quantile Regression Models: A Monte Carlo Study. *Journal of Econometrics*, 68, 303-338.

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.

Chen, X. and Z. Liao (2015). Sieve semiparametric two-step GMM under weak dependence. *Journal of Econometrics*, 189 (1), 163-186.

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.

Chen, X., and D. Pouzo (2009). Efficient Estimation of Semiparametric Conditional Moment Models With Possibly Nonsmooth Residuals. *Journal of Econometrics*, 152, 46-60.

Chen, X., and D. Pouzo (2012). Estimation of Nonparametric Conditional Moment Models With Possibly Nonsmooth Generalized Residuals. *Econometrica*, 80(1), 277-321.

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

Chernozhukov, V., and C. Hansen (2008). Instrumental variable quantile regression: a robust inference approach. J. Econometrics 142, 379-398.

Chernozhukov, V., Hansen, C., and Jansson, M. (2009). Finite sample inference for quantile regression models. *Journal of Econometrics*, 152(2), 93-103.

Chernozhukov, V., Hansen, C., Kong L., Wang, W. (2023). Plausible GMM via Avenue Bayes, working paper.

Chernozhukov, V., C. Hansen, and K. Wüthrich. (2018). Instrumental Variable Quantile Regression. in: Koenker, R., Chernozhukov, V., He, X., Peng, L. (Eds.), Handbook of Quantile Regression, Chapman & Hall/CRC.

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

Cho, J. S., and H. White (2018). Directionally differentiable econometric models. *Econometric Theory*, 34(5), 1101-1131.

Conley T. G., C. B. Hansen, and P. E. Rossi (2012). Plausibly Exogenous. *Review of Economics and Statistics*, 94, 260–272.

Giacomini, R., D. N. Politis, H. White (2013). A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. *Econometric Theory*, 29(3), 567-589.

Giné, E., and Zinn, J. (1990). Bootstrapping general empirical measures. *The Annals of Probability*, 851-869.

Gutenbrunner, C., and J. Jurečková (1992). Regression rank scores and regression quantiles. *Annals of Statistics*, 20, 305-330.

Hahn, J. (1995). Bootstrapping Quantile Regression Estimators. *Econometric Theory*, 11, 105-121.

Hahn, J. (1996). A Note on Bootstrapping Generalized Method of Moments Estimators. *Econometric Theory*, 12, 187-191.

Hahn, J. and Z. Liao (2021). Bootstrap Standard Error Estimates and Inference. *Econometrica*, 89(4), 1963-1977.

Hall, A. R., and A. Inoue (2003). The large sample behavior of the generalized method of moments estimator in misspecified models. *Journal of Econometrics*, 114(2), 361-394.

Hall, P., and S. J. Sheather (1988). On the Distribution of a Studentized Quantile, *Journal of the Royal Statistical Society: Series B*, 50(3), 381-391.

Hansen, B., and S. Lee. (2021). Inference for Iterated GMM under Misspecification. *Econometrica*, 89(3), 1419-1447.

He, X. (1997). Quantile curves without crossing. *The American Statistician*, 51, 186-192.

Hong, H., and J. Li (2023). Rate-Adaptive Bootstrap for Possibly Misspecified GMM. *Econometric Theory*, 1-51.

Hong, H., A. Mahajan, and D. Nekipelov (2015). Extremum estimation and numerical derivatives. *Journal of Econometrics*, 188 (1), 250-263.

Honoré, B. E., and L. Hu (2004). Estimation of Cross Sectional and Panel Data Censored Regression Models with Endogeneity. *Journal of Econometrics*, 122 (2), 293-316.

Huber, P.J. (1967). The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions. *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley: University of California Press, Vol. 1, 221-223.

Kaido, H., and K. Wüthrich (2021). Decentralization Estimators for Instrumental Variable Quantile Regression Models. *Quantitative Economics*, 12(2), 443-475.

Kaplan, D.M. (2022). Smoothed instrumental variables quantile regression. *Stata Journal* 22(2), 379-403.

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

Koenker, R (1994). Confidence intervals for regression quantiles. *Asymptotic statistics*, 349-359. Springer.

Koenker, R., and G. Bassett (1982). Robust tests for heteroscedasticity based on regression quantiles. *Econometrica*, 50, 43-61.

Koenker, R., and Z. Xiao (2002). Inference on the Quantile Regression Process. *Econometrica*, 70, 1583-1612.

Koenker, R. and Q. Zhao (1994). L-estimation for linear heteroscedastic models. *Journal of Nonparametric Statistics*, 3, 223-235.

Lee, S. (2014). Asymptotic Refinements of a Misspecification-Robust Bootstrap for generalized method of moments (GMM) Estimators. *Journal of Econometrics*, 178(3), 398-413

Machado, J.A.F., and J.M.C Santos Silva (2019). Quantiles via moments. *Journal of Econometrics*, 213, 145–173.

Newey, W. K., and D. McFadden (1994) Large sample estimation and hypothesis testing. *Handbook of econometrics*, 4, Elsevier North Holland, 2111-2245.

Pakes, A., and D. Pollard (1989). Simulation and the Asymptotics of Optimization Estimators. *Econometrica*, 57(5), 1027-1057.

Pollard, D. (1985). New Ways to Prove Central Limit Theorems. *Econometric Theory*, 1, 295-314.

van der Vaart, A. W. and J. A. Wellner (1996). Weak Convergence and Empirical Processes. Springer, New York.

Vapnik, V.N., and A. Ya. Chervonenkis (1971). On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities. *Theory of Probability and Its Applications*, 26, 532-553.

Wüthrich, K (2020). A comparison of two quantile models with endogeneity. *Journal of Business and Economic Statistics*, 38 (2), 443-456

Zhao, Q. (2000). Restricted regression quantiles. *Journal of Multivariate Analysis*, 72(1), 78-99.

Zhu, Y. (2019). Learning non-smooth models: instrumental variable quantile regressions and related problems, working paper, https://arxiv.org/abs/1805.06855.