Journal of Economic Theory and Econometrics, Vol. 33, No. 4, Dec. 2022, 31–53

Nonparametric Estimation of a Triangular System of Equations for Quantile Regression*

Sungwon Lee†

Abstract We consider a class of nonparametric quantile regression (QR) models with endogenous regressors. Building upon the semiparametric QR model in Lee (2007), we develop a nonparametric framework for quantile regression in a triangular system of equations. We provide a set of conditions under which the parameters are nonparametrically identified. Then, we propose to use the penalized sieve minimum distance (PSMD) estimation approach of Chen and Pouzo (2012) to estimate the parameters. We establish the consistency and convergence rate of the PSMD estimator. Since the identification is based on a control function approach, the PSMD estimator does not suffer from an ill-posed inverse problem. A Monte-Carlo simulation study confirms that the PSMD estimator performs well in finite samples.

Keywords Quantile Regression, Endogeneity, Nonparametric Simultaneous Equations Model, Sieve Estimation

JEL Classification C13, C14, C31

*The author is grateful to the two anonymous reviewers for their thoughtful comments and suggestions.

†Deparment of Economics, Sogang University, E-mail: sungwonlee@sogang.ac.kr

Submitted September 2, 2022, Accepted November 5, 2022

1. INTRODUCTION

Since the seminal paper by Koenker and Bassett (1978), quantile regression (QR) has received a lot of attention from the literature on both theoretical and applied econometrics. QR provides a useful framework in which researchers are allowed to investigate heterogeneous effects of variables across the distribution of an outcome variable. As heterogeneous effects are one of the most important issues in empirical studies in economics, QR has also become a popular approach in empirical economics.

It is common that a variable of interest is endogenous in many observational studies, and one of popular approaches to dealing with endogeneity issues is to use instrumental variables (IVs). Endogeneity in the QR framework has been studied by numerous papers in the literature (e.g., Chernozhukov and Hansen, 2005, 2006; Chernozhukov *et al.*, 2007; Horowitz and Lee, 2007; Lee, 2007; Imbens and Newey, 2009; Chen and Pouzo, 2009, 2012; Chernozhukov *et al.*, 2015).

In this paper, we consider a nonparametric QR model with endogenous regressors. Building upon the semiparametric triangular model of Lee (2007), we develop a fully nonparametric triangular model similar to that of Newey *et al.* (1999). Then, we study the identification of model parameters in the nonparametric triangular model for QR. The identification conditions are similar to those in Newey *et al.* (1999), utilizing the additively separable structure of the model and the existence of instrumental variables. The identification strategy presented in this paper is a control function approach that relies on the structure of a triangular system of equations (cf. Newey *et al.*, 1999; Blundell and Powell, 2003, 2004; Lee, 2007; Imbens and Newey, 2009; Blundell *et al.*, 2013; Chernozhukov *et al.*, 2015). In this regard, the identification strategy of this paper differs from that of the previous studies that use nonparametric quantile IV (NPQIV) regression models, where models for endogenous regressors are absent, and identification is achieved by some conditional moment restriction on the IVs (e.g., Chernozhukov *et al.*, 2007; Chen and Pouzo, 2009, 2012).

We propose to use a sieve method to estimate the parameters in the model that are nonparametrically specified. We adopt the penalized sieve minimum distance (PSMD) approach for nonparametric models, which was developed by Chen and Pouzo (2012). This estimation approach consists of a one-step procedure using conditional moment restrictions, and it is very easy to implement in practice. Moreover, we do not require a certain type of identifying assumptions on reduced-form parameters by using the one-step procedure. This mitigates the issue about model misspecification. We establish the consistency and convergence rate of the PSMD estimator of the true parameters under a set of low-level conditions.

One important and practical advantage of the PSMD estimator in the setting of this paper is that it does not suffer from an ill-posed inverse problem. It is well known that in a model where nonparametric parameters are identified through a set of conditional moment restrictions and endogenous variables are arguments of the nonparametric objects, many semi-/non-parametric estimators are subject to an ill-posed inverse problem that may lead to a slower convergence rate (cf. Carrasco *et al.*, 2007; Horowitz, 2014). While the additively separable NPQIV model considered in Chen and Pouzo (2012) is subject to the ill-posed inverse problem, our sieve estimator does not suffer from such an ill-posed inverse problem by the virtue of the triangular system of equations, where we specify models for endogenous regressors.[1]

Our estimation strategy is a one-step procedure, which facilitates the analysis on the asymptotic properties of the estimators. When the identification of model parameters is based on a control function approach, it is natural to use a multiple-step estimation procedure (e.g., Newey *et al.*, 1999; Das *et al.*, 2001; Imbens and Newey, 2009; Chernozhukov *et al.*, 2015). However, it is important to account for the effects of estimators in a prior step on the final estimator when establishing the asymptotic theory, including convergence rates, for such estimators, and this issue is sometimes referred to as "generated regressors problems" (cf. Hahn and Ridder, 2013; Hahn *et al.*, 2018; Chen *et al.*, 2021). On the other hand, the PSMD estimation procedure used in this paper allows us to effectively circumvent such issues.[2]

We conduct a small Monte-Carlo simulation study to investigate whether the proposed estimator performs well in finite samples. Our simulation results

[1]Related to this issue, there are studies in the literature using some other regularization than sieve methods to resolve the ill-posed inverse problem. One of the most popular regularization methods is the Tikhonov-type regularization (Tikhonov (1963a,b)), and the Tikhonov regularization has been widely used for estimation of nonparametric IV or NPQIV models (e.g., Hall and Horowitz, 2005; Horowitz and Lee, 2007; Darolles *et al.*, 2011).

[2]Although we do not consider inference in this paper, our one-step PSMD estimator has several advantages over a conventional two-step estimator in terms of inference. First, we can effectively circumvent estimation of asymptotic variances for inference by employing the bootstrap. We believe that the inference results developed by Chen and Pouzo (2015) are applicable to our PSMD estimator, and one of their main results is the bootstrap validity. Second, the inference results of Chen and Pouzo (2015) allow us to consider inference for functionals that are not $\sqrt{n}$-estimable. To the best of our knowledge, these important and practically useful results have not been established for two-step sieve estimation procedures. We leave the inference for the PSMD estimator proposed in this paper as future work.

indicate that the PSMD estimator has negligible biases and small standard deviations. Moreover, we compare the finite-sample performance of the one-step PSMD estimator with that of a conventional two-step sieve estimator. The simuation results indicate that our one-step PSMD estimator has a lower variance than the conventional two-step sieve estimator, whereas the bias of the one-step estimator tends to be greater than the bias of the two-step estimator. The Monte-Carlo simulation study also confirms that our one-step estimator is robust to misspecification of the reduced-form equation.

The main contribution of this paper is twofold. First, we extend the semiparametric model for QR with endogenous regressors in Lee (2007) to a fully nonparametric framework for QR while incorporating endogenous regressors and provide a set of conditions under which the model parameters are nonparametrically identified. Second, we establish the consistency and convergence rate of nonparametric PSMD estimator in the framework. The PSMD estimator proposed in this paper has many desirable properties and it is tractable and easy to implement. Therefore, we believe that the nonparametric framework and QR estimator of this paper are widely applicable to various empirical studies.

The additively separable NPQIV model in Chen and Pouzo (2012) is probably the most closely related to the model in this paper. We highlight the main differences between the model in this paper and that in Chen and Pouzo (2012). First, we rely on a triangular system of equations where all endogenous variables are determined within the system, whereas the NPQIV model in Chen and Pouzo (2012) does not require a specification for endogenous regressors. These models are nonnested, as pointed out by Horowitz (2014, p.28); and therefore, the additively separable NPQIV model in Chen and Pouzo (2012) is not more general than ours, and our model is not more general than the model in Chen and Pouzo (2012). Second, the identification strategy of this paper is completely different from that of Chen and Pouzo (2012). While we use a control function approach to identification of the model parameters, Chen and Pouzo (2012) impose a high-level condition for identification (cf. Condition 6.2(ii) in Chen and Pouzo (2012)). This high-level condition in Chen and Pouzo (2012) is hard to verify in practice. Lastly, the NPQIV model in Chen and Pouzo (2012) needs to assume the degree of ill-posedness to derive the convergence rate in their Proposition 6.2 (cf. Condition 6.3(ii) in Chen and Pouzo (2012)). On the other hand, we do not require such a restriction on the data generating process (DGP). In all, the assumptions for identification and estimation of this paper are quite different from those of Chen and Pouzo (2012); and therefore, the properties of the PSMD estimator proposed in this paper are established in a different way from

Chen and Pouzo (2012).

The rest of this paper is organized as follows. In Section 2, we develop the model and consider identification of the model. Section 3 introduces the PSMD estimation for the triangular QR model. We establish the consistency and convergence rate of the PSMD estimator in Sections 4 and 5, respectively. We present the Monte-Carlo simulation results in Section 6. Then, Section 7 concludes and discusses future work. All mathematical proofs are presented in the appendix.

Notation For a generic random variable A, the support of A is denoted by $Supp(A)$. For two random variables A and B, and for any $\tau \in (0,1)$, $Q_{A|B}(\tau|b)$ indicates the τ-th conditional quantile of A on $B = b$, and $F_{A|B}(a|b)$ is the conditional distribution function of A given $B = b$. $\mathbb{E}[\cdot]$ is the expectation operator. For any positive real sequences $\{a_n\}$ and $\{b_n\}$, $a_n \lesssim b_n$ means that there exist a finite constant $C > 0$ and $N \in \mathbb{N}$ such that $a_n \leq Cb_n$ for all $n \geq N$. If $a_n \lesssim b_n$ and $b_n \lesssim a_n$, it is denoted by $a_n \asymp b_n$.

2. THE MODEL AND IDENTIFICATION

We consider the following nonparametric triangular model: for each $\tau \in (0,1)$,

$$\begin{aligned} Y &= g(X, Z_1; \tau) + U(\tau), \\ X &= h(Z) + V, \end{aligned} \tag{1}$$

where $X \in \mathbb{R}^{d_x}$, $Z_1 \in \mathbb{R}^{d_{z_1}}$, $Z \equiv (Z_1', Z_2')' \in \mathbb{R}^{d_{z_1}+d_{z_2}}$. $U(\tau)$ and V are unobserved error terms that are scalar, and Z_2 is a vector of excluded variables such that $Z_2 \in \mathbb{R}^{d_{z_2}}$ and $d_{z_2} \geq d_x$. We call the first equation in model (1) the outcome equation, and the second equation is referred to as the reduced-form equation.

To allow for endogeneity of X, we assume that $U(\tau)$ and V can be correlated. The functions g and h are the parameters of interest that are nonparametrically specified, and researchers can only observe $(Y, X', Z')'$ from the data.

The model in (1) is closely related to that of Newey *et al.* (1999) in the sense that both equations have an additively separable structure between observed and unobserved variables. It is also a natural nonparametric extension of the model considered in Lee (2007), with a minor change in the specification for the reduced-form equation. Specifically, Lee (2007) assumes that $g(X, Z_1; \tau) = X'\beta(\tau) + Z_1'\gamma(\tau)$ and $h(Z) = \mu + Z'\pi$, while allowing that the reduced-form equation is also a QR model.

Suppose that $h(\cdot)$ is identified from the reduced-form equation and that

$$Q_{U(\tau)|Z,V}(\tau|Z,V) = Q_{U(\tau)|V}(\tau|V). \tag{2}$$

Equation (2) is qualitatively the same as the model restriction (2) in Lee (2007). A similar restriction was also considered by Newey *et al.* (1999) in a framework for conditional mean regression. Equation (2) is implied by that Z is independent of $U(\tau)$ conditional on V. When X is endogenous and Z is independent of the error terms, the conditional τ-th quantile of $U(\tau)$ on V is not a constant function. Therefore, we assume that $Q_{U(\tau)|V}(\tau|V)$ is a non-trivial function of V, which is denoted by $r(V;\tau)$:

$$\begin{aligned} Q_{U(\tau)|X,Z}(\tau|X,Z) &= Q_{U(\tau)|X,V}(\tau|X,V) \\ &= r(V;\tau). \end{aligned} \tag{3}$$

As a result, we have the following model restriction:

$$Q_{Y|X,Z_1,V}(\tau|X,Z_1,V) = g(X,Z_1;\tau) + r(V;\tau), \tag{4}$$

where $V = X - h(Z)$.

We now consider identifying conditions needed for identification of g, r, and h.

Assumption 1. *There exists a known* $(\bar{x}', \bar{z}_1')' \in Supp(X,Z_1)$ *such that* $g(\bar{x},\bar{z}_1;\tau) = 0$.

Assumption 2. *(i)* $g(X,Z_1)$*,* $r(V)$*, and* $h(Z)$ *are differentiable; (ii) the boundary of* $Supp(Z,V)$ *has zero probability.*

Assumption 3. *The function* $h(\cdot)$ *is identified over the support of* Z*.*

Assumption 4. *(i) For each* $\tau \in (0,1)$*,* $Q_{U(\tau)|Z,V}(\tau|Z,V) = Q_{U(\tau)|V}(\tau|V)$ *almost surely; (ii)* $\Pr\left(rank\left(\frac{\partial h(Z)}{\partial Z_2}\right) = d_x\right) = 1$.

Assumption 1 is a normalization condition, which is standard in the literature on nonparametric identification (cf. Matzkin (2007)).

Assumption 2 imposes conditions on the smoothness of the parameters and the distribution of (Z,V). The latter condition was also considered by Newey *et al.* (1999).

Assumption 3 is a high-level, but mild condition. Since $h(\cdot)$ is a reduced-form parameter, it is nonparametrically identified under a standard condition for

Z and V. Specifically, one can consider the conditional mean independence (i.e., $E[V|Z] = 0$ almost surely) as in Newey *et al.* (1999), or the $\tilde{\tau}$-th conditional quantile independence (i.e., $Q_{V|Z}(\tilde{\tau}|Z) = 0$ almost surely) for some $\tilde{\tau} \in (0,1)$ as in Lee (2007). As will be shown in the Monte-Carlo simulation in Section 6, the estimation procedure proposed in this paper yields a consistent estimator of h, regardless of what identifying assumption for h is imposed.

Assumption 4 is crucial for identification. The first condition of Assumption 4 requires that the common regressor Z_1 and the vector of excluded variables, Z_2, are conditional quantile independent of the error term $U(\tau)$ given V. This is a quantile version of the conditional mean independence considered by Newey *et al.* (1999) and a nonparametric version of the conditional τ-th quantile independence considered by Lee (2007). The second condition of Assumption 4 is a nonparametric version of the rank condition in the linear simultaneous equations model. This condition implicitly imposes a restriction that $d_{z_2} \geq d_x$.

The next theorem shows that under Assumptions 1–4, one can nonparametrically identify g, r, and h.

Theorem 1. *Suppose that Assumptions 1–4 hold. Then, g, r, and h are nonparametrically identified over their supports.*

3. ESTIMATION

Let g_0, r_0, and h_0 be the true parameter values for g, r, and h, respectively. We assume that $\mathscr{G}$, $\mathscr{R}$, and $\mathscr{H}$ are the parameter spaces for g, r, and h, respectively. We denote the vector of parameters by α (i.e., $\alpha \equiv (g,r,h)'$), and the true parameter vector is denoted by α_0. The parameter space for α is denoted by $\mathscr{A}$, which is the Cartesian product of the parameter spaces for g, r, and h (i.e., $\mathscr{A} = \mathscr{G} \times \mathscr{R} \times \mathscr{H}$).

Let $\{W_i \equiv (Y_i, X_i', Z_i')' : i = 1,2,...,n\}$ be the data. For each $\tau \in (0,1)$, consider the following conditional moment restriction:

$$\begin{aligned} m(X,Z;\alpha) &\equiv \mathbb{E}\left[\rho_\tau(W;\alpha)|X,Z\right] \\ &= \mathbb{E}\left[\mathbf{1}\left(Y \leq g(X,Z_1;\tau) + r(X - h(Z);\tau)\right) - \tau|X,Z\right], \end{aligned} \tag{5}$$

where $\mathbf{1}(\cdot)$ is an indicator function. Under the identification conditions, we have that $m(X,Z;\alpha) = 0$ almost surely in X and Z if and only if $\alpha = \alpha_0$.[3]

We consider the sieve minimum distance (SMD) estimation approach proposed by Chen and Pouzo (2012) that allows for a penalty function, which we

[3]Obviously, m also depends on τ, but we drop τ from the expression of m.

refer to as PSMD estimation approach. The PSMD sieve estimator of α_0, $\hat{\alpha}_n$, is defined as

$$\hat{\alpha}_n \equiv \arg \inf_{\alpha \in \mathscr{A}_n} \left\{ \frac{1}{n} \sum_i^n \hat{m}_n(X_i, Z_i; \alpha)' \left[\hat{\Sigma}_n(X_i, Z_i)\right]^{-1} \hat{m}_n(X_i, Z_i; \alpha) + \lambda_n \hat{P}_n(\alpha) \right\}, \tag{6}$$

where $\hat{m}_n(x,z;\alpha)$ is a consistent estimator of $m(x,z;\alpha)$, $\hat{\Sigma}_n(x,z)$ is a consistent estimator of positive definite matrix $\Sigma(x,z)$, and $\hat{P}_n(\alpha) \geq 0$ is a possibly random penalty function, λ_n is a positive real sequence such that $\lambda_n \downarrow 0$. $\mathscr{A}_n$ is a sieve space for the parameter space $\mathscr{A}$. Due to the flexibility and tractability of the PSMD approach, it has been widely used to estimate semi-nonparametric models for various empirical studies (e.g., Chen and Ludvigson (2009); Chen *et al.* (2013); Compiani (2022)).[4]

We consider a series estimator of $m(X,Z,;\alpha)$, $\hat{m}_n(X,Z;\alpha)$. That is, we define

$$\hat{m}_n(X,Z;\alpha) \equiv p^{J_n}(X,Z)'(P'P)^{-} \sum_{i=1}^{n} p^{J_n}(X_i,Z_i)\rho_\tau(W_i;\alpha), \tag{7}$$

where $\{p_j(\cdot,\cdot)\}_{j=1}^{\infty}$ is a sequence of some basis functions,

$$p^{J_n}(x,z) \equiv (p_1(x,z), p_2(x,z), ..., p_{J_n}(x,z))'$$

, and $P \equiv \left[p^{J_n}(X_1,Z_1), p^{J_n}(X_2,Z_2), ..., p^{J_n}(X_n,Z_n)\right]'$. When the argument of the basis functions is multi-dimensional, one can construct a sequence of basis functions by using tensor-product of univariate basis functions (cf. Chen (2007)).

It is worth pointing out that the nonparametric objects in the quantile model restriction in (5) do not depend on any endogenous regressors once we include $r(X - h(Z))$. Therefore, the PSMD estimator of α_0, $\hat{\alpha}_n$, does not suffer from an ill-posed inverse problem.

We introduce some class of functions. Let $f : \mathbb{D} \to \mathbb{R}$ where $\mathbb{D} \subseteq \mathbb{R}^{d_x}$ for some integer $d_x \geq 1$. Let $\omega = (\omega_1, ..., \omega_{d_x})$ be a d_x-tuple of nonnegative integers, and define the differential operator as $\nabla^\omega f \equiv \frac{\partial^{|\omega|}}{\partial x_1^{\omega_1} \partial x_2^{\omega_2} \cdots \partial x_{d_x}^{\omega_{d_x}}} f(x)$, where $x = (x_1, x_2, ..., x_{d_x}) \in \mathbb{D}$ and $|\omega| \equiv \sum_{i=1}^{d_x} \omega_i$. Let $[p]$ be the integer part of $p \in \mathbb{R}_+$, then a function $f : \mathscr{X} \to \mathbb{R}$ is called p-smooth if it is $[p]$ times continuously differentiable on $\mathscr{X}$ and for all ω such that $|\omega| = [p]$ and for some $\nu \in (0,1]$ and

[4] The sieve approach is also useful for estimating unknown distribution or density functions. For example, Song (2015) proposes a simulated Kolmogorov-Smirnov test based on sieve density estimators.

constant $c > 0$, $|\nabla^{\omega} f(x) - \nabla^{\omega} f(y)| \leq c \cdot ||x-y||_E^{\nu}$ for all $x, y \in \mathscr{X}$, where $||\cdot||_E$ is the Euclidean norm. Let $\mathscr{C}^{[p]}(\mathscr{X})$ denote the space of all $[p]$ times continuously differentiable real-valued functions on $\mathscr{X}$. A Hölder ball with smoothness p is defined as follows:

$$\Lambda_C^p(\mathscr{X}) \equiv \{f \in \mathscr{C}^{[p]}(\mathscr{X}) : \sup_{|\omega| \leq [p]} \sup_{x \in \mathscr{X}} |\nabla^{\omega} f(x)| \leq C,$$
$$\sup_{|\omega| = [p]} \sup_{x,y \in \mathscr{X}, x \neq y} \frac{|\nabla^{\omega} f(x) - \nabla^{\omega} f(y)|}{||x-y||_E^{\nu}} \leq C\},$$

where C is a positive finite constant.

We introduce norms on $\mathscr{A}$. For a generic function defined on the support of a random variable X with its distribution function F_X, $\mathscr{X}$, let $||f||_{\infty} \equiv \operatorname{ess\,sup}_{x \in \mathscr{X}} |f(x)|$ and $||f||_2^2 \equiv \int f(x)^2 dF_X(x)$ denote the supremum-norm (or sup-norm) and L_2-norm, respectively, while ess sup denotes the essential supremum. For any $\alpha, \tilde{\alpha} \in \mathscr{A}$, define $||\alpha - \tilde{\alpha}||_{\mathscr{A},\infty} \equiv ||g(\cdot,\cdot) - \tilde{g}(\cdot,\cdot)||_{\infty} + ||r(\cdot) - \tilde{r}(\cdot)||_{\infty} + ||h(\cdot) - \tilde{h}(\cdot)||_{\infty}$ and $||\alpha - \tilde{\alpha}||_{\mathscr{A},2}^2 \equiv ||g(\cdot,\cdot) - \tilde{g}(\cdot,\cdot)||_2^2 + ||r(\cdot) - \tilde{r}(\cdot)||_2^2 + ||h(\cdot) - \tilde{h}(\cdot)||_2^2$. For a (random) vector A, $||A||_E$ is the Euclidean norm of A.

4. CONSISTENCY

We now establish that the sieve estimator $\hat{\alpha}_n$ is consistent for the true parameter value α_0 in the sup-norm $||\cdot||_{\mathscr{A},\infty}$. For the simplicity of notation, we drop τ from the expressions for g and r.

We impose the following assumptions to show that the sieve estimator $\hat{\alpha}_n$ is consistent for α_0 with respect to $||\cdot||_{\mathscr{A},\infty}$.

Assumption 5. *(i) The data $\{W_i : i = 1,2,...n\}$ are i.i.d; (ii) The conditional distribution of Y on X and Z admits its conditional density function $f_{Y|X,Z}$ such that $f_{Y|X,Z}(g_0(X,Z_1;\tau) + r_0(X - h_0(Z);\tau)|X,Z) > 0$ almost surely, $f_{Y|X,Z}(y|x,z)$ is continuous in $(y,x',z')'$ and $\sup_y f_{Y|X,Z}(y|x,z) < \infty$ over $(x',z')' \in Supp(X,Z)$; (iii) $Supp(X,Z)$ is a compact subset of $\mathbb{R}^{d_x+d_z}$ with Lipschitz continuous boundary; (iv) the density function of (X,Z), $f_{XZ}(x,z)$, is bounded and bounded away from zero over $Supp(X,Z)$.*

Assumption 6. *(i) $g_0 \in \mathscr{G} \equiv \Lambda_{c_g}^{p_g}(Supp(X,Z_1))$, $r_0 \in \mathscr{R} \equiv \Lambda_{c_r}^{p_r}(Supp(X - h_0(Z))$, and $h_0 \in \mathscr{H} \equiv \Lambda_{c_h}^{p_h}(Supp(Z))$ with $p_g > d_x + d_{z_1}$, $p_r > d_x$, and $p_h > d_z$; (ii) all first-order partial derivatives of g_0, r_0, and h_0 are uniformly bounded.*

Assumption 7. *(i) $(p_j(\cdot))_{j=1}^{\infty}$ is a sequence of polynomial functions; (ii) the sieve spaces for $\mathscr{G}$, $\mathscr{R}$, and $\mathscr{H}$ are given by*

$$\mathscr{G}_n \equiv \left\{ g_n(x,z_1) = p^{k_{g,n}}(x,z_1)' \beta_{g,n} : ||g_n||_\infty \leq c_g \right\},$$
$$\mathscr{R}_n \equiv \left\{ r_n(v) = p^{k_{r,n}}(v)' \beta_{r,n} : ||r_n||_\infty \leq c_r \right\},$$
$$\mathscr{H}_n \equiv \left\{ h_n(z) = p^{k_{h,n}}(z)' \beta_{h,n} : ||h_n||_\infty \leq c_h \right\},$$

where $k_{g,n}$, $k_{r,n}$ and $k_{h,n}$ are some positive non-decreasing integer sequences such that $k_{g,n}, k_{r,n}, k_{h,n} \to \infty$, $\max(k_{g,n}, k_{r,n}, k_{h,n}) = o(n)$;
(iii) let $\mathbb{Q}_{g,n} \equiv \mathbb{E}\left[p^{k_{g,n}}(X,Z_1) \cdot p^{k_{g,n}}(X,Z_1)'\right]$, $\mathbb{Q}_{r,n} \equiv \mathbb{E}\left[p^{k_{r,n}}(V) p^{k_{r,n}}(V)'\right]$, and $\mathbb{Q}_{h,n} \equiv \mathbb{E}\left[p^{k_{h,n}}(Z) p^{k_{h,n}}(Z)'\right]$, then the eigenvalues of $\mathbb{Q}_{g,n}$, $\mathbb{Q}_{r,n}$ and $\mathbb{Q}_{h.n}$ are bounded above and away from zero uniformly over all n.

Assumption 8. *(i) $\Pr\left(\Sigma(X,Z) = \hat{\Sigma}_n(X,Z) = 1\right) = 1$ for all n; (ii) $\lambda_n = 0$ for all n.*

Assumption 9. *$F_{Y|X,Z}(g_0(\cdot,\cdot;\tau) + r_0(\cdot - h_0(\cdot;\tau)|X = \cdot, Z = \cdot) \in \Lambda_{c_m}^{p_m}(Supp(X,Z))$ with $p_m > 1/2$.*

Assumption 10. *(i) $(p_j(\cdot))_{j=1}^{\infty}$ is a sequence of polynomial functions;*
(ii) $\max_{j \leq J_n} \mathbb{E}\left[||p_j(X,Z)||_E^2\right] < C < \infty$ for some constant C; (iii) the smallest eigenvalue of $\mathbb{E}\left[p^{J_n}(X,Z) p^{J_n}(X,Z)'\right]$ is bounded away from zero for all J_n; (iv) $J_n \to \infty$ as $n \to \infty$ and $J_n^3 = o(n)$.

Condition (i) in Assumption 5 imposes conditions on the DGP. Condition (ii) in Assumption 5 is standard in the literature on QR. Condition (iii) and (iv) in Assumption 5 impose conditions on the DGP of X and Z.

Assumption 6 specifies the parameter spaces for the structural functions g, r, and h. The parameter spaces are a Hölder ball with some degree of smoothness and boundedness.

Assumption 7 defines sieve spaces for $\mathscr{G}$, $\mathscr{R}$, and $\mathscr{H}$. We use polynomial sieve spaces to approximate the parameter functions. The choice of sieve spaces depends on the parameter spaces and support conditions. When a function with a compact support belongs to a Hölder ball, it is well known that finite-dimensional linear sieve spaces, such as polynomial, trigonometric, or spline sieve spaces, can well approximate functions in the Hölder ball.[5] Condition

[5]For more details on the choice of sieve spaces, one may refer to Chen (2007).

(ii) imposes some rate condition on $k_{g,n}$, $k_{r,n}$, and $k_{h,n}$. It is obvious that $\mathscr{A}_n \subseteq \mathscr{A}_{n+1}$ for all $n \geq 1$ since $k_{g,n}, k_{r,n}, k_{h,n} \to \infty$. Conditions (i) and (ii) in Assumption 7, together with Assumption 6, ensure that there exist sequence of functions $\{\pi_n g_0\}_n$, $\{\pi_n r_0\}$, and $\{\pi_n h_0\}$ such that $(\pi_n g_0, \pi_n r_0, \pi_n h_0) \in \mathscr{A}_n$ and $||g_0 - \pi_n g_0||_\infty = O\left(k_{g,n}^{-p_g/(d_x+d_{z_1})}\right)$, $||r_0 - \pi_n r_0||_\infty = O\left(k_{r,n}^{-p_r/d_x}\right)$, and $||h_0 - \pi_n h_0||_\infty = O\left(k_{h,n}^{-p_h/d_z}\right)$ (cf. Newey (1997)). In addition, we have $\overline{\cup_n \mathscr{A}_n} = \mathscr{A}$, where, for a set A, $\overline{A}$ is the closure of A, under Assumptions 6 and 7. Condition (iii) of Assumption 7 is standard in the literature on sieve or series estimation (e.g., Newey (1997); Belloni *et al.* (2015, 2019)).

Assumption 8 defines the weighting matrix and the parameter for the degree of penalty, λ_n. Since Chen and Pouzo (2012) show that the use of a slowly growing finite-dimensional sieve space (i.e., $\max(k_{g,n}, k_{r,n}, k_{h,n}) = o(n)$) is easy to compute estimators and that estimators using a slowly growing finite-dimensional sieves with a low penalty (or without penalization) perform well in finite samples, we do not use penalization in this paper.[6] The PSMD estimation procedure without penalization can be carried out by setting λ_n to zero for all n, as in condition (ii) of Assumption 8.

Assumption 9 specifies the space of functions where the conditional moment function m belongs to. Since $Supp(X,Z)$ is compact by Assumption 5, we use polynomial sieve spaces to construct the series estimator of m, as in (7).

Assumption 10 is a set of sufficient conditions under which the sieve estimator $\hat{m}_n$ behaves well in the sense that some conditions for consistency of the sieve estimator $\hat{\alpha}_n$ are satisfied (cf. Assumption 3.3 in Chen and Pouzo (2012)). In particular, condition (iii) of Assumption 10 restricts the rate of J_n. If $J_n \asymp k_n$, where $k_n = \max(k_{g,n}, k_{r,n}, k_{h,n})$, with $J_n \geq C \cdot k_n$ for some constant $C > 0$, condition (iii) requires that $k_n^3/n = o(1)$.

The following theorem demonstrates that the sieve estimator $\hat{\alpha}_n$ is consistent for α_0 with respect to $||\cdot||_{\mathscr{A},\infty}$.

Theorem 2. *Suppose that Assumptions 1–4 hold. If Assumptions 5 – 10 also hold, then we have*

$$||\hat{\alpha}_n - \alpha_0||_{\mathscr{A},\infty} = o_p(1).$$

[6] One can refer to Chen and Pouzo (2012) for the precise definition of slowly growing sieve spaces and comparison between slowly growing and large sieves.

5. CONVERGENCE RATES

Given that the sieve estimator $\hat{\alpha}_n$ is consistent for α_0 with respect to $||\cdot||_{\mathscr{A},\infty}$, we consider a shrinking $||\cdot||_{\mathscr{A},\infty}$ neighborhood around α_0. For given small $\varepsilon > 0$ and large $M > 0$, we define

$$\mathscr{A}_{os} \equiv \left\{\alpha \in \mathscr{A} : ||\alpha - \alpha_0||_{\mathscr{A},\infty} \leq \varepsilon, ||\alpha||_{\mathscr{A},\infty} \leq M\right\},$$
$$\mathscr{A}_{osn} \equiv \mathscr{A}_{os} \cap \mathscr{A}_n.$$

Define

$$\frac{dm(X,Z;\alpha_0)}{d\alpha}[\alpha - \alpha_0] \equiv \frac{d\mathbb{E}\left[\rho(W;(1-t)\alpha_0 + t\alpha|X,Z\right]}{dt}\Big|_{t=0}$$

as the pathwise derivative of m in the direction $[\alpha - \alpha_0]$ evaluated at α_0. Let $||\cdot||$ denote a pseudo metric on $\mathscr{A}_{os}$, where for any $\alpha_1, \alpha_2 \in \mathscr{A}_{os}$,

$$||\alpha_1 - \alpha_2|| \equiv \sqrt{\mathbb{E}\left[\left(\frac{dm(X,Z;\alpha_0)}{d\alpha}[\alpha_1 - \alpha_2]\right)'(\Sigma(X,Z))^{-1}\left(\frac{dm(X,Z;\alpha_0)}{d\alpha}[\alpha_1 - \alpha_2]\right)\right]}.$$

In this section, we establish the convergence rate of the sieve estimator $\hat{\alpha}_n$ with respect to $||\cdot||_2$. We impose the following additional condition.

Assumption 11. *(i) $\mathscr{A}_{os}$ and $\mathscr{A}_{osn}$ are convex; (ii) $\mathbb{E}\left[||m(X,Z;\alpha)||_E^2\right] \asymp ||\alpha - \alpha_0||^2$ for all $\alpha \in \mathscr{A}_{osn}$.*

Condition (i) in Assumption 11, together with Assumption 5-(ii), guarantees that the norm $||\cdot||$ is well defined over $\mathscr{A}_{os}$ and $\mathscr{A}_{osn}$. Condition (ii) in Assumption 11 is mild when we focus on the shrinking neighborhood of α_0, $\mathscr{A}_{osn}$ (cf. van der Geer (2000, section 12.3)).

The following theorem provides the L_2-convergence rate of the sieve estimator $\hat{\alpha}_n$:

Theorem 3. *Suppose that the conditions in Theorem 2 are satisfied. If Assumption 11 is additionally satisfied, then*

$$||\hat{\alpha}_n - \alpha_0||_{\mathscr{A},2} = O_p\left(\max\left\{||\alpha_0 - \pi_n\alpha_0||_{\mathscr{A},2}, \sqrt{\frac{J_n}{n}}, J_n^{-\frac{p_m}{d_x+d_z}}\right\}\right).$$

If $J_n \asymp k_n = \max\left(k_{g,n}, k_{r,n}, k_{h,n}\right)$, then the convergence rate in Theorem 3 is reduced to

$$||\hat{\alpha}_n - \alpha_0||_{\mathscr{A},2} = O_p\left(\max\left\{k_{g,n}^{-p_g/(d_x+d_{z_1})}, k_{r,n}^{-p_r/d_x}, k_{h,n}^{-p_h/d_z}\sqrt{\frac{k_n}{n}}, k_n^{-\frac{p_m}{d_x+d_z}}\right\}\right),$$

which is the convergence rate of the standard nonparametric estimator without endogeneity. This is because the parameters in (5) do not depend on endogenous regressors anymore, which allows us to circumvent the ill-posed inverse problem.

6. MONTE CARLO SIMULATION

We conduct a small Monte Carlo simulation study to investigate the performance of the PSMD estimator $\hat{\alpha}_n$ in finite samples. To this end, we consider the following DGP:

$$Y = F_B(X; a_g(\tau), b_g(\tau)) - F_B(0.5; a_g(\tau), b_g(\tau) + \Phi(\varepsilon),$$
$$X = F_B(Z/2 + 0.5; a_h, b_h) + \Phi(\eta) - \tilde{\tau}_0,$$

where $(\varepsilon, \eta)' \sim BVN\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & 0.3\\0.3 & 1\end{bmatrix}\right)$ with BVN standing for bivariate normal distributions, $F_B(\cdot; a, b)$ is the beta distribution function with parameters a and b, $\Phi(\cdot)$ is the standard normal distribution function, and $\tilde{\tau}_0 \in \{0.25, 0.5, 0.75\}$. Under this DGP, $U = \Phi(\varepsilon)$, $V = \Phi(\eta) - \tilde{\tau}_0$, and $Q_{V|Z}(\tilde{\tau}_0|Z) = 0$ almost surely. We allow for $a_g(\tau)$ and $b_g(\tau)$ to vary across the quantile level τ. Specifically, we set $a_g(\tau) = 4 + \Phi^{-1}(\tau)$, $b_g(\tau) = 4 - \Phi^{-1}(\tau)$, and $a_h = b_h = 2$. Note that the normalization value $\bar{x}$ in Assumption 1 is 0.5 in the simulation.

The sample size is set to be 500. We use polynomial sieve spaces to approximate g, r, h, and m with $k_n \propto n^{1/7}$ and $J_n = (k_{g,n} + 1) \cdot (k_{h,n} + 1)$, where $k_n = \max\left(k_{g,n}, k_{r,n}, k_{h,n}\right)$. In our simulation, we set $k_n = 4$ by choosing a proper constant C_0 such that $k_n = C_0 \cdot n^{1/7} = 4$. For measures of the performance of the PSMD estimator $\hat{\alpha}_n$, we focus on the integrated bias (*IBIAS*) and integrated variance (*IVAR*). Specifically, we define $(IBIAS)^2 \equiv \int_{\mathscr{X}} \left(\frac{1}{m}\sum_{j=1}^{m} \hat{g}_n^{(j)}(x;\tau) - g_0(x;\tau)\right)^2 dF_X(x)$, where m is the number of simulations and $\hat{g}_n^{(j)}$ is the sieve estimator of g_0 from the j-th simulation. The *IVAR* is defined as $IVAR \equiv \frac{1}{m}\sum_{j=1}^{m}\left[\int_{\mathscr{X}}\left(\hat{g}_n^{(j)}(x;\tau) - \frac{1}{m}\sum_{j=1}^{m}\hat{g}_n^{(j)}(x;\tau)\right)^2 dF_X\right]$. The *IBIAS* and *IVAR* are calculated by numerical integration over the unit interval. The integrated mean squared error (*IMSE*) is defined as the sum of $(IBIAS)^2$ and *IVAR*. We consider $\tau \in \{0.25, 0.5, 0.75\}$, and all results are obtained from 1000 simulations.

In the simulation, we compare the performance of the one-step estimator proposed in this paper to that of a conventional two-step estimator. Specifically, a two-step estimator of $g_0(\cdot;\tau)$ is obtained as follows: In the first step, we run a nonparametric QR of X on Z using the model restriction $Q_{V|Z}(\tilde{\tau}|Z) = 0$ almost surely. Then, we generate $\hat{V} \equiv X - \hat{h}_n(Z)$ and run a nonparametric QR of Y on X and $\hat{V}$ in the second stage. The quantile level for the first stage regression, $\tilde{\tau}$, is chosen by the researcher, and we allow for misspecification of the reduced-form equation by assuming that $Q_{V|Z}(\tilde{\tau}|Z) = 0$ almost surely for some $\tilde{\tau} \neq \tilde{\tau}_0$. We consider both cases where the reduced-form equation is correctly specified (i.e., $\tilde{\tau} = \tilde{\tau}_0$) and it is misspecified (i.e., $\tilde{\tau} \neq \tilde{\tau}_0$) to investigate the performance of the PSMD estimator of $g_0(\cdot;\tau)$.

Table 1 reports the *IBIAS* and *IVAR* of the one-step and two-step sieve estimators of $g_0(\cdot;\tau)$ under correct specification for the reduced-form equation. We find that both sieve estimators of $g_0(\cdot;\tau)$ perform well in finite samples. Moreover, the simulation results indicate that the one-step estimator has a lower *IVAR* than the two-step estimator, whereas *IBIAS* of the one-step estimator tends to be larger than that of the

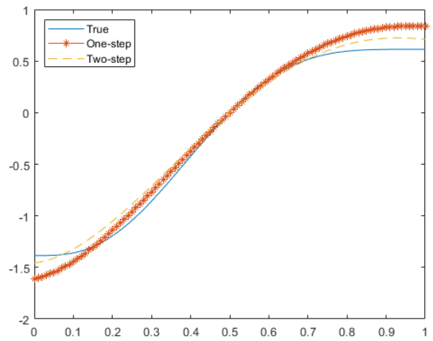
Table 1: Simulation Results: One-step and Two-step Estimators under Correct Specification for the Reduced-form Equation $(n = 500, k_{g,n} = 4)$

	$Q_{V\|Z}(0.25\|Z) = 0$					
	$\hat{g}_n(\cdot;0.25)$		$\hat{g}_n(\cdot;0.5)$		$\hat{g}_n(\cdot;0.75)$	
	One-step	Two-step	One-step	Two-step	One-step	Two-step
$(IBIAS)^2$	0.0158	0.0060	0.0068	0.0028	0.0171	0.0188
$IVAR$	0.0016	0.0025	0.0015	0.0039	0.0021	0.0048
$IMSE$	0.0174	0.0085	0.0083	0.0068	0.0191	0.0237
	$Q_{V\|Z}(0.5\|Z) = 0$					
	$\hat{g}_n(\cdot;0.25)$		$\hat{g}_n(\cdot;0.5)$		$\hat{g}_n(\cdot;0.75)$	
	One-step	Two-step	One-step	Two-step	One-step	Two-step
$(IBIAS)^2$	0.0141	0.0073	0.0127	0.0060	0.0110	0.0076
$IVAR$	0.0008	0.0024	0.0008	0.0030	0.0009	0.0024
$IMSE$	0.0149	0.0097	0.0135	0.0090	0.0119	0.0100
	$Q_{V\|Z}(0.75\|Z) = 0$					
	$\hat{g}_n(\cdot;0.25)$		$\hat{g}_n(\cdot;0.5)$		$\hat{g}_n(\cdot;0.75)$	
	One-step	Two-step	One-step	Two-step	One-step	Two-step
$(IBIAS)^2$	0.0178	0.0190	0.0054	0.0028	0.0094	0.0061
$IVAR$	0.0023	0.0070	0.0016	0.0047	0.0015	0.0026
$IMSE$	0.0201	0.0259	0.0070	0.0075	0.0109	0.0087

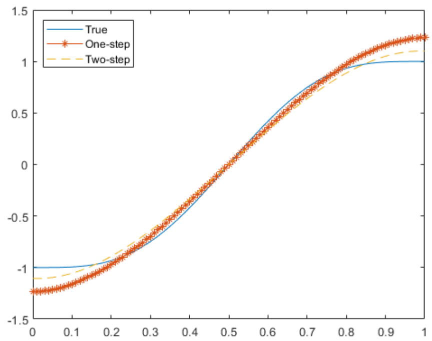
two-step estimator. In terms of the $IMSE$, we can find that the performance depends on the quantile level of interest for the outcome equation. Specifically, when $\tau = 0.25$ or $\tau = 0.5$, the two-step estimator has a slightly lower $IMSE$ than the one-step estimator. However, when $\tau = 0.75$, then the one-step estimator tends to achieve a lower $IMSE$ than the two-step estimator. Figure 1 depicts the true structural function $g_0(\cdot;\tau)$ and the means of the one-step and two-step sieve estimators of $g_0(\cdot;\tau)$ over the simulations (i.e., $\frac{1}{m}\sum_{j=1}^{m}\hat{g}_n^{(j)}(x;\tau)$). The solid lines represent the true structural function, the marked lines are the mean of the one-step sieve estimators, and the dashed lines are the mean of the two-step sieve estimators from 1000 simulations. We can see that in terms of the bias, the one-step estimator performs poorer at the tails of the support of X than the two-step estimator, while the one-step estimator outperforms the two-step estimator in the interior of the support of X.

We now turn our attention to the case where the reduced-form equation is misspecified. Table 2 reports the estimation results when the reduced-form equation is misspecified. We do not find discernible changes in the finite-sample performance of the one-step and two-step estimators even when the reduced-form equation is misspecified. The results in Table 2 are qualitatively the same to those in Table 1. Although we do not report

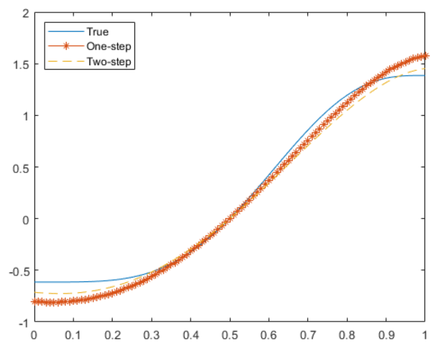
Figure 1: Simulation Results $(n = 500,\ k_{g,n} = 4)$



$\tau = 0.25$



$\tau = 0.5$



$\tau = 0.75$

Note: The solid lines represent the true structural function, and the dashed lines are the mean of sieve estimators from 1000 simulations.

Table 2: Simulation Results: One-step and Two-step Estimators under Misspecification for the Reduced-form Equation $(n = 500, k_{g,n} = 4)$

	$\hat{g}_n(\cdot;0.5)$							
	$\tilde{\tau}_0 = 0.25,\ \tilde{\tau} = 0.5$		$\tilde{\tau}_0 = 0.25,\ \tilde{\tau} = 0.75$		$\tilde{\tau}_0 = 0.75,\ \tilde{\tau} = 0.25$		$\tilde{\tau}_0 = 0.75,\ \tilde{\tau} = 0.5$	
	One-step	Two-step	One-step	Two-step	One-step	Two-step	One-step	Two-step
$(IBIAS)^2$	0.0067	0.0028	0.0068	0.0028	0.0053	0.0028	0.0053	0.0028
$IVAR$	0.0016	0.0045	0.0016	0.0041	0.0016	0.0051	0.0015	0.0049
$IMSE$	0.0083	0.0073	0.0083	0.0070	0.0069	0.0079	0.0068	0.0077

the estimation results for the first-stage parameter (i.e., h_0) in this paper, the simulation results show that the one-step procedure yields a better estimator of h_0 than the two-step procedure in terms of both bias and variance. These results suggest that the one-step procedure is robust to misspecification of the reduced-form equation in the sense that regardless of whether the reduced-form equation is correctly specified or not, we can obtain a consistent estimator of h_0. In all, our one-step PSMD estimators perform well in finite samples in the sense that they have a reasonable bias and a small variance.

7. CONCLUSIONS

In this paper, we consider a nonparametric triangular system of equations for QR. Our model is a nonparametric extension of the semiparametric model of Lee (2007) and shares some common features with the model of Newey *et al.* (1999). We provide a set of conditions under which the model parameters are nonparametrically identified. Then, we propose to use the PSMD estimation approach developed by Chen and Pouzo (2012) to estimate the parameters, and establish the consistency with respect to the sup-norm and the L_2-convergence rate. The Monte-Carlo simulation study confirms that the PSMD estimator performs well in finite samples. When comparing the one-step PSMD estimator with a conventional two-step sieve estimator, the one-step estimator has a lower variance than the two-step estimator, whereas the bias of the one-step estimator tends to be higher than the bias of the two-step estimator.

While we show that the PSMD estimator in this paper is consistent and establish its L_2-convergence rate, it is needed to develop some distributional theory for statistical inference. In our work in progress, we adopt the sieve inference method developed by Chen and Pouzo (2015) to derive the theory for inference on functionals of the nonparametric functions in the model of this paper.

A. MATHEMATICAL PROOFS

In this section, we provide mathematical proofs of the main results. We introduce notation that will be used in the proofs. Let $(\mathscr{F}, ||\cdot||_{\mathscr{F}})$ be a metric space of real valued function $f : \mathscr{X} \to \mathbb{R}$. The covering number $N(\varepsilon, \mathscr{F}, ||\cdot||_{\mathscr{F}})$ is the minimum number of $||\cdot||_{\mathscr{F}}$ ε-balls that cover $\mathscr{F}$. The entropy is the logarithm of the covering number. An ε-bracket in $(\mathscr{F}, ||\cdot||_{\mathscr{F}})$ is a pair of functions $l, u \in \mathscr{F}$ such that $||l||_{\mathscr{F}}, ||u||_{\mathscr{F}} < \infty$ and $||u - l||_{\mathscr{F}} \leq \varepsilon$. The covering number with bracketing $N_{[]}(\varepsilon, \mathscr{F}, ||\cdot||_{\mathscr{F}})$ is the minimum number of $||\cdot||_{\mathscr{F}}$ ε-brackets that cover $\mathscr{F}$. The entropy with bracketing is the logarithm of the covering number with bracketing. The bracketing integral is defined as $\int_0^\delta \sqrt{\log N_{[]}(\varepsilon, \mathscr{F}, ||\cdot||_{\mathscr{F}})} d\varepsilon$. Let C denote a generic positive and finite constant. It can be different across where it appears. Some empirical processes may not be measurable, and thus the expectation operator cannot be applied to those processes. In such a case, one can replace the expectation operator with the outer expectation operator. We use the notation $\mathbb{E}[\cdot]$ mainly to indicate the expectation operator, but it may also stand for the outer expectation if its argument is not measurable.

A.1. PROOF OF THEOREM 1

Proof. Note that $h(\cdot)$ is identified over $Supp(Z)$ under Assumption 3. Under Assumptions 2 and 4, the structural functions g and r are nonparametrically identified up to additive constants by Theorem 2.3 in Newey *et al.* (1999). Finally, the normalization condition in Assumption 1 identifies the structural functions. □

A.2. PROOF OF THEOREM 2

We first recall that

$$m(X, Z; \alpha) = F_{Y|X,Z}(g(X, Z_1) + r(X - h(Z))|X, Z) - \tau. \tag{8}$$

Lemma 1. *Suppose that Assumptions 1–4 hold. If Assumptions 5–8 hold, then Assumptions 3.1 and 3.2-(a) in Chen and Pouzo (2012) are satisfied.*

Proof. Condition (i) of Assumption 3.1 in Chen and Pouzo (2012) is satisfied by condition (i) in Assumption 8.

Condition (ii) of Assumption 3.1 in Chen and Pouzo (2012) is implied by the identification conditions (Assumptions 1–4).

Condition (iii) of Assumption 3.1 in Chen and Pouzo (2012) is guaranteed by Assumptions 6 and 7. Specifically, since we use polynomial sieve spaces for $\mathscr{G}$, $\mathscr{R}$, and $\mathscr{H}$, there exist $\{\pi_n g_0\}_n$, $\{\pi_n r_0\}$, and $\{\pi_n h_0\}$ such that

$$||\pi_n \alpha_0 - \alpha_0||_{\mathscr{A},\infty} = O\left(k_{g,n}^{-\frac{p_g}{d_x + d_{z_1}}}\right) + O\left(k_{r,n}^{-\frac{p_r}{d_x}}\right) + O\left(k_{h,n}^{-\frac{p_h}{d_z}}\right)$$

by Newey (1997). Moreover, $\max\left(k_{g,n}^{-\frac{p_g}{d_x+d_{z_1}}}, k_{r,n}^{-\frac{p_r}{d_x}}, k_{h,n}^{-\frac{p_h}{d_z}}\right) = o(1)$ by Assumption 7.

Assumption 5-(ii) implies that $F_{Y|X,Z}(\cdot|x,z)$ is continuous for all $(x',z')' \in Supp(X,Z)$, and therefore, we have $||\pi_n\alpha_0 - \alpha_0||_{\mathscr{A},\infty} = o(1)$ under Assumptions 6 and 7. Therefore, condition (iv) of Assumption 3.1 in Chen and Pouzo (2012) is met.

Assumption 3.2-(a) in Chen and Pouzo (2012) is satisfied under condition (ii) in Assumption 8. □

Lemma 2. *Suppose that Assumptions 1–4 hold. If Assumptions 5–10 hold, then Assumption 3.3 in Chen and Pouzo (2012) is satisfied with*

$$\bar{\delta}_{m,n}^2 = \eta_{0,n} = \max\left\{\frac{J_n}{n}, J_n^{-\frac{2p_m}{d_x+d_z}}\right\} = o(1).$$

Proof. We verify the conditions of Lemma C.2 in Chen and Pouzo (2012). Conditions (i) and (ii) of Assumption C.1 in Chen and Pouzo (2012) are imposed by Assumption 5. Conditions (iii) of Assumption C.1 in Chen and Pouzo (2012) is imposed by Assumption 10. Since $\sup_{(x,z)\in Supp(X,Z)} ||p^{J_n}(x,z)||_E^2 \asymp J_n^2$ when $\left(p^{J_n}\right)_n$ is a sequence of polynomial functions (Newey (1997)), condition (iv) of Assumption 10 implies condition (iv) of Assumption C.1 in Chen and Pouzo (2012). Lastly, condition (v) of Assumption C.1 in Chen and Pouzo (2012) is implied by Assumption 8.

We now consider Assumption C.2 in Chen and Pouzo (2012). Since

$$\sup_{\alpha\in\mathscr{A}_n} |\rho(W,\alpha)| \leq 2$$

, condition (i) of Assumption C.2 in Chen and Pouzo (2012) is satisfied. Condition (ii) of Assumption C.2 in Chen and Pouzo (2012) is guaranteed by Assumptions 9 and 10. Specifically, under these assumptions, there exist a sequence of functions $(\pi_n m)_n$ such that $\pi_n m(x,z) = p^{J_n}(x,z)'\beta_{m,n}^*$ for each n and $\mathbb{E}\left[\left(m(X,Z;\alpha) - p_m^{J_n}(X,Z)'\beta_{m,n}^*\right)^2\right] = O\left(J_n^{-\frac{2p_m}{d_x+d_z}}\right)$. Obviously, $J_n^{-\frac{2p_m}{d_x+d_z}} = o(1)$.

Let $\delta > 0$ such that $\delta = o(1)$. For each $1 \leq j \leq J_n$ and for any $\alpha \in \mathscr{A}_n$, it is straightforward to see that

$$\begin{aligned}
&\mathbb{E}[(p_j(X,Z))^2 \sup_{\tilde{\alpha}\in\mathscr{A}_n:||\alpha-\tilde{\alpha}||_\infty\leq\delta} |\rho(W;\alpha) - \rho(W;\tilde{\alpha})|^2] \\
\lesssim &\mathbb{E}[(p_j(X,Z))^2 \sup_{\tilde{\alpha}\in\mathscr{A}_n:||\alpha-\tilde{\alpha}||_\infty\leq\delta} \{\mathbf{1}(Y \leq g(X,Z_1) + r(X - h(Z))) \\
&- \mathbf{1}\left(Y \leq \tilde{g}(X,Z_1) + \tilde{r}(X - \tilde{h}(Z))\right)\}^2].
\end{aligned}$$

We note that if $||\alpha - \tilde{\alpha}||_{\mathscr{A},\infty} \leq \delta$, then,

$$\begin{aligned}\left|(g(X,Z_1)+r(X-h(Z))) - \left(\tilde{g}(X,Z_1)+\tilde{r}(X-\tilde{h}(Z))\right)\right| &\leq ||g-\tilde{g}||_\infty + ||r-\tilde{r}||_\infty \\ &\quad + C\cdot||h-\tilde{h}||_\infty \\ &\lesssim \delta\end{aligned}$$

by using the fact that r and $\tilde{r}$ are differentiable and their derivatives are uniformly bounded under Assumptions 6 and 7. Therefore, by the law of iterated expectations and the argument of Chen *et al.* (2003, p.1600), we have

$$\begin{aligned}&\mathbb{E}[(p_j(X,Z))^2 \sup_{\tilde{\alpha}\in\mathscr{A}_n:||\alpha-\tilde{\alpha}||_\infty\leq\delta} |\mathbf{1}(Y \leq g(X,Z_1)+r(X-h(Z))) \\ &\qquad - \mathbf{1}\left(Y \leq \tilde{g}(X,Z_1)+\tilde{r}(X-\tilde{h}(Z))\right)|] \\ &\lesssim \mathbb{E}[(p_j(X,Z))^2 \cdot (F_{Y|X,Z}(g(X,Z_1)+r(X-h(Z))+\delta|X,Z) \\ &\qquad - F_{Y|X,Z}(g(X,Z_1)+r(X-h(Z))-\delta|X,Z))] \\ &\lesssim \mathbb{E}[(p_j(X,Z))^2]\delta\end{aligned}$$

by Assumption 5. Since $\mathbb{E}\left[(p_j(X,Z))^2\right] < \infty$ for all $j = 1,2,...,J_n$, we have

$$\mathbb{E}\left[(p_j(X,Z))^2 \sup_{\tilde{\alpha}\in\mathscr{A}_n:||\alpha-\tilde{\alpha}||_\infty\leq\delta} ||\rho(W;\alpha)-\rho(W;\tilde{\alpha})||_E^2\right] \leq K^2\delta$$

for some $K > 0$. Therefore, condition (C.1) in Chen and Pouzo (2012) is satisfied with $\kappa = 1/2$.

We now need to calculate the entropy $\log N\left(w^{1/\kappa}, \mathscr{A}_n^{M_0}, ||\cdot||_{\mathscr{A},\infty}\right)$ for $\kappa = 1/2$, where, for some $M_0 > 0$, $\mathscr{A}_n^{M_0} \equiv \{\alpha \in \mathscr{A}_n : \lambda_n P(\alpha) \leq \lambda_n M_0\}$. Since $\lambda_n = 0$ for all n by Assumption 8, we have $\mathscr{A}_n^{M_0} = \mathscr{A}_n$ for all $M_0 > 0$. Moreover, we have $\mathscr{A}_n \subseteq \mathscr{A}_{n+1}$ for all $n \geq 1$, and $\overline{\cup_n \mathscr{A}_n} = \mathscr{A}$, and thus, it is enough to calculate $\log N\left(w^{1/\kappa}, \mathscr{A}, ||\cdot||_{\mathscr{A},\infty}\right)$. By Lemma 9.18 and Theorem 9.19 in Kosorok (2008), we have

$$\begin{aligned}&\log N\left(w^{1/\kappa}, \mathscr{A}, ||\cdot||_{\mathscr{A},\infty}\right) \\ \leq&\log N\left(w^{1/\kappa}, \mathscr{A}, ||\cdot||_{\mathscr{A},\infty}\right) \\ \lesssim&\log N\left(Cw^{1/\kappa}, \mathscr{G}, ||\cdot||_{\mathscr{A},\infty}\right) + \log N\left(Cw^{1/\kappa}, \mathscr{R}, ||\cdot||_{\mathscr{A},\infty}\right) + \log N\left(w^{1/\kappa}, \mathscr{H}, ||\cdot||_{\mathscr{A},\infty}\right) \\ \lesssim& w^{-\frac{2(d_x+d_{z_1})}{p_g}} + w^{-\frac{2d_x}{p_r}} + w^{-\frac{2d_z}{p_h}}.\end{aligned}$$

Therefore, it is straightforward to obtain that

$$\int_0^1 \sqrt{1+\log N\left(w^{1/\kappa}, \mathscr{A}, ||\cdot||_{\mathscr{A},\infty}\right)}dw \lesssim \int_0^1 \left(w^{-\frac{(d_x+d_{z_1})}{p_g}} + w^{-\frac{d_x}{p_r}} + w^{-\frac{d_z}{p_h}}\right)dw < \infty$$

under Assumption 6. By Remark C.1-(iii) in Chen and Pouzo (2012), Assumption C.2-(iv) in Chen and Pouzo (2012) is satisfied. Applying Lemma C.2-(iii) yields that Assumption 3.3 in Chen and Pouzo (2012) is satisfied with

$$\bar{\delta}_{m,n}^2 = \eta_{0,n} = \max\left\{\frac{J_n}{n}, J_n^{-\frac{2p_m}{d_x+d_z}}\right\} = o(1)$$

under Assumption 10. □

Proof of Theorem 2

Proof. Note that for any $\varepsilon > 0$,

$$\liminf_{n\to\infty}\left(\inf_{\alpha\in\mathscr{A}_n:||\alpha-\alpha_0||_{\mathscr{A},\infty}\geq\varepsilon}\mathbb{E}\left[|m(X,Z;\alpha)|^2\right]\right) = \inf_{\alpha\in\mathscr{A}:||\alpha-\alpha_0||_{\mathscr{A},\infty}\geq\varepsilon}\mathbb{E}\left[|m(X,Z;\alpha)|^2\right].$$

Since it follows from Theorems 1 and 2 in Freyberger and Masten (2019) that $(\mathscr{A}, ||\cdot||_{\mathscr{A},\infty})$ is compact under Assumption 6, we have

$$\inf_{\alpha\in\mathscr{A}:||\alpha-\alpha_0||_{\mathscr{A},\infty}\geq\varepsilon}\mathbb{E}\left[|m(X,Z;\alpha)|^2\right] > 0$$

under Assumptions 1–4. Moreover, $\mathbb{E}\left[|m(X,Z;\alpha)|^2\right]$ is continuous in α. In all, by Lemmas 1 and 2, all conditions of Theorem 3.1 in Chen and Pouzo (2012) are satisfied. Therefore, $||\hat{\alpha}_n - \alpha_0||_{\mathscr{A},\infty} = o_p(1)$. □

A.3. PROOF OF THEOREM 3

Proof. Theorem 3 is a direct consequence of Theorem 4.1 and Remark 4.1 in Chen and Pouzo (2012). Assumption 4.1-(i) in Chen and Pouzo (2012) is directly imposed by Assumption 11. Assumption 4.1-(ii) and (iii) in Chen and Pouzo (2012) are implied by Assumption 11-(ii). Since the sieve measure of ill-posedness $\sup_{\alpha\in\mathscr{A}_{os}:||\alpha-\alpha_0||\neq 0}\frac{||\alpha-\alpha_0||_{\mathscr{A},2}}{||\alpha-\alpha_0||}$ is equal to a constant under Assumption 5-(ii), applying Theorem 4.1 and Remark 4.1 in Chen and Pouzo (2012) establishes the convergence rate with respect to $||\cdot||_{\mathscr{A},2}$. □

REFERENCES

Belloni, Alexandre, V. Chernozhukov, D. Chetverikov, and K. Kato (2015). "Some new asymptotic theory for least squares series: Pointwise and uniform results," *Journal of Econometrics* 186, 345-366.

Belloni, Alexandre, V. Chernozhukov, D. Chetverikov, and I. Fernández-Val (2019). "Conditional quantile processes based on series or many regressors," *Journal of Econometrics* 213, 4-29.

Blundell, R., D. Kristensen, and R. L. Matzkin (2013). "Control functions and simultaneous equations methods," *American Economic Review* 103, 563-569.

Blundell, R. and J.L. Powell (2003). "Endogeneity in nonparametric and semiparametric regression models," in *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, eds., M. Dewatripont, L. P. Hansen, and S. J. Turnovsky, Cambridge University Press, 312-357.

Blundell, R. W. and J. L. Powell (2004). "Endogeneity in semiparametric binary response models," *Review of Economic Studies* 71, 655-679.

Carrasco, M., J-P. Florens, and E. Renault (2007)."Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization," *Handbook of Econometrics*, 6B, 5633-5751.

Chen, L., A. Galvao, and S. Song (2021). "Quantile regression with generated regressors," *Econometrics* 9.

Chen, X. (2007). "Large sample sieve estimation of semi-nonparametric models," *Handbook of Econometrics*, 6B, 5549-5632.

Chen, X., J. Favilukis, and S.C. Ludvigson (2013). "An estimation of economic models with recursive preferences," *Quantitative Economics* 4, 39-83.

Chen, X., O. Linton, and I. van Keilegom (2003). "Estimation of semiparametric models when the criterion function is not smooth," *Econometrica* 71, 1591-1608.

Chen, X. and S.C. Ludvigson (2009). "Land of addicts? An empirical investigation of habit-based asset pricing models," *Journal of Applied Econometrics* 24, 1057-1093.

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, 277-321.

Chen, X. and D. Pouzo (2015). “Sieve Wald and QLR inferences on semi/nonparametric conditional moment models,” *Econometrica* 83, 1013-1079.

Chernozhukov, V., I. Fernández-Val, and A. Kowalski (2015). “Quantile regression with censoring and endogeneity,” *Journal of Econometrics* 186, 201-221.

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

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

Chernozhukov, V., G. W. Imbens, and W. K. Newey (2007). “Instrumental variable estimation of nonseparable models,” *Journal of Econometrics* 139, 4-14.

Compiani, G. (2022). “Market counterfactuals and the specification of multiproduct demand: A nonparametric approach,” *Quantitative Economics* 13, 545-591.

Darolles, S., Y. Fan, J.-P. Florens, and E. Renault (2011). “Nonparametric instrumental regression,” *Econometrica* 79, 1541-1565.

Das, M., W. K. Newey, and F. Vella (2003). “Nonparametric estimation of sample selection models,” *Review of Economic Studies* 70, 33-58.

Freyberger, J. and M. A. Masten (2019). “A practical guide to compact infinite dimensional parameter spaces,” *Econometric Reviews* 38, 979-1006.

Hahn, J., Z. Liao, and G. Ridder (2018). “Nonparametric two-step sieve M estimation and inference,” *Econometric Theory* 34, 1281-1324.

Hahn, J. and G. Ridder (2013). “Asymptotic variance of semiparametric estimators with generated regressors,” *Econometrica* 81, 315-340.

Hall, P. and J. L. Horowitz (2005). “Nonparametric methods for inference in the presence of instrumental variables,” *Annals of Statistics* 33, 2904-2929.

Horowitz, J.L. (2014). “Ill-posed inverse problems in economics,” *Annual Review of Economics* 6, 21-51.

Horowitz, J. L. and S. Lee (2007). “Nonparametric instrumental variables estimation of a quantile regression model,” *Econometrica* 75, 1191-1208.

Imbens, G. W. and W. K. Newey (2009). “Identification and estimation of triangular simultaneous equations models without additivity,” *Econometrica* 77, 1481-1512.

Koenker, R. and G. Bassett (1978). “Regression quantiles,” *Econometrica* 46, 33-50.

Kosorok, M. (2008). *Introduction to empirical processes and semiparametric inference*, Springer.

Lee, S. (2007). "Endogeneity in quantile regression models: A control function approach," *Journal of Econometrics* 141, 1131-1158.

Matzkin, R. (2007). "Nonparametric identification," *Handbook of Econometrics*, 6B, 5307-5368.

Newey, W. K. (1997). "Convergence rates and asymptotic normality for series estimators," *Journal of Econometrics* 79, 147-168.

Newey, W. K., J. L. Powell, and F. Vella (1999). "Nonparametric estimation of triangular simultaneous equations models," *Econometrica* 67, 565-603.

Song, H. (2015). "A univariate sieve density estimation based on a simulated Kolmogorov-Smirnov test," *Journal of Economic Theory and Econometrics* 26, 26-43.

Tikhonov, A. (1963a). "Solution of incorrectly formulated problems and the regularization method," *Soviet Mathematics Doklady* 4, 1035-1038.

Tikhonov, A. (1963b). "Regularization of incorrectly posed problems," *Soviet Mathematics Doklady* 4, 1624-1627.

van der Geer, S. (2000). *Empirical processes in M-estimation*, Cambridge University Press.