

Singapore Management University

Institutional Knowledge at Singapore Management University

Research Collection School Of Economics

School of Economics

2-2015

Sieve Instrumental Variable Quantile Regression Estimation of Functional Coefficient Models

Liangjun SU

Singapore Management University, ljsu@smu.edu.sg

Tadao HOSHINA

Waseda University

Follow this and additional works at: https://ink.library.smu.edu.sg/soe_research



Part of the [Econometrics Commons](#)

Citation

SU, Liangjun and HOSHINA, Tadao. Sieve Instrumental Variable Quantile Regression Estimation of Functional Coefficient Models. (2015). 1-64.

Available at: https://ink.library.smu.edu.sg/soe_research/1716

This Working Paper is brought to you for free and open access by the School of Economics at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Economics by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email cherylds@smu.edu.sg.

Sieve Instrumental Variable Quantile Regression Estimation of Functional Coefficient Models

Liangjun Su and Tadao Hoshino

February 2015

Paper No. 01-2015

Sieve Instrumental Variable Quantile Regression Estimation of Functional Coefficient Models*

Liangjun Su^a and Tadao Hoshino^b

^a School of Economics, Singapore Management University

^b Global Education Center, Waseda University

February 9, 2015

Abstract

In this paper, we consider sieve instrumental variable quantile regression (IVQR) estimation of functional coefficient models where the coefficients of endogenous regressors are unknown functions of some exogenous covariates. We approximate the unknown functional coefficients by some basis functions and estimate them by the IVQR technique. We establish the uniform consistency and asymptotic normality of the estimators of the functional coefficients. Based on the sieve estimates, we propose a nonparametric specification test for the constancy of the functional coefficients, study its asymptotic properties under the null hypothesis, a sequence of local alternatives and global alternatives, and propose a wild-bootstrap procedure to obtain the bootstrap p -values. A set of Monte Carlo simulations are conducted to evaluate the finite sample behavior of both the estimator and test statistic. As an empirical illustration of our theoretical results, we present the estimation of quantile Engel curves.

JEL Classifications: C12, C13, C14, C21, C23, C26

Key Words: Endogeneity; Functional coefficient; Heterogeneity; Instrumental variable; Panel data; Sieve estimation; Specification test; Structural quantile function

1 Introduction

This paper focuses on sieve estimation of functional coefficient quantile regression (FCQR) models with endogeneity. As an effective way to model random coefficients and to allow the marginal effect of a regressor in a regression to be varying along with some other covariates, functional coefficient models

*The authors gratefully thank a co-editor, an associate editor, and three anonymous referees for their many constructive comments on the previous version of the paper. Su acknowledges support from the Singapore Ministry of Education for Academic Research Fund under grant number MOE2012-T2-2-021. Hoshino's work is supported by a JSPS Grant-in-Aid for Scientific Research PD-247943. Address correspondence to: Liangjun Su, School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903, Singapore; Tel: +65 6828 0386; E-mail: ljsu@smu.edu.sg. Tadao Hoshino, Global Education Center, Waseda University, 1-104 Totsuka, Shinjuku-ku, Tokyo, 169-8050, Japan; E-mail: thoshino@aoni.waseda.jp.

have been studied extensively in the last two decades; see Chen and Tsay (1993), Hastie and Tibshirani (1993), Fan and Zhang (1999), Cai et al. (2000), Fan and Huang (2005), and Su et al. (2009), among others. The coefficients in these models are modeled as unknown functions of the observed variables which can be estimated nonparametrically. But most of these works focus on conditional mean regression models with exogenous regressors. Recently, Cai et al. (2006), Cai and Li (2008), Tran and Tsionas (2010), Cai and Xiong (2012), and Su et al. (2014) focus on functional coefficient conditional mean regression models with endogenous regressors. On the other hand, the quantile regression model, which was pioneered by Koenker and Bassett (1978), has been widely used in various disciplines, including economics, finance, biology, and medicine. Despite the popularity of linear quantile regression models in the early literature (see, e.g., Koenker (2005) for an overview), the last two decades also witnessed a rapid growth of nonparametric and semiparametric quantile regression models. More recently, Honda (2004) and Kim (2007) study FCQR models for independent and identically distributed (IID) data using local polynomials and splines, respectively; and Cai and Xu (2008) and Cai and Xiao (2012) study local polynomial estimation of FCQR models and partially linear FCQR models for time series data, respectively; Wang et al. (2009) consider sieve estimation of partially linear FCQR models with longitudinal data. Compared with fully nonparametric quantile regression models, FCQR models serve as an intermediate class of models that are robust to model misspecification of functional coefficients and alleviate the notorious “curse of dimensionality” problem in the nonparametric literature. Unfortunately, none of these FCQR models allow for endogeneity.

In a series of papers, Chernozhukov and Hansen (2005, 2006, 2008) and Chernozhukov et al. (2009) address the important endogeneity issue in linear quantile regression models. They introduce an instrumental variable quantile regression (IVQR) estimator for heterogeneous treatment effect models to evaluate the impact of endogenous variables or treatments on the entire distribution of outcomes. Since then, their estimation strategy has been widely applied in the literature on quantile regression models with endogenous regressors. For example, Kaplan and Sun (2012) consider smoothed-estimating-equations IVQR estimator that improve over the original IVQR estimator in terms of computational speed and asymptotic efficiency; Chernozhukov et al. (2015) develop a new censored quantile IV estimator by extending the algorithm for censored quantile regression developed by Chernozhukov and Hong (2002). Extension to panel and spatial data models have also been done; see, e.g., Galvao and Montes-Rojas (2010), Galvao (2011) and Harding and Lamarche (2009, 2012, 2014), and Su and Yang (2012), respectively.

The purpose of this paper is to extend Chernozhukov and Hansen’s IVQR estimator further to the literature on functional coefficient models. There are several advantages associated with this extension. First, by adopting a functional coefficient quantile regression modeling strategy, we can model heterogeneous effects, account for both observed and unobserved heterogeneity, and put our model in the general framework of random coefficient models. We allow the heterogeneous effect of a regressor of interest on the outcome variable to vary across both the quantile indices and some observed covariates; see the examples in Section 2.1. Secondly, like Chernozhukov and Hansen (2006) the endogeneity issue in our model can be handled through a quantile analog of the two stage least squares. In particular, we can approximate the functional coefficients by basis functions and then obtain the sieve IVQR estimator as in the parametric case. So the computation for our estimator is as easy as that for the usual parametric IVQR estimation. Third, in the estimation context, the advantage of using the traditional constant coefficient IVQR models rest on their validity. Nevertheless, to the best of our knowledge, there is no

specification test available for this class of models. Using our sieve estimates of the functional coefficients, we provide a consistent nonparametric specification test for the constancy of the functional coefficients. If we fail to reject the null of constancy, then we can continue to rely on the traditional constant coefficient IVQR models. Otherwise we may have to consider the functional coefficients with unknown form.

Specifically, we develop nonparametric sieve estimation for a class of functional coefficient IVQR models where some or all the regressors are endogenous and their coefficients are varying with respect to some exogenous variables. In comparison with the widely used kernel estimation, the greatest advantage of sieve estimation lies in its computational simplicity, which can be a valid concern when bootstrap-based specification tests are considered and no closed solutions are available for the estimates. More importantly, it is well known that the kernel estimates (either the local polynomial, local constant, or nearest-neighborhood estimates) of nonparametric quantile functions tend to be rough, particularly for small or large values of quantile indices because only a small number of data points are essentially used in those regions. In this regard, sieve estimation might work better as it employs all observations in its global estimation procedure despite the fact that it may not be rich enough to characterize some local properties of the functional coefficients (c.f., Cai and Xu (2008)). After we study the asymptotic properties of the sieve IVQR estimates, we develop a new Wald-type test statistic for testing the hypothesis that a subvector of the functional coefficients is constant. The consistency, asymptotic null distribution, and asymptotic local power of the proposed test are established. In view of the well observed phenomenon that nonparametric tests based on the critical values from their asymptotic normal distributions may perform poorly in finite samples, we also provide a wild-bootstrap procedure to approximate the asymptotic null distribution of our test statistic and justify its asymptotic validity. To assess the finite sample properties of the proposed sieve IVQR estimator and the test statistic, we conduct a set of Monte Carlo simulations. The results show that our estimator performs well in finite samples and our test has approximately correct size and good power properties as the sample size increases, for various data generating processes under investigation. As an empirical illustration, we consider the estimation of quantile Engel curves for food using the U.K. Family Expenditure Survey data. We find that the effects of total expenditure on the food share vary over both the proportion of food expenditure and the age of household child, and they are significantly heterogeneous with respect to the age of household child, at the middle and higher quantiles.

The paper is organized as follows. In Section 2 we introduce our functional coefficient IVQR model and propose a sieve estimator for the functional coefficients. The asymptotic properties of the proposed estimator and its extensions to partially linear FCQR models and panel data models are studied in Section 3. We propose a nonparametric specification test for the widely used linear IVQR model and study its asymptotic properties in Section 4. We conduct a set of Monte Carlo studies to evaluate the finite sample performance of the proposed estimator and test in Section 5. Section 6 provides empirical data analysis and Section 7 concludes. All technical details are relegated to the appendix.

Notation. For natural numbers n_1 and n_2 , we use I_{n_1} to denote an $n_1 \times n_1$ identity matrix, and $0_{n_1 \times n_2}$ an $n_1 \times n_2$ matrix of zeros. We use $\mathbf{1}\{\cdot\}$ to denote the usual indicator function which takes value 1 if the condition inside the curly bracket holds and 0 otherwise, and c to signify a generic constant whose exact value may vary from case to case. For a matrix A , $\|A\|$ denotes its Frobenius norm: $\|A\| = \{\text{tr}(AA')\}^{1/2}$ where $\text{tr}(\cdot)$ is the trace operator and prime denotes transpose. When A is a symmetric matrix, we use $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ to denote its largest and smallest eigenvalues, respectively. We use \xrightarrow{d} and \xrightarrow{P} to denote convergence in distribution and probability, respectively. For any two conformable matrices or

vectors \tilde{G} and G , we write $\tilde{G} = G + \mathbf{o}_P(1)$ to denote $\|\tilde{G} - G\| = o_P(1)$. Of course, when the dimensions of \tilde{G} and G are fixed as we increase the sample size, we can interchange the use of $\mathbf{o}_P(1)$ and $o_P(1)$.

2 The model and estimator

In this section we introduce the functional coefficient quantile regression (FCQR) model with endogeneity and propose a sieve IVQR estimator for the vector of functional coefficients.

2.1 Functional coefficient quantile regression model with endogeneity

We consider the following structural quantile regression model

$$Y = \alpha_\tau(U)'D + \beta_\tau(U)'X + \varepsilon, \quad \tau \in (0, 1), \quad (2.1)$$

where Y is a scalar outcome variable, $D = (D_1, \dots, D_{k_1})'$ is a $k_1 \times 1$ vector of endogenous variables, $X = (X_1, \dots, X_{k_2})'$ is a $k_2 \times 1$ vector of exogenous variables, $\alpha_\tau(\cdot)$ and $\beta_\tau(\cdot)$ are $k_1 \times 1$ and $k_2 \times 1$ vectors of functional coefficients that vary with U , respectively, U is a $d \times 1$ vector of exogenous continuous variables, and $\varepsilon \equiv \varepsilon_\tau$ is the quantile error term such that

$$P(\varepsilon \leq 0 | U, X, Z) = \tau \text{ almost surely (a.s.)} \quad (2.2)$$

for a $k_3 \times 1$ vector of instrumental variables $Z = (Z_1, \dots, Z_{k_3})'$.¹ We assume that Z is correlated with D but is independent of ε , and $k_3 \geq k_1$. The model defined in (2.1) and (2.2) can be considered as a quantile counterpart to the models studied by Das (2005), Cai et al. (2006), Cai et al. (2010), Cai and Xiong (2012), and Su et al. (2014). The latter authors consider IV estimation of functional coefficient models under conditional mean restrictions. As Powell (2013) remarks, many empirical applications suggest that quantile regressions are useful as they provide richer empirical evidence than conditional mean regressions by allowing for *heterogenous* effects and estimating the *distributional* impacts of the explanatory variables.

To proceed, we provide two motivating examples for the model defined in (2.1) and (2.2) used in economics. See the empirical application in Section 6 for a third example.

Example 1. (Heterogeneous returns to education) Labor economists are interested in estimating the Mincer equation which describes the relationship between log-wage and schooling. Traditionally, they include schooling, experience, experience squared, and possibly some other control variables on the right hand side of the Mincer equation, regard schooling as an endogenous variable due to the unobserved heterogeneity in ability, and apply linear IV regression model. Nevertheless, Card (2001) finds that the returns to education tend to be underestimated by using the 2SLS method when one ignores the nonlinearity and the interaction between schooling and working experience, and Schultz (2003) argues that the marginal returns to education may vary with different levels of working experience and schooling. This motivates Su et al. (2014) to consider the following functional coefficient IV model

$$\ln(Wage) = \alpha(U) \cdot Schooling + \beta(U) + e$$

¹We can consider a slightly more general model than that in (2.1) where $\alpha_\tau(\cdot)$ and $\beta_\tau(\cdot)$ are functions of the $d_1 \times 1$ and $d_2 \times 1$ random vectors U_1 and U_2 , respectively. It is easy to see that our sieve estimation method works for this case with a little modification.

where U is a vector of control variables that includes working experience and some discrete demographic variables, and e denotes the error term. Similarly, Cai et al. (2010) consider the following partially linear functional coefficient model

$$\ln(Wage) = \alpha(U) \cdot Schooling + \beta'W + e$$

where U denotes working experience, and W denotes a vector of exogenous control variables that include marital status, union status, etc. Apparently, both models allow the impact of education on the log-wage to vary with working experience. Both groups of authors consider the IV estimation of their models under the usual conditional mean restrictions. Here we consider the IVQR analogue of the above models which allows us to estimate the *heterogeneous distributional impacts* of schooling on earnings through quantile regressions. Of course, we can also allow β in the second model to vary over U .

Example 2. (Heterogeneous effects of FDI on economic growth) Macro and international economists are typically interested in exploring the role of foreign direct investment (FDI) in economic growth. They usually regard the FDI inflows as an endogenous variable. Kottaridi and Stengos (2010) find that a beneficial effect of FDI on economic growth exists only for countries at higher levels of initial income. That is, the effect of FDI on economic growth varies across initial income levels. This motivates Cai et al. (2010) to consider the following partially linear functional coefficient model

$$Y = \alpha(U) \cdot fdi + \beta'W + e$$

where Y denotes the growth rate of income per capita, fdi is the ratio of FDI to the gross domestic product (GDP), U is the income level at some initial period, and W is a vector of exogenous control variables that include the logarithm of investment rate and population growth rate. Again, we can allow β in the above model to vary over U and the IVQR analogue of the resulting model allows us to study the *heterogeneous distributional impacts* of FDI on economic growth.

As Chernozhukov and Hansen (2006) argue, solving (2.1)-(2.2) as an IV quantile regression problem is to find a function $(u, d, x) \mapsto \alpha_\tau(u)'d + \beta_\tau(u)'x$ such that 0 is a solution to the ordinary quantile regression of $Y - \alpha_\tau(U)'D - \beta_\tau(U)'X$ on (U, X, Z) , i.e.,

$$0 \in \arg \min_{g_\tau \in \mathcal{G}_\tau} E[\rho_\tau(Y - \alpha_\tau(U)'D - \beta_\tau(U)'X - g_\tau(U, X, Z))], \quad (2.3)$$

where $\rho_\tau(w) = w(\tau - \mathbf{1}\{w \leq 0\})$ and \mathcal{G}_τ is a class of measurable functions of (U, X, Z) . In this paper, we restrict our attention to the class of functions $\mathcal{G}_\tau = \{g_\tau : g_\tau(U, X, Z) = \gamma_\tau(U)'Z \text{ for some measurable function } \gamma_\tau(\cdot)\}$. The resulting function $s(u, d, x) \equiv \alpha_\tau(u)'d + \beta_\tau(u)'x$ defines a *structural quantile function* (SQF) such that

$$P(Y \leq s(U, D, X) | U, X, Z) = \tau \text{ a.s.} \quad (2.4)$$

2.2 Sieve IVQR estimation

The main idea for our sieve IVQR estimation is simple. At the population level, for a given $\alpha \in \mathcal{A} \subset \mathbb{R}^{k_1}$ define

$$(\beta_\tau(\alpha, u), \gamma_\tau(\alpha, u)) \equiv \arg \min_{(\beta, \gamma) \in \mathcal{B} \times \Gamma} E[\rho_\tau(Y - \alpha'D - \beta'X - \gamma'Z) | U = u], \quad (2.5)$$

where $\mathcal{B} \times \Gamma \subset \mathbb{R}^{k_2+k_3}$. $\alpha_\tau(u)$ can be defined as $\alpha^* \in \mathcal{A} \subset \mathbb{R}^{k_1}$ such that $\|\gamma_\tau(\alpha^*, u)\| = 0$ under certain identification restriction. Then we have $(\beta_\tau(u), \gamma_\tau(u)) = (\beta_\tau(\alpha^*, u), \gamma_\tau(\alpha^*, u))$. In practice, one has to replace the above conditional expectation operator by its sample analogue. In principle, one can apply either kernel estimation or sieve estimation. Even though, as a referee kindly points out, kernel estimation has the advantage of its capability of capturing the local properties of the coefficient functionals and its asymptotic properties are also well documented in the literature, it is computationally demanding especially if one also considers bootstrap-based specification tests for the above structural quantile regression model. For this reason, we propose to estimate the functional coefficients by the method of sieve estimation in this paper. For an excellent review on sieve methods, see Chen (2007) or the book by Li and Racine (2007).

For expositional simplicity, we focus on the case where U takes value in a compact set $\mathcal{U} \subset \mathbb{R}^d$. Let $\alpha_{\tau,l}(u)$ and $\beta_{\tau,l}(u)$ denote the l 'th component of $\alpha_\tau(u)$ and $\beta_\tau(u)$, respectively, for $l = 1, \dots, k_1$ or k_2 . For each $u \in \mathcal{U}$, we approximate the functional coefficients $\alpha_{\tau,l}(u)$ and $\beta_{\tau,l}(u)$ by $p^K(u)'A_{\tau,l}$ and $p^K(u)'B_{\tau,l}$, respectively, for $l = 1, \dots, k_1$ or k_2 , where $p^K(u) = [p_1(u), \dots, p_K(u)]'$ is a $K \times 1$ vector of known basis functions, and $A_{\tau,l}$ and $B_{\tau,l}$ are $K \times 1$ vectors of unknown parameters to be estimated. Define

$$P_K(D, U) = [D_1 p^K(U)', \dots, D_{k_1} p^K(U)']' \text{ and } P_K(X, U) = [X_1 p^K(U)', \dots, X_{k_2} p^K(U)']'.$$

Then we can rewrite (2.1) as

$$Y = P_K(D, U)'A_\tau + P_K(X, U)'B_\tau + v + \varepsilon, \quad \tau \in (0, 1), \quad (2.6)$$

where $A_\tau \equiv (A'_{\tau,1}, \dots, A'_{\tau,k_1})'$, $B_\tau \equiv (B'_{\tau,1}, \dots, B'_{\tau,k_2})'$, and v is the approximation error term defined by $v \equiv v_\tau = (\alpha_\tau(U)'D - P_K(D, U)'A_\tau) + (\beta_\tau(U)'X - P_K(X, U)'B_\tau)$. Similarly, for $l = 1, \dots, k_3$ we approximate $\gamma_{\tau,l}(u)$, the l 'th component of $\gamma_\tau(u)$, by $p^K(u)'C_{\tau,l}$. Let $P_K(Z, U) = [Z_1 p^K(U)', \dots, Z_{k_3} p^K(U)']'$ and $C_\tau = (C'_{\tau,1}, \dots, C'_{\tau,k_3})'$.

Combining (2.3) and (2.6), (A_τ, B_τ, C_τ) can be characterized as follows. For a given $k_1 K \times 1$ vector A , let

$$\Theta_\tau(A) \equiv (B_\tau(A)', C_\tau(A)')' = \arg \min_{(B,C) \in \mathcal{B}_K \times \mathcal{C}_K} Q_\tau^K(A, B, C), \quad (2.7)$$

where

$$Q_\tau^K(A, B, C) \equiv E[\rho_\tau(Y - P_K(D, U)'A - P_K(X, U)'B - P_K(Z, U)'C)],$$

and \mathcal{B}_K and \mathcal{C}_K are compact parameter spaces in $\mathbb{R}^{k_2 K}$ and $\mathbb{R}^{k_3 K}$, respectively. By the continuity and convexity of the function $\rho_\tau(\cdot)$, we know that $\Theta_\tau(A)$ is continuous and uniquely defined for any $A \in \mathcal{A}_K$.

Then we have

$$A_\tau = \arg \min_{A \in \mathcal{A}_K} \|C_\tau(A)\|_{M_K}^2, \quad (2.8)$$

where $\mathcal{A}_K \subset \mathbb{R}^{k_1 K}$ is a compact parameter set, $\|C\|_{M_K}^2 = C'M_K C$ for a $k_3 K \times k_3 K$ symmetric positive definite weight matrix M_K (e.g., $M_K = I_{k_3 K}$). Then (B_τ, C_τ) can be represented by $\Theta_\tau \equiv (B'_\tau, C'_\tau)' \equiv (B_\tau(A_\tau)', C_\tau(A_\tau)')$.

Let $\{(Y_i, U_i, D_i, X_i, Z_i)\}_{i=1}^n$ be a random sample drawn from the distribution of (Y, U, D, X, Z) . For a sample analogue to the above procedure, we define our IVQR estimator of $\delta_\tau(u) \equiv (\alpha_\tau(u)', \beta_\tau(u)')$ for any $u \in \mathcal{U}$ as follows:

1. For a given $k_1 K \times 1$ parameter vector A , run a quantile regression to obtain

$$\hat{\Theta}_\tau(A) \equiv (\hat{B}_\tau(A)', \hat{C}_\tau(A)')' = \arg \min_{(B,C) \in \mathcal{B}_K \times \mathcal{C}_K} Q_{n\tau}^K(A, B, C) \quad (2.9)$$

where $Q_{n\tau}^K(A, B, C) \equiv n^{-1} \sum_{i=1}^n \rho_\tau(Y_i - P_K(D_i, U_i)'A - P_K(X_i, U_i)'B - P_K(Z_i, U_i)'C)$.

2. Minimize the weighted norm of $\hat{C}_\tau(A)$ over \mathcal{A}_K to obtain an estimator of $A_\tau = (A'_{\tau,1}, \dots, A'_{\tau,k_1})'$, i.e.,

$$\hat{A}_\tau \equiv (\hat{A}'_{\tau,1}, \dots, \hat{A}'_{\tau,k_1})' \equiv \arg \min_{A \in \mathcal{A}_K} \left\| \hat{C}_\tau(A) \right\|_{M_K}^2. \quad (2.10)$$

3. Obtain the estimator of $\Theta_\tau \equiv (B'_\tau, C'_\tau)'$ as $\hat{\Theta}_\tau \equiv \hat{\Theta}_\tau(\hat{A}_\tau) \equiv (\hat{B}_\tau(\hat{A}_\tau)', \hat{C}_\tau(\hat{A}_\tau)')'$. In particular, $\hat{B}_\tau \equiv \hat{B}_\tau(\hat{A}_\tau) \equiv (\hat{B}'_{\tau,1}, \dots, \hat{B}'_{\tau,k_2})'$ is an estimator of $B_\tau = (B'_{\tau,1}, \dots, B'_{\tau,k_2})'$.
4. For any $u \in \mathcal{U}$, the estimators of $\alpha_\tau(u)$ and $\beta_\tau(u)$ are given by $\hat{\alpha}_\tau(u) = [p^K(u)' \hat{A}_{\tau,1}, \dots, p^K(u)' \hat{A}_{\tau,k_1}]'$ and $\hat{\beta}_\tau(u) = [p^K(u)' \hat{B}_{\tau,1}, \dots, p^K(u)' \hat{B}_{\tau,k_2}]'$, respectively.

Remark 1. As mentioned above, a convenient choice of M_K in Step 2 is given by $I_{k_3 K}$. As in the IV literature, if $k_3 = k_1$ so that the IVQR model is just identified, we can demonstrate that the choice of M_K does not affect the asymptotic distributions of $\hat{\delta}_\tau(u) \equiv (\hat{\alpha}_\tau(u)', \hat{\beta}_\tau(u)')$ and our test statistic; see Remarks 4 and 9 below. If $k_3 > k_1$, one can apply $M_K = I_{k_3 K}$ to obtain preliminary estimators of those unknown parameters, based on which one obtains a consistent estimate for the variance-covariance of $\hat{C}_\tau(A)$ and use its inverse as M_K to obtain an asymptotically more efficient estimator of \hat{A}_τ . Since the asymptotics for the case of known M_K is already quite involved, we will not consider the case of estimated M_K in the following study.

3 Asymptotic properties of the sieve IVQR estimator

In this section we first provide assumptions for the identification and estimation, and then study the asymptotic properties of the sieve IVQR estimator proposed in the last section. Extensions to partially linear functional coefficient models and panel data models are also discussed later in this section.

3.1 Basic assumptions for identification and estimation

A real-valued function q on \mathcal{U} is said to satisfy a Hölder condition with exponent r if there is c_q such that $|q(u) - q(\tilde{u})| \leq c_q \|u - \tilde{u}\|^r$ for all $u, \tilde{u} \in \mathcal{U}$. Given a d -tuple nonnegative integers, $b = (b_1, \dots, b_d)$, set $[b] = b_1 + \dots + b_d$ and let ∇^b denote the differential operator defined by $\nabla^b = \frac{\partial^{[b]}}{\partial u_1^{b_1} \dots \partial u_d^{b_d}}$. A real-valued function q on \mathcal{U} is said to be λ -smooth, $\lambda = r + m$, if it is m -times continuously differentiable on \mathcal{U} and $\nabla^b q$ satisfies a Hölder condition with exponent r for all b with $[b] = m$. The λ -smooth class of functions are popular in econometrics because a λ -smooth function can be approximated well by various linear sieves; see, e.g., Chen (2007).

Let $W = (X', Z')'$, $\phi_W^K(U) \equiv [P_K(X, U)', P_K(Z, U)']'$ and $\psi_\tau(\varepsilon) \equiv \tau - \mathbf{1}\{\varepsilon \leq 0\}$. Define

$$\begin{aligned} \Pi_\tau(A, B) &\equiv E[\psi_\tau(Y - P_K(D, U)'A - P_K(X, U)'B) \phi_W^K(U)], \text{ and} \\ \Pi_{\tau|A}(B, C) &\equiv E[\psi_\tau(Y - P_K(D, U)'A - P_K(X, U)'B - P_K(Z, U)'C) \phi_W^K(U)]. \end{aligned}$$

Let $W_i = (X'_i, Z'_i)'$ and $\mathcal{W} = \mathcal{X} \times \mathcal{Z}$ be the support of W . We make the following set of basic assumptions.

Assumption A1. (i) (Y_i, U_i, D_i, W_i) , $i = 1, \dots, n$, are IID random variables that are defined on a probability space (Ω, \mathcal{F}, P) , share the same distribution as (Y, U, D, W) , and take values in $\mathcal{Y} \times \mathcal{U} \times \mathcal{D} \times \mathcal{W} \subset \mathbb{R}^{1+d+k_1+(k_2+k_3)}$, where \mathcal{U} and \mathcal{W} are compact and $k_3 \geq k_1$.

(ii) $P(Y \leq \alpha_\tau(U)'D + \beta_\tau(U)'X|U, W) = \tau$ a.s.

(iii) The cumulative distribution function (CDF) of ε conditional on $(U, D, W) = (u, d, w)$, $F_\varepsilon(\cdot|u, d, w)$, exhibits a probability density function (PDF) $f_\varepsilon(\cdot|u, d, w)$ that is bounded from above by \bar{c}_{f_ε} for all $(u, d, w) \in \mathcal{U} \times \mathcal{D} \times \mathcal{W}$; $f_\varepsilon(\cdot|u, d, w)$ is continuously differentiable in the neighborhood of 0 with first derivative bounded from above by $\bar{c}_{f'_\varepsilon}$ for all $(u, d, w) \in \mathcal{U} \times \mathcal{D} \times \mathcal{W}$; $E\{\sup_{v \in \mathbb{R}} \mathbf{1}\{|\varepsilon - v| \leq \vartheta(\chi)\}|\chi\} \leq 2\bar{c}_{f'_\varepsilon}\vartheta(\chi)$ for any measurable function $\vartheta(\cdot)$ where $\chi \equiv (U, D, W)$.

(iv) The distribution of U is absolutely continuous on \mathcal{U} with respect to the Lebesgue measure.

Assumption A2.(i) For $l = 1, \dots, k_1, k_2$, or k_3 , $\alpha_{\tau,l}(\cdot)$, $\beta_{\tau,l}(\cdot)$, and $\gamma_{\tau,l}(\cdot)$ belong to the class of λ -smooth functions with $\lambda > 0$.

(ii) For any λ -smooth function ϑ defined on \mathcal{U} , there exists a function $\Pi_{\infty K} \vartheta(\cdot) \equiv \pi'_{\vartheta} p^K(\cdot)$ in the sieve space $\mathcal{G}_K \equiv \{g(\cdot) = \pi' p^K(\cdot) \text{ for some } \pi \in \mathbb{R}^K\}$ such that $\sup_{u \in \mathcal{U}} \|\vartheta(u) - \Pi_{\infty K} \vartheta(u)\| = O(K^{-\lambda/d})$. In particular, there exist $A_{\tau,l}$, $B_{\tau,l}$, and $C_{\tau,l}$ such that $\sup_{u \in \mathcal{U}} |\alpha_{\tau,l}(u) - p^K(u)'A_{\tau,l}| = O(K^{-\lambda/d})$ for $l = 1, \dots, k_1$, $\sup_{u \in \mathcal{U}} |\beta_{\tau,l}(u) - p^K(u)'B_{\tau,l}| = O(K^{-\lambda/d})$ for $l = 1, \dots, k_2$, and $\sup_{u \in \mathcal{U}} |p^K(u)'C_{\tau,l}| = O(K^{-\lambda/d})$ for $l = 1, \dots, k_3$.

(iii) Let $A_\tau \equiv (A'_{\tau,1}, \dots, A'_{\tau,k_1})'$, $B_\tau \equiv (B'_{\tau,1}, \dots, B'_{\tau,k_2})'$, and $C_\tau \equiv (C'_{\tau,1}, \dots, C'_{\tau,k_3})'$. (A_τ, B_τ, C_τ) lies in the interior of $\mathcal{A}_K \times \mathcal{B}_K \times \mathcal{C}_K$, where $\mathcal{A}_K \subset \mathbb{R}^{k_1 K}$, $\mathcal{B}_K \subset \mathbb{R}^{k_2 K}$ and $\mathcal{C}_K \subset \mathbb{R}^{k_3 K}$ are compact and convex for all K , and \mathcal{C}_K contains 0 for all K .

Assumption A3. (i) For all K , the Jacobian matrices $\frac{\partial}{\partial(A', B')} \Pi_\tau(A, B)$ and $\frac{\partial}{\partial(B', C')} \Pi_{\tau|A}(B, C)$ exist, are continuous and have full rank uniformly over $\mathcal{A}_K \times \mathcal{B}_K \times \mathcal{C}_K$.

(ii) The image $\Pi_\tau(\mathcal{A}_K, \mathcal{B}_K)$ is simply connected for all K .

Assumption A4. Let $\mathcal{N}_{K,\epsilon} \equiv \{A \in \mathcal{A}_K, \|A - A_\tau\| < \epsilon\}$ be an ϵ -open subset of \mathcal{A}_K containing A_τ and $\mathcal{N}_{K,\epsilon}^c$ its complement. Let $\mathbf{Q}_K(A) \equiv \|C_\tau(A)\|_{M_K}^2$. Assume that $\liminf_{n \rightarrow \infty} [\min_{A \in \mathcal{A}_K \cap \mathcal{N}_{K,\epsilon}^c} \mathbf{Q}_K(A) - \mathbf{Q}_K(A_\tau)] > 0$ for all $\epsilon > 0$.

Assumption A1(i) imposes IID sampling and compactness on the support of the exogenous independent variables. As Chernozhukov and Hansen (2006) remark, compactness is not restrictive in microeconomic applications but can be relaxed at lengthy arguments. A1(ii) specifies the conditional quantile restriction which is used to construct our sieve IVQR estimator. A1(iii) imposes conditions on the quantile error term ε that are standard in the quantile regression literature. A1(iv) requires that the variables in U be continuously valued, which is standard in the literature on functional coefficient models. The extension to allow for both continuous and discrete variables in U is possible but will not be pursued in this paper.

Assumption A2(i) imposes smoothness conditions on the relevant functions and A2(ii) quantifies the approximation error of λ -smooth functions. These conditions are satisfied, for example, for polynomials, splines, and wavelets. A2(iii) imposes compactness on the parameter space. Such an assumption is needed at least for the parameter space \mathcal{A}_K because the objective function in (2.8) is not convex in A . A3 parallels Assumption 2.R3 in Chernozhukov and Hansen (2006) which is needed for the global identification. A4

specifies the *identifiable uniqueness* condition as defined in White (1994, p.28) or Gallant and White (1988, p.19). Note that they allow both the pseudo-true parameter (A_τ here) and the nonstochastic objective function ($\mathbf{Q}_K(\cdot)$ here) to depend on the sample size but restrict their attention to the case where the dimension of the parameter is fixed. Clearly A4 imposes some restrictions on the choice of basis functions that determine the solution $C_\tau(A)$ for any fixed $A \in \mathcal{A}_K$.

The following theorem describes the identification of the functional coefficients in the IVQR model.

Theorem 3.1 *Suppose that Assumptions A1-A4 hold. Then $\alpha_\tau(u)$, $\beta_\tau(u)$ and $\gamma_\tau(u)$ can be identified for all $u \in \mathcal{U}$ as $K \rightarrow \infty$.*

3.2 Asymptotic properties of the sieve IVQR estimators

To study the asymptotic properties of the sieve IVQR estimators, let $v_i(A, \Theta) = P_K(D_i, U_i)'A + \phi_{W_i}^K(U_i)'\Theta - \alpha_\tau(U_i)'D_i - \beta_\tau(U_i)'X_i$ and $g_i(A, \Theta) = \phi_{W_i}^K(U_i)\psi_\tau(\varepsilon_i - v_i(A, \Theta))$. Define

$$\begin{aligned}\Psi_K &= E[\phi_{W_i}^K(U_i)\phi_{W_i}^K(U_i)'], \\ J_{K,A}(A) &= -E\left[f_\varepsilon(v_i(A, \Theta_\tau(A)) | U_i, D_i, W_i) \phi_{W_i}^K(U_i) P_K(D_i, U_i)'\right], \\ \Phi_K(A) &= E\left[f_\varepsilon(v_i(A, \Theta_\tau(A)) | U_i, D_i, W_i) \phi_{W_i}^K(U_i) \phi_{W_i}^K(U_i)'\right], \\ \bar{\Phi}_{n,K}(A) &= \frac{1}{n} \sum_{i=1}^n f_\varepsilon(v_i(A, \Theta_\tau(A)) | U_i, D_i, W_i) \phi_{W_i}^K(U_i) \phi_{W_i}^K(U_i)'.\end{aligned}$$

Further, write $[\bar{\Phi}_{K,B}(A)' \quad \bar{\Phi}_{K,C}(A)']'$ as a conformable partition of $[\Phi_K(A)]^{-1}$, where $\bar{\Phi}_{K,B}(A)$ and $\bar{\Phi}_{K,C}(A)$ are $k_2 K \times (k_2 + k_3) K$ and $k_3 K \times (k_2 + k_3) K$ matrices, respectively. Let $\bar{\Phi}_{K,B} \equiv \bar{\Phi}_{K,B}(A_\tau)$, $\bar{\Phi}_{K,C} \equiv \bar{\Phi}_{K,C}(A_\tau)$, and $J_{K,A} \equiv J_{K,A}(A_\tau)$. Finally, let $\Omega_\tau \equiv (\Omega'_{A_\tau}, \Omega'_{B_\tau})'$, where

$$\begin{aligned}\Omega_{A_\tau} &\equiv - (J'_{K,A} \bar{\Phi}'_{K,C} M_K \bar{\Phi}_{K,C} J_{K,A})^{-1} J'_{K,A} \bar{\Phi}'_{K,C} M_K \bar{\Phi}_{K,C} \\ \Omega_{B_\tau} &\equiv \bar{\Phi}_{K,B} [I_{(k_2+k_3)K} + J_{K,A} \Omega_{A_\tau}].\end{aligned}$$

We add the following assumptions.

- Assumption A5.** (i) $0 < \underline{c}_M \leq \lambda_{\min}(M_K) \leq \lambda_{\max}(M_K) \leq \bar{c}_M < \infty$ uniformly in K ;
(ii) $0 < \underline{c}_\Psi \leq \lambda_{\min}(\Psi_K) \leq \lambda_{\max}(\Psi_K) \leq \bar{c}_\Psi < \infty$ uniformly in K ;
(iii) $0 < \underline{c}_\Phi \leq \inf_{A \in \mathcal{A}_K} \lambda_{\min}(\Phi_K(A)) \leq \sup_{A \in \mathcal{A}_K} \lambda_{\max}(\Phi_K(A)) \leq \bar{c}_\Phi < \infty$ uniformly in K ;
(iv) $0 < \underline{c}_J \leq \lambda_{\min}(J_{K,A} J'_{K,A}) \leq \lambda_{\max}(J_{K,A} J'_{K,A}) \leq \bar{c}_J < \infty$ uniformly in K .

Assumption A6. (i) Let $\zeta_K \equiv \sup_{u \in \mathcal{U}} \|p^K(u)\|$. As $n \rightarrow \infty$, $\zeta_K^2 K^3 (\ln n)^2 / n \rightarrow 0$ and $n K^{-(1+2\lambda/d)} / \ln n \rightarrow c_0 \in [0, \infty)$.

Assumption A5(i) imposes conditions on the weight matrix M_K and is trivially satisfied for $I_{k_3 K}$ with $\underline{c}_M = \bar{c}_M = 1$. The condition in A5(ii) is standard in the literature on sieve estimation (e.g., Newey (1997)). For fixed A , A5(iii) reduces to the typical requirement for sieve estimation of conditional quantiles without endogeneity (e.g., Lee and Horowitz (2005)). The uniform requirement on $A \in \mathcal{A}_K$ pertains to our sieve IVQR estimation and can be satisfied under A5(iii) if $f_\varepsilon(\cdot | U_i, D_i, W_i)$ is uniformly bounded away from 0 and infinity a.s. A5(iv) requires that $J_{K,A} \equiv J_{K,A}(A_\tau)$ has full rank for all K in large samples. In Lemma A.1 in the appendix, we show that Assumptions A5(i)-A5(iv) imply that $\Omega_\tau \Omega'_\tau$

has eigenvalues that are bounded away from infinity and zero uniformly in K in large samples. This is important as Ω_τ appears in the asymptotic variance-covariance matrix for our sieve IVQR estimator; see Theorem 3.5 below.

Assumption A6(i) imposes conditions on K , ζ_K and λ . For some basis functions, the order ζ_K is well known. For example, $\zeta_K = O(K)$ for power series and $\zeta_K = O(K^{1/2})$ for splines (see Newey (1997)). The first condition in A6(i) requires that K should not diverge to infinity too fast while the second requires that K should not diverge too slowly and it is sufficient to ensure that the asymptotic bias term of our first stage sieve estimator $\hat{\Theta}_\tau(A)$ is at most as large as its variance term uniformly in $A \in \mathcal{A}_K$. The larger value λ/d takes (i.e., the smoother the class of functional coefficients), the less stringent condition needed for K in order for both conditions in A6(i) to be satisfied.

The next theorem studies the uniform convergence and the uniform Bahadur representation of the first stage estimator $\hat{\Theta}_\tau(A)$.

Theorem 3.2 *Suppose that Assumptions A1-A4, A5(i)-(iii), and A6(i) hold. Then*

- (i) $\sup_{A \in \mathcal{A}_K} \|\hat{\Theta}_\tau(A) - \Theta_\tau(A)\| = O_P[(K \ln n/n)^{1/2}]$,
 - (ii) $\hat{\Theta}_\tau(A) - \Theta_\tau(A) = \Phi_K(A)^{-1} n^{-1} \sum_{i=1}^n g_i(A, \Theta_\tau(A)) + \mathbf{o}_P(n^{-1/2}) + r_n$ uniformly in $A \in \mathcal{A}_K$,
- where $\|r_n\| = O_P(\zeta_K^{1/2} K^{5/4} n^{-3/4} \ln n)$.

Remark 2. The first condition in Assumption A6(i) ensures $\|r_n\| = o_P((K/n)^{1/2})$. This condition ensures that r_n is of smaller order term than the dominant term in the Bahadur representation for $\hat{\Theta}_\tau(A)$. If one requires $\|r_n\| = o_P(n^{-1/2})$ to simplify the expression in Theorem 3.2, one needs to strengthen that condition to $\zeta_K^2 K^5 (\ln n)^4 / n \rightarrow 0$. In the case of spline estimation, $\zeta_K = O(K^{1/2})$, the latter condition is simplified to $K^6 (\ln n)^4 / n \rightarrow 0$, which means that K cannot increase at a rate faster than $n^{1/6}$.

We study the consistency of the $(\hat{A}_\tau, \hat{B}_\tau, \hat{C}_\tau)$ and derive the influence functions for \hat{A}_τ and \hat{B}_τ in the following theorem.

Theorem 3.3 *Suppose that Assumptions A1-A4, A5(i)-(iii), and A6(i) hold. Then*

- (i) $\|\hat{A}_\tau - A_\tau\| = o_P(1)$, $\|\hat{B}_\tau - B_\tau\| = o_P(1)$, and $\|\hat{C}_\tau - C_\tau\| = o_P(1)$;
- (ii) $\hat{A}_\tau - A_\tau = \Omega_{A_\tau} n^{-1} \sum_{i=1}^n \phi_{W_i}^K(U_i) \psi_\tau(\varepsilon_i - v_i(A_\tau, \Theta_\tau)) + \mathbf{o}_P(n^{-1/2})$;
- (iii) $\hat{B}_\tau - B_\tau = \Omega_{B_\tau} n^{-1} \sum_{i=1}^n \phi_{W_i}^K(U_i) \psi_\tau(\varepsilon_i - v_i(A_\tau, \Theta_\tau)) + \mathbf{o}_P(n^{-1/2})$.

The next theorem gives the uniform rate of convergence of our sieve IVQR estimator.

Theorem 3.4 *Suppose that Assumptions A1-A4, A.5(i)-(iii), and A6(i) hold. Then*

- (i) $\sup_{u \in \mathcal{U}} \|\hat{\alpha}_\tau(u) - \alpha_\tau(u)\| = O_P[\zeta_K((K/n)^{1/2} + K^{-\lambda/d})]$;
- (ii) $\sup_{u \in \mathcal{U}} \|\hat{\beta}_\tau(u) - \beta_\tau(u)\| = O_P[\zeta_K((K/n)^{1/2} + K^{-\lambda/d})]$.

Remark 3. Despite the complication of our estimation strategy, Theorem 3.4 indicates that we can obtain the same uniform convergence rate as obtained in the sieve estimation of conditional mean function; see Theorem 1 in Newey (1997). For the selection of K , Newey (1997) mainly requires that $\zeta_K^2 K/n \rightarrow 0$, which is much weaker than our first requirement that $\zeta_K^2 K^3 (\ln n)^2 / n \rightarrow 0$. This is as expected because our estimator is essentially a two-stage estimator and we have to apply some uniform convergence results to demonstrate our first-stage estimator $\hat{\Theta}_\tau(A)$ is uniformly consistent in Frobenius norm and exhibits

certain uniform Bahadur representation. Following the proof of Theorem 3.4, one can also obtain the mean square convergence rate:

$$\begin{aligned} \int \|\hat{\alpha}_\tau(u) - \alpha_\tau(u)\|^2 dF_u(u) &= O_P(K/n + K^{-2\lambda/d}) \text{ and} \\ \int \|\hat{\beta}_\tau(u) - \beta_\tau(u)\|^2 dF_u(u) &= O_P(K/n + K^{-2\lambda/d}), \end{aligned}$$

where $F_u(\cdot)$ denotes the CDF of U . We skip the details to conserve space.

To study the asymptotic distribution of our estimator, we introduce some additional notation. Let $\Pi_\alpha(u)$ and $\Pi_\beta(u)$ be $k_1 \times k_1 K$ and $k_2 \times k_2 K$ block diagonal matrices, respectively, whose diagonal block

$$\begin{aligned} \text{is } p^K(u)'; \text{ e.g., } \Pi_\alpha(u) &= \begin{bmatrix} p^K(u)' & \cdots & 0_{1 \times K} \\ \vdots & \ddots & \vdots \\ 0_{1 \times K} & \cdots & p^K(u)' \end{bmatrix}. \text{ Define the } (k_1 + k_2) \times (k_1 + k_2) K \text{ matrix} \\ \Pi(u) &\equiv \begin{pmatrix} \Pi_\alpha(u) & 0_{k_1 \times k_2 K} \\ 0_{k_2 \times k_1 K} & \Pi_\beta(u) \end{pmatrix}. \end{aligned} \quad (3.1)$$

We add the following assumption.

Assumption A6. (ii) As $n \rightarrow \infty$, $n\zeta_{K,u}^{-2} K^{-2\lambda/d} \rightarrow 0$ where $\zeta_{K,u} \equiv \|\Pi(u)\| > 0$.

Note that Assumption A6(ii) is similar to the second requirement in A6(i) and it ensures that the asymptotic bias term for our sieve IVQR estimator is of smaller order than the asymptotic variance term. In general, one expects that $\|\Pi(u)\| = O(K^{1/2})$, and thus A6(ii) reduces to the typical requirement that $nK^{-(1+2\lambda/d)} \rightarrow 0$; see, e.g., Huang (2003).

The following theorem studies the asymptotic normality of our sieve IVQR estimator.

Theorem 3.5 *Suppose that Assumptions A1-A6 hold. Then*

$$\left\{ \tau(1-\tau) \Pi(u) \Omega_\tau \Psi_K \Omega_\tau' \Pi(u)' \right\}^{-1/2} \sqrt{n} \begin{pmatrix} \hat{\alpha}_\tau(u) - \alpha_\tau(u) \\ \hat{\beta}_\tau(u) - \beta_\tau(u) \end{pmatrix} \xrightarrow{d} N(0, I_{k_1+k_2}).$$

Remark 4. In the above study we restrict our attention to the case where the weight matrix M_K used in (2.10) is nonrandom. In the case of just-identification (i.e., $k_3 = k_1$), $\Omega_{A_\tau} \equiv -(\bar{\Phi}_{K,C} J_{K,A})^{-1} \bar{\Phi}_{K,C}$, and $\Omega_{B_\tau} \equiv \bar{\Phi}_{K,B} [I_{(k_2+k_3)K} - J_{K,A} (\bar{\Phi}_{K,C} J_{K,A})^{-1} \bar{\Phi}_{K,C}]$. Therefore the choice of M_K has no effect on the asymptotic distribution of our sieve IVQR estimator, and one can always set $M_K = I_{k_3 K}$. In the case of over-identification, however, the choice of M_K generally matters. It affects the asymptotic efficiency of \hat{A}_τ and $\hat{\alpha}_\tau(u)$ and that of \hat{B}_τ and $\hat{\beta}_\tau(u)$. Here we focus on the estimation of the structural functional coefficient $\alpha_\tau(u)$. For the general choice of M_K , the asymptotic variance-covariance (AVC) matrix of $\sqrt{n}(\hat{\alpha}_\tau(u) - \alpha_\tau(u))$ is given by

$$\begin{aligned} \Omega_{\alpha_\tau}(M_K) &\equiv \tau(1-\tau) \Pi_\alpha(u) \Omega_{A_\tau} \Psi_K \Omega_{A_\tau}' \Pi_\alpha(u)' \\ &= \tau(1-\tau) \Pi_\alpha(u) (J'_{K,A} \bar{\Phi}'_{K,C} M_K \bar{\Phi}_{K,C} J_{K,A})^{-1} J'_{K,A} \bar{\Phi}'_{K,C} M_K \bar{\Phi}_{K,C} \Psi_K \bar{\Phi}'_{K,C} M_K \bar{\Phi}_{K,C} J_{K,A} \\ &\quad \times (J'_{K,A} \bar{\Phi}'_{K,C} M_K \bar{\Phi}_{K,C} J_{K,A})^{-1} \Pi_\alpha(u)'. \end{aligned}$$

By (A.10) and the proof of Theorem 3.5 in the appendix, the AVC matrix of $\sqrt{n}(\hat{C}_\tau - C_\tau)$ is given by $\Sigma_{C_\tau} \equiv \tau(1-\tau) \bar{\Phi}_{K,C} \Psi_K \bar{\Phi}'_{K,C}$. Then, if we choose $M_K = \Sigma_{C_\tau}^{-1}$, the above AVC matrix reduces to

$$\Omega_{\alpha_\tau}(\Sigma_{C_\tau}^{-1}) \equiv \tau(1-\tau) \Pi_\alpha(u) \left[J'_{K,A} \bar{\Phi}'_{K,C} (\bar{\Phi}_{K,C} \Psi_K \bar{\Phi}'_{K,C})^{-1} \bar{\Phi}_{K,C} J_{K,A} \right]^{-1} \Pi_\alpha(u)'.$$

Standard arguments show that $\Omega_{\alpha_\tau}(M_K) \geq \Omega_{\alpha_\tau}(\Sigma_{C_\tau}^{-1})$. It follows that by setting $M_K = \Sigma_{C_\tau}^{-1}$ we can obtain the most efficient sieve IVQR estimator of the structural functional coefficient $\alpha_\tau(u)$.

Remark 5. In practice, Σ_{C_τ} is not feasible and one may estimate it based on some preliminary consistent estimators. For example, one can first choose $M_K = I_{k_3K}$ to obtain a preliminary estimate $\hat{\delta}_\tau(u)$ of $\delta_\tau(u)$ at all data points and the resulting quantile regression residuals $\hat{\varepsilon}_i$. Let $\hat{\Sigma}_{C_\tau} = \tau(1-\tau)\hat{\Phi}_{K,C}\hat{\Psi}_K\hat{\Phi}_{K,C}'$, where $\hat{\Psi}_K$ and $\hat{\Phi}_K$ are defined below and $\hat{\Phi}_{K,C}$ is the lower $k_3K \times (k_2+k_3)K$ submatrix of $\hat{\Phi}_K^{-1}$. Then a feasible version of the optimal choice of M_K is given by $\hat{M}_K = \hat{\Sigma}_{C_\tau}^{-1}$. Under some regularity conditions, we can show that $\|\hat{\Sigma}_{C_\tau}^{-1} - \Sigma_{C_\tau}^{-1}\| = O_P[K^{1/2}((K/n)^{1/2} + K^{-\lambda/d})]$ and the estimation error does not affect the distributional theory of our estimator.

For statistical inference, it is necessary to obtain consistent estimators of Φ_K , $J_{K,A}$ and Ψ_K . A natural estimator for Ψ_K is $\hat{\Psi}_K \equiv \frac{1}{n} \sum_{i=1}^n \phi_{W_i}^K(U_i) \phi_{W_i}^K(U_i)'$. Following Powell (1991), we can estimate Φ_K and $J_{K,A}$ respectively by

$$\hat{\Phi}_K \equiv \frac{1}{2nh} \sum_{i=1}^n 1\{|\hat{\varepsilon}_i| \leq h\} \phi_{W_i}^K(U_i) \phi_{W_i}^K(U_i)' \text{ and } \hat{J}_{K,A} \equiv \frac{-1}{2nh} \sum_{i=1}^n 1\{|\hat{\varepsilon}_i| \leq h\} \phi_{W_i}^K(U_i) P_K(D_i, U_i)',$$

where $h \equiv h_n$ is a bandwidth parameter such that as $n \rightarrow \infty$, $h \rightarrow 0$ and $K^2/(nh) \rightarrow 0$. To compute $\hat{\Phi}_K$ and $\hat{J}_{K,A}$, we need to choose the smoothing parameter h . Following Koenker (2005, pp. 80-81), we can set $h = \hat{\kappa} [\Phi^{-1}(\tau + cn^{-1/3}) - \Phi^{-1}(\tau - cn^{-1/3})]$, where Φ^{-1} is the inverse of the standard normal CDF, $\hat{\kappa}$ is a robust estimate of scale/standard deviation, e.g., $\hat{\kappa} = \text{median}|\hat{\varepsilon}_i - \text{median}(\hat{\varepsilon}_i)|/0.6745$ (Hogg and Craig, 1995, p. 390), and c is a proportional constant. In the simulation and application below we set $c = 0.5$.² Following the consistency of these estimators (as demonstrated in the proof of Theorem 4.3 below), we can readily obtain a consistent estimator for the asymptotic variance-covariance matrix in the above theorem.

3.3 Extension to partially linear functional coefficient models

Now we consider extending the model in (2.1) to the partially linear functional coefficient model:

$$Y = \alpha_\tau(U)'D + \beta_{1\tau}(U)'X_1 + \beta_{2\tau}'X_2 + \varepsilon, \quad \tau \in (0, 1), \quad (3.2)$$

where Y , U , D , and $\alpha_\tau(\cdot)$ are defined as above, $X_1 = (X_1, \dots, X_{k_{21}})'$ and $X_2 = (X_{k_{21}+1}, \dots, X_{k_{21}+k_{22}})'$ are $k_{21} \times 1$ and $k_{22} \times 1$ vectors of exogenous variables, respectively, $\beta_{1\tau}(\cdot)$ is a $k_{21} \times 1$ vector of functional coefficients that vary with U , $\beta_{2\tau}$ is $k_{22} \times 1$ vector of coefficients that do not vary with U , and $\varepsilon \equiv \varepsilon_\tau$ is the quantile error term such that

$$P(\varepsilon \leq 0 | U, X_1, X_2, Z) = \tau \text{ a.s.} \quad (3.3)$$

for a $k_3 \times 1$ vector of instrumental variables Z . In the absence of the endogenous component $\alpha_\tau(U)'D$, the model in (3.2) was recently studied by Wang et al. (2009) for longitudinal data and Cai and Xiao (2012) for time series data. Because of the presence of the endogenous component, the methodology developed in neither paper applies to our framework.

²Alternatively, one could follow Pakes and Pollard (1989) and Honoré and Hu (2004) and estimate Φ_K and $J_{K,A}$ using numerical derivatives.

We remark that the results developed in previous sections extend straightforwardly to the model specified in (3.2) and (3.3). To conserve space, we only consider the asymptotic normality of the estimator of the finite dimensional parameter $\beta_{2\tau}$ as the results for the other functional parameters, namely, $\alpha_\tau(\cdot)$ and $\beta_{1\tau}(\cdot)$, are almost the same as before. We follow closely the notation defined above and only suggest necessary changes as follows:

$$\begin{aligned} P_K(\mathbf{X}_1, U) &\equiv [X_{1p}^K(U)', \dots, X_{k_{21}p}^K(U)']', \\ W &\equiv [\mathbf{X}'_1, \mathbf{X}'_2, Z']', \\ \phi_W^K(U) &\equiv [P_K(\mathbf{X}_1, U)', \mathbf{X}'_2, P_K(Z, U)']', \\ B_\tau &\equiv (B'_{1\tau}, \beta'_{2\tau})', \end{aligned}$$

where $B_{1\tau} \equiv (B'_{1\tau,1}, \dots, B'_{1\tau,k_{21}})'$, and $B_{1\tau,l}$, $l = 1, \dots, k_{21}$, are defined as in Section 2.2. Clearly $\phi_W^K(U)$ is a $[(k_{21} + k_3)K + k_{22}] \times 1$ vector, and B_τ is $(k_{21}K + k_{22}) \times 1$ vector. Let \hat{A}_τ and $\hat{B}_\tau \equiv (\hat{B}'_{1\tau}, \hat{\beta}'_{2\tau})'$ be the sieve IVQR estimator of A_τ and B_τ , respectively, by following the exact procedure specified in Section 2.2. Let Ψ_K , $J_{K,A}(A)$, $\Phi_K(A)$, and Ω_{B_τ} be defined as in Section 3.2 with the newly defined $\phi_W^K(U)$ in place of the original one. Now we need that Assumptions A.5(ii)-(iv) hold for these newly defined objects. Then the results in Theorems 3.2 and 3.3 continue to hold. Consequently, we have

$$\begin{aligned} \hat{B}_\tau - B_\tau &= \Omega_{B_\tau} n^{-1} \sum_{i=1}^n \phi_{W_i}^K(U_i) \psi_\tau(\varepsilon_i - v_i(A_\tau, \Theta_\tau)) + \mathbf{o}_P(n^{-1/2}) \\ &= \Omega_{B_\tau} n^{-1} \sum_{i=1}^n \phi_{W_i}^K(U_i) \psi_\tau(\varepsilon_i) + \mathbf{o}_P(n^{-1/2}) \end{aligned}$$

provide that the approximation error $v_i(A_\tau, \Theta_\tau)$ is asymptotically negligible (which holds if $nK^{-2\lambda/d} \rightarrow 0$). Let \mathbb{S}_2 be a $k_2 \times (k_{21}K + k_{22})$ selection matrix such that $\mathbb{S}_2 B$ selects only the last k_2 elements in the $(k_{21}K + k_{22}) \times 1$ vector B . Then we can easily demonstrate that

$$\begin{aligned} \sqrt{n} \left(\hat{\beta}_{2\tau} - \beta_{2\tau} \right) &= \sqrt{n} \mathbb{S}_2 \left(\hat{B}_\tau - B_\tau \right) = \mathbb{S}_2 \Omega_{B_\tau} n^{-1/2} \sum_{i=1}^n \phi_{W_i}^K(U_i) \psi_\tau(\varepsilon_i) + \mathbf{o}_P(1) \\ &\xrightarrow{d} N \left(0, \tau(1-\tau) \lim_{K \rightarrow \infty} \mathbb{S}_2 \Omega_{B_\tau} \Psi_K \Omega'_{B_\tau} \mathbb{S}'_2 \right). \end{aligned}$$

Consequently, one can conduct statistical inference on $\beta_{2\tau}$ as usual by estimating the AVC matrix given above. Alternatively, one can apply the bootstrap method to obtain standard errors and make inference.

3.4 Extension to panel data models with individual fixed effects

Now, we consider the extension of the model in (2.1) to a panel data model with individual fixed effects:

$$Y_{it} = \alpha_\tau(U_{it})' D_{it} + \beta_{1\tau}(U_{it})' \mathbf{X}_{1,it} + \beta_{2\tau,i} + \varepsilon_{it}, \quad \tau \in (0, 1), \quad (3.4)$$

where $i = 1, \dots, N$, $t = 1, \dots, T$, $\mathbf{X}_{1,it} = (X_{1,it}, \dots, X_{k_2,it})'$, $\beta_{2\tau,i}$'s are individual fixed effects that will be treated as parameters to be estimated, and the definitions of other objects are the same as before. Note that we follow Kato et al. (2012) and allow the individual effects to vary across the quantile index

τ .³ Since quantile regression with individual effects is subject to the incidental parameter problem and so far there is no general transformation that can suitably eliminate the individual effects in quantile regressions, here we follow the literature, assume that both N and T go to infinity, and focus on the estimation of the functional coefficients $\delta_\tau(\cdot) \equiv (\alpha_\tau(\cdot)', \beta_{1\tau}(\cdot)')$.

Let $\mathbf{X}_{2,it}$ denote the i th column of I_N for each t . Then we can rewrite the model in (3.4) as

$$Y_{it} = \alpha_\tau(U_{it})'D_{it} + \beta_{1\tau}(U_{it})'\mathbf{X}_{1,it} + \beta_{2\tau}'\mathbf{X}_{2,it} + \varepsilon_{it}, \quad \tau \in (0, 1),$$

where $\beta_{2\tau} = (\beta_{2\tau,1}, \dots, \beta_{2\tau,N})'$. If we assume that $\varepsilon_{it} = \varepsilon_{it,\tau}$ satisfies

$$P(\varepsilon_{it} \leq 0 | U_{it}, \mathbf{X}_{1,it}, \beta_{2\tau,i}, Z_{it}) = \tau \text{ a.s.}$$

for a $k_3 \times 1$ vector of instrumental variables Z_{it} , we can estimate the model using the same estimation procedure as in the partially linear FCQR model presented above. Let

$$\begin{aligned} P_K(\mathbf{X}_{1,it}, U_{it}) &\equiv [X_{1,it}p^K(U_{it})', \dots, X_{k_2,it}p^K(U_{it})']', \\ P_K^\dagger(\mathbf{X}_{it}, U_{it}) &\equiv [P_K(\mathbf{X}_{1,it}, U_{it})', \mathbf{X}_{2,it}']', \\ B_\tau &\equiv (B_{1\tau}', \beta_{2\tau}')', \end{aligned}$$

where $B_{1\tau} \equiv (B_{1\tau,1}', \dots, B_{1\tau,k_2}')'$, and $B_{1\tau,l}$, $l = 1, \dots, k_2$, are defined as in Section 2.2. Here, B_τ is $(k_2K + N) \times 1$ vector. The sieve IVQR estimators can be defined analogously to those in section 2.2 with the objective function in (2.9) replaced by

$$Q_{NT\tau}^K(A, B, C) \equiv (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T \rho_\tau(Y_{it} - P_K(D_{it}, U_{it})'A - P_K^\dagger(\mathbf{X}_{it}, U_{it})'B - P_K(Z_{it}, U_{it})'C). \quad (3.5)$$

Let $\Theta = (B', C')' \in \mathbb{R}^{k_2K+N} \times \mathbb{R}^{k_3K}$. Let $\Theta_\tau(A) \equiv (B_\tau(A)', C_\tau(A)')'$ and A_τ are as defined in (2.7) and (2.8), respectively. Let $\hat{A}_\tau, \hat{B}_\tau \equiv (\hat{B}_{1\tau}', \hat{\beta}_{2\tau}')'$ and \hat{C}_τ be the sieve IVQR estimator of A_τ, B_τ , and C_τ , respectively. The sieve IVQR estimators of $\alpha_\tau(u)$ and $\beta_{1\tau}(u)$ are given by $\hat{\alpha}_\tau(u) = [p^K(u)' \hat{A}_{\tau,1}, \dots, p^K(u)' \hat{A}_{\tau,k_1}]'$ and $\hat{\beta}_{1\tau}(u) = [p^K(u)' \hat{B}_{1\tau,1}, \dots, p^K(u)' \hat{B}_{1\tau,k_2}]'$, respectively.

Let $W_{it} \equiv (\mathbf{X}_{1,it}', \mathbf{X}_{2,it}', Z_{it}')', \phi_{W_{it}}^K(U_{it}) \equiv [P_K^\dagger(\mathbf{X}_{it}, U_{it})', P_K(Z_{it}, U_{it})']'$, and $v_{it}(A, \Theta) = P_K(D_{it}, U_{it})'A + \phi_{W_{it}}^K(U_{it})' \Theta - \alpha_\tau(U_{it})'D_{it} - \beta_{1\tau}(U_{it})'\mathbf{X}_{1,it} - \beta_{2\tau}'\mathbf{X}_{2,it}$. Let $\chi_{it} = (U_{it}', D_{it}', W_{it}')'$, $H = \text{diag}(I_{k_2K}, NK^{-1/2}I_N, I_{k_3K})$ and $H_1 = \text{diag}(I_{k_2K}, N^{1/2}I_N, I_{k_3K})$. Define

$$\begin{aligned} \Psi_K &= (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T E \left[H_1 \phi_{W_{it}}^K(U_{it}) \phi_{W_{it}}^K(U_{it})' H_1 \right], \\ J_{K,A}(A) &= -(NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T E \left[f_{\varepsilon_{it}}(v_{it}(A, \Theta_\tau(A)) | \chi_{it}) H_1 \phi_{W_{it}}^K(U_{it}) P_K(D_{it}, U_{it})' \right], \\ \Phi_K(A) &= (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T E \left[f_{\varepsilon_{it}}(v_{it}(A, \Theta_\tau(A)) | \chi_{it}) H_1 \phi_{W_{it}}^K(U_{it}) \phi_{W_{it}}^K(U_{it})' H_1 \right], \end{aligned}$$

³Galvao (2011) applies the IVQR method to estimate dynamic panel data models by using lagged regressors or lagged differences of regressors as instruments for the lagged variable $Y_{i,t-1}$ which plays the role of D_{it} here. Galvao and Montes-Rojas (2010) consider penalized IVQR of dynamic panel data models by imposing ℓ_1 -penalty on the fixed effects, following the lead of Koenker (2004). Both papers provide proofs based on the *heuristic* arguments used in Koenker (2004). In particular, they claim that their \sqrt{NT} -consistency and asymptotic normality results hold as long as $N^a/T \rightarrow 0$ for some $a > 0$ under some regularity conditions, which is apparently not the case. Indeed, Kato et al. (2012) establish the \sqrt{NT} -consistency and asymptotic normality results for the conventional panel quantile regression estimators under the conditions $N^2(\ln N)^3/T \rightarrow 0$ as $(N, T) \rightarrow \infty$.

where $f_{\varepsilon_{it}}(\cdot|\chi_{it})$ denotes the PDF of ε_{it} given χ_{it} . As in Section 3.2, we write $[\bar{\Phi}_{K,B}(A)' \bar{\Phi}_{K,C}(A)']'$ as a conformable partition of $[H_1 H^{-1} \Phi_K(A) H_1^{-1} H]^{-1}$, where $\bar{\Phi}_{K,B}(A)$ and $\bar{\Phi}_{K,C}(A)$ are now $(k_2 K + N) \times [(k_2 + k_3) K + N]$ and $k_3 K \times [(k_2 + k_3) K + N]$ matrices, respectively. Let $\bar{\Phi}_{K,B} \equiv \bar{\Phi}_{K,B}(A_\tau)$, $\bar{\Phi}_{K,C} \equiv \bar{\Phi}_{K,C}(A_\tau)$, and $J_{K,A} \equiv J_{K,A}(A_\tau)$. Further, let $\Omega_\tau \equiv (\Omega'_{A_\tau}, \Omega'_{B_{1\tau}})'$, where

$$\begin{aligned}\Omega_{A_\tau} &\equiv - (J'_{K,A} H^{-1} H_1 \bar{\Phi}'_{K,C} M_K \bar{\Phi}_{K,C} H_1 H^{-1} J_{K,A})^{-1} J'_{K,A} H^{-1} H_1 \bar{\Phi}'_{K,C} M_K \bar{\Phi}_{K,C}, \\ \Omega_{B_{1\tau}} &\equiv \bar{\Phi}_{K,B_1} [I_{(k_2+k_3)K+N} + H_1 H^{-1} J_{K,A} \Omega_{A_\tau}],\end{aligned}$$

where $\bar{\Phi}_{K,B_1}$ denotes the upper $k_2 K \times [(k_2 + k_3) K + N]$ submatrix of $[H_1 H^{-1} \Phi_K(A_\tau) H_1^{-1} H]^{-1}$ (or equivalently that of $\bar{\Phi}_{K,B}$).

We only state two theorems that parallel Theorems 3.2 and 3.5 under Assumptions D1-D6 stated in Supplementary Appendix D. The counterparts of other theorems in Section 3.2 continue to hold under these assumptions.

Theorem 3.6 *Suppose that Assumptions D1-D4, D5(i)-(iii), and D6(i) and (iii) in Appendix D hold. Then*

$$\begin{aligned}(i) \sup_{A \in \mathcal{A}_K} \|H^{-1}[\hat{\Theta}_\tau(A) - \Theta_\tau(A)]\| &= O_P([K \ln(NT)/(NT)]^{1/2}), \\ (ii) H^{-1}[\hat{\Theta}_\tau(A) - \Theta_\tau(A)] &= (NT)^{-1} [H_1 H^{-1} \Phi_K(A) H_1^{-1} H]^{-1} \bar{H}_1 \sum_{i=1}^N \sum_{t=1}^T g_{it}(A, \Theta_\tau(A)) + o_P((NT)^{-1/2}) \\ &+ r_{NT} \text{ uniformly in } A \in \mathcal{A}_K, \\ \text{where } \bar{H}_1 = H_1 H^{-1} H_1, \text{ and } \|r_{NT}\| &= O_P(\zeta_K^{1/2} K^{5/4} (NT)^{-3/4} \ln(NT)).\end{aligned}$$

Theorem 3.7 *Suppose that Assumptions D1-D6 in Appendix D hold. Then as $(N, T) \rightarrow \infty$,*

$$\{\tau(1-\tau)\Pi(u)\Omega_\tau\Psi_K\Omega'_\tau\Pi(u)'\}^{-1/2}\sqrt{NT}\begin{pmatrix}\hat{\alpha}_\tau(u) - \alpha_\tau(u) \\ \hat{\beta}_{1\tau}(u) - \beta_{1\tau}(u)\end{pmatrix} \xrightarrow{d} N(0, I_{k_1+k_2}),$$

where $\Pi(u)$ is a $(k_1 + k_2) \times (k_1 + k_2) K$ matrix defined in (3.1).

For the proofs of the above theorems, see Appendix D. As demonstrated in Appendix D, the proofs of the above results are quite involved. The complications arise for several reasons. First, the fixed effects parameter $\beta_{2\tau}$ is of dimension N , which generally diverges to infinity at a much faster rate than K . Second, the estimator of $\hat{\beta}_{2\tau}$ has a convergence rate (in Frobenius norm) that is different from that of $\hat{B}_{1\tau}$ and \hat{C}_τ , which explains the need for the normalization matrix H defined above. Third, without using the second normalization matrix H_1 in the definitions of $J_{K,A}(A)$ and $\Phi_K(A)$, the latter matrices would be degenerate asymptotically. See, e.g., the proof of Lemma D.1 for the case of $\Phi_K(A)$.

4 A specification test

In this section, we consider testing the hypothesis that some of the functional coefficients are constant. The test can be applied to any nonempty subset of the full set of functional coefficients.

4.1 Hypotheses and test statistic

Let \mathbb{S} be an $r \times (k_1 + k_2)$ matrix that selects r elements from $\delta_\tau(u) = (\alpha_\tau(u)', \beta_\tau(u)')'$, where $1 \leq r \leq k_1 + k_2$. For example, if $\mathbb{S} = (I_{k_1}, 0_{k_1 \times k_2})$, then $\mathbb{S}\delta_\tau(u) = \alpha_\tau(u)$; if $\mathbb{S} = (0_{k_2 \times k_1}, I_{k_2})$, then $\mathbb{S}\delta_\tau(u) = \beta_\tau(u)$;

and if $\mathbb{S} = I_{k_1+k_2}$, then $\mathbb{S}\delta_\tau(u) = \delta_\tau(u)$. We are interested in testing the null hypothesis

$$\mathbb{H}_0 : \delta_{1\tau}(U_i) \equiv \mathbb{S}\hat{\delta}_\tau(U_i) = \delta_{1\tau} \text{ a.s. for some parameter } \delta_{1\tau} \in \mathbb{R}^r. \quad (4.1)$$

The alternative hypothesis \mathbb{H}_1 is the negation of \mathbb{H}_0 . That is, under \mathbb{H}_0 , r of the $(k_1 + k_2)$ functional coefficients are constant, whereas under \mathbb{H}_1 , at least one of the functional coefficients in $\delta_{1\tau}(\cdot)$ is not constant.

In principle one can consider various ways to test the null hypothesis in (4.1). For example, one can estimate the restricted semiparametric functional coefficient IVQR model under the null, and construct a Lagrangian multiplier (LM) type of test based on the estimation of the restricted model only. Alternatively, one can adopt the likelihood ratio (LR) principle to estimate both the unrestricted and restricted models and construct various test statistics, say, by comparing the estimates of either $\delta_{1\tau}(\cdot)$ or $\delta_\tau(\cdot)$ in both models through certain distance measure, or by extending the generalized likelihood ratio (GLR) test of Fan et al. (2001) to our IV quantile regression framework. Both the LM and LR types of tests require estimation under the null and one needs to estimate the restricted model multiple times in order to test for multiple null hypotheses for different subsets of functional coefficients.

In this paper, we propose a Wald-type statistic that requires only consistent estimation of the unrestricted model. Let $\hat{\delta}_{1\tau}(U_i) = \mathbb{S}\hat{\delta}_\tau(U_i)$ and $\bar{\delta}_{1\tau} = \frac{1}{n} \sum_{i=1}^n \hat{\delta}_{1\tau}(U_i)$. We propose the following test statistic

$$T_n = \sum_{i=1}^n \left\| \hat{\delta}_{1\tau}(U_i) - \bar{\delta}_{1\tau} \right\|^2 a(U_i), \quad (4.2)$$

where $a(\cdot)$ is a uniformly bounded nonnegative weight function defined on the support \mathcal{U} of U_i . Our theory allows one to take $a(u) = 1$ for all $u \in \mathcal{U}$, in which case one obtains an unweighted version of the test. By specifying a weight function that is positive only in a subset of \mathcal{U} , one may focus the test on a specific region of \mathcal{U} in applications. In the next subsection, we show that after being suitably normalized, T_n is asymptotically distributed as $N(0, 1)$ under \mathbb{H}_0 and diverges to infinity under \mathbb{H}_1 .

4.2 Asymptotic distribution of the test statistic

To proceed, we first consider the consistent estimation of $\delta_{1\tau}$ under \mathbb{H}_0 . We estimate it by

$$\bar{\delta}_{1\tau} = \frac{1}{n} \sum_{i=1}^n \hat{\delta}_{1\tau}(U_i). \quad (4.3)$$

Let $\bar{\Pi} \equiv E[\Pi(U_1)]$ and $\Sigma_{\delta_{1\tau}}^K \equiv \tau(1-\tau) \mathbb{S}\bar{\Pi}\Omega_\tau\Psi_K\Omega_\tau'\bar{\Pi}'\mathbb{S}'$. We make the following additional assumptions.

Assumption A6*. As $n \rightarrow \infty$, $\zeta_K^2 K^3 (\ln n)^2 / n \rightarrow 0$ and $nK^{-2\lambda/d} \rightarrow 0$.

Assumption A7. (i) $0 < \underline{c}_{\bar{\Pi}} \leq \lambda_{\min}(\bar{\Pi}\bar{\Pi}') \leq \lambda_{\max}(\bar{\Pi}\bar{\Pi}') \leq \bar{c}_{\bar{\Pi}} < \infty$ for each K .

(ii) $0 < \underline{c}_{\Pi} \leq \lambda_{\min}(E[\Pi(U_i)'\Pi(U_i)a(U_i)]) \leq \lambda_{\max}(E[\Pi(U_i)'\Pi(U_i)a(U_i)]) \leq \bar{c}_{\Pi} < \infty$ for each K .

Assumption A6* strengthens Assumption A6(i). The second requirement in A6* ensures that the asymptotic bias term of $\bar{\delta}_{1\tau}$ under \mathbb{H}_0 is $o(n^{-1/2})$ so that it has an asymptotically negligible effect on the asymptotic distribution of $\bar{\delta}_{1\tau}$. A7 requires that $\bar{\Pi}$ and $E[\Pi(U_i)'\Pi(U_i)a(U_i)]$ be full rank. In conjunction with Assumptions A5(ii) and (iv), it also ensures that the minimum and maximum eigenvalues of $\Sigma_{\delta_{1\tau}}^K$

are bounded and bounded away from 0 with probability approaching 1 (w.p.a.1.) in the case $a(U_i) = 1$ a.s..

The following theorem establishes the \sqrt{n} -consistency and asymptotic normality of $\bar{\delta}_{1\tau}$ under \mathbb{H}_0 .

Theorem 4.1 *Suppose Assumptions A1-A5, A6*, and A7(i) hold. Suppose that $\Sigma_{\bar{\delta}_{1\tau}} \equiv \lim_{K \rightarrow \infty} \Sigma_{\bar{\delta}_{1\tau}}^K$ exists. Then under \mathbb{H}_0 , $\sqrt{n}(\bar{\delta}_{1\tau} - \delta_{1\tau}) \xrightarrow{d} N(0_{r \times 1}, \Sigma_{\bar{\delta}_{1\tau}})$.*

Remark 6. Clearly Theorem 4.1 says that under \mathbb{H}_0 , $\bar{\delta}_{1\tau}$ can consistently estimate $\delta_{1\tau}$ at the parametric rate. The second requirement in A6* indicates that one needs to select a larger number of sieve approximation terms than usual in order to achieve this rate. On the other hand, if $r = k_1 + k_2$, i.e., all functional coefficients take constant values under \mathbb{H}_0 , and the sieve basis includes the constant term, then the bias term from the sieve approximation vanishes automatically and the second requirement in Assumption A6* becomes redundant. In this case, a small value of K can be selected.

Let $\Upsilon_i \equiv \Pi(U_i)' S' S \Pi(U_i) a(U_i)$ and $\bar{\Omega}_\tau \equiv \Omega'_\tau E(\Upsilon_1) \Omega_\tau$. Define

$$\mathbb{B}_n \equiv \tau(1-\tau) \text{tr} \left(\Omega'_\tau E(\Upsilon_1) \Omega_\tau \hat{\Psi}_K \right) \text{ and } \sigma_n^2 \equiv 2\tau^2(1-\tau)^2 \text{tr} \left(\bar{\Omega}_\tau \Psi_K \bar{\Omega}_\tau \Psi_K \right). \quad (4.4)$$

To state the next result, we modify Assumption A6* as follows.

Assumption A6.** As $n \rightarrow \infty$, $\zeta_K^2 K^3 (\ln n)^2 / n \rightarrow 0$ and $nK^{-(1/2+2\lambda/d)} \rightarrow 0$.

Intuitively, one does not need the constrained estimator $\bar{\delta}_{1\tau}$ to be \sqrt{n} -consistent to derive the asymptotic distribution for our test statistic. Assumption A6** is sufficient to ensure that the bias term from the sieve approximation plays an asymptotically negligible role in the asymptotic distribution.

The next theorem studies the asymptotic distribution of T_n under \mathbb{H}_0 .

Theorem 4.2 *Suppose Assumptions A1-A5, A6**, and A7 hold. Then under \mathbb{H}_0 , $\sigma_n^{-1}(T_n - \mathbb{B}_n) \xrightarrow{d} N(0, 1)$.*

Remark 7. The above theorem also holds if one replaces $\hat{\Psi}_K$ in the definition of \mathbb{B}_n by its population analogue Ψ_K . We use $\hat{\Psi}_K$ because \mathbb{B}_n appears as a term in the decomposition of T_n .

To implement the test, we need consistent estimates of both \mathbb{B}_n and σ_n^2 . Let $\hat{\varepsilon}_i \equiv \hat{\varepsilon}_{i\tau} \equiv Y_i - \hat{\alpha}_\tau(U_i)' D_i - \hat{\beta}_\tau(U_i)' X_i$, $\hat{\Upsilon} = \frac{1}{n} \sum_{i=1}^n \Pi(U_i)' S' S \Pi(U_i) a(U_i)$, $\hat{\Omega}_\tau = (\hat{\Omega}'_{A_\tau}, \hat{\Omega}'_{B_\tau})'$, where

$$\begin{aligned} \hat{\Omega}_{A_\tau} &\equiv - \left(\hat{J}'_{K,A} \hat{\Phi}'_{K,C} M_K \hat{\Phi}_{K,C} \hat{J}_{K,A} \right)^{-1} \hat{J}_{K,A} \hat{\Phi}'_{K,C} M_K \hat{\Phi}_{K,C}, \\ \hat{\Omega}_{B_\tau} &\equiv \hat{\Phi}_{K,B} \left[I_{(k_2+k_3)K} + \hat{\Omega}_{A_\tau} \right], \end{aligned}$$

and $\left[\hat{\Phi}'_{K,B} \quad \hat{\Phi}'_{K,C} \right]'$ is a conformable partition of $\hat{\Phi}_K^{-1}$ with $\hat{\Phi}_{K,B}$ and $\hat{\Phi}_{K,C}$ being $k_2 K \times (k_2 + k_3) K$ and $k_3 K \times (k_2 + k_3) K$ matrices, respectively. We propose to estimate \mathbb{B}_n and σ_n^2 respectively by

$$\hat{\mathbb{B}}_n \equiv \tau(1-\tau) \text{tr} \left(\hat{\Omega}'_\tau \hat{\Upsilon} \hat{\Omega}_\tau \hat{\Psi}_K \right) \text{ and } \hat{\sigma}_n^2 \equiv 2\tau^2(1-\tau)^2 \text{tr} \left(\hat{\Omega}'_\tau \hat{\Upsilon} \hat{\Omega}_\tau \hat{\Psi}_K \hat{\Omega}'_\tau \hat{\Upsilon} \hat{\Omega}_\tau \hat{\Psi}_K \right).$$

In the proof of Theorem 4.3 below, we show that $\sigma_n^{-1}(\hat{\mathbb{B}}_n - \mathbb{B}_n) = o_P(1)$ and $\sigma_n^{-1}(\hat{\sigma}_n - \sigma_n) = o_P(1)$ under the following additional assumption.

Assumption A8. As $n \rightarrow \infty$, $h \rightarrow 0$ and $K^2/(nh) \rightarrow 0$.

Then we have

$$\hat{T}_n \equiv \hat{\sigma}_n^{-1} \left(T_n - \hat{\mathbb{B}}_n \right) \xrightarrow{d} N(0, 1) \text{ under } \mathbb{H}_0. \quad (4.5)$$

When n is sufficiently large, we can compare the feasible test statistic \hat{T}_n to the one-sided critical value z_α , the upper α percentile from the standard normal distribution, and reject the null at asymptotic level α if $\hat{T}_n > z_\alpha$.

To examine the asymptotic local power, we consider the following sequence of Pitman local alternatives

$$\mathbb{H}_1 \left(\sigma_n^{1/2} n^{-1/2} \right) : \delta_{1\tau}(U_i) = \delta_{1\tau} + \sigma_n^{1/2} n^{-1/2} \Delta_n(U_i) \text{ a.s.}$$

where Δ_n 's are a sequence of real continuous vector-valued functions such that $\mu_0 \equiv \lim_{n \rightarrow \infty} E[\|\Delta_n(U_i) - E[\Delta_n(U_i)]\|^2 a(U_i)] < \infty$. The following theorem establishes the asymptotic local power of the \hat{T}_n test.

Theorem 4.3 *Suppose Assumptions A1-A5, A6**, and A7-A8 hold. Then under $\mathbb{H}_1(\sigma_n^{1/2} n^{-1/2})$, $\hat{T}_n \xrightarrow{d} N(\mu_0, 1)$.*

Remark 8. Theorem 4.3 shows that the \hat{T}_n test has nontrivial power against Pitman local alternatives that converge to zero at rate $n^{-1/2} K^{1/4}$ because $\sigma_n \propto K^{1/2}$ as demonstrated in the proof of the above theorem. The asymptotic local power function is given by $\lim_{n \rightarrow \infty} P(\hat{T}_n \geq z \mid \mathbb{H}_1(\sigma_n^{1/2} n^{-1/2})) = 1 - \Phi(z - \mu_0)$, where Φ is the standard normal CDF.

The next theorem establishes the consistency of the test.

Theorem 4.4 *Suppose Assumptions A1-A5, A6**, and A7-A8 hold. Then under \mathbb{H}_1 , $n^{-1} \sigma_n \hat{T}_n = \mu_A + o_P(1)$ where $\mu_A \equiv E[\|\delta_{1\tau}(U_i) - E[\delta_{1\tau}(U_i)]\|^2 a(U_i)]$, so that $P(\hat{T}_n > c_n) \rightarrow 1$ under \mathbb{H}_1 for any nonstochastic sequence $c_n = o(n/\sigma_n)$.*

Remark 9. In the above study we restrict our attention to the case where the weight matrix M_K used in (2.10) is nonrandom. If efficiency is also of concern, we can consider efficient choice of M_K . As we have seen from Remark 4, an optimal choice of M_K for the efficient estimation of the structural functional coefficient $\alpha_\tau(u)$ is given by $\Sigma_{C_\tau}^{-1}$. But this choice of M_K may not be optimal for the testing problem on hand. Despite the importance of optimal test, a formal study is highly complicated and beyond the scope of the current paper. Therefore we leave it for future research.

Remark 10. If we fail to reject \mathbb{H}_0 in (4.1), one may consider more efficient estimation of the null-restricted model. The simplest approach is to impose the null restriction and estimate both the finite dimensional coefficient parameter ($\delta_{1\tau}$) and the functional coefficients (if any) in a single step. One can readily establish the convergence rates and asymptotic normality for the estimates of both the parametric and nonparametric components and show the resulting estimates of the functional coefficients are more efficient than those obtained under the alternative. Alternatively, one can follow the above procedure to first estimate the unrestricted model and then to obtain the estimate of $\delta_{1\tau}$ by $\bar{\delta}_{1\tau}$. If there are remaining functional coefficients to be estimated, one can estimate them by substituting $\delta_{1\tau}$ by $\bar{\delta}_{1\tau}$ in the original SQF under \mathbb{H}_0 and treating it as if it were known. In the special case where we fail to reject $\mathbb{H}_0 : \alpha_\tau(U_i) = \alpha_\tau$ a.s. for some parameter $\alpha_\tau \in \mathbb{R}^{k_1}$, after obtaining $\bar{\alpha}_\tau \equiv n^{-1} \sum_{i=1}^n \hat{\alpha}_\tau(U_i)$, in the second step we can estimate the functional coefficient $\beta_\tau(u)$ by considering the ordinary functional coefficient quantile regression (FCQR) of $Y_i - \bar{\alpha}_\tau' D_i$ on X_i . Similar approach is also taken by Cai and Xiao (2012) in their kernel estimation of partially linear FCQR models without endogeneity. To conserve space, we do not report the asymptotic properties of these estimates.

4.3 A bootstrap version of our test

It is well known that a nonparametric test based on its asymptotic normal null distribution may perform poorly in finite samples. So we suggest using a bootstrap method to obtain the bootstrap approximation to the finite-sample distribution of our test statistic under the null. Härdle and Mammen (1993) show that a two-point wild bootstrap is valid in the context of nonparametric specification tests for conditional mean models. A similar procedure has been extended to the time series framework (e.g., Hansen (2000) and Su and White (2010)) or functional coefficient IV regression (e.g., Su et al. (2014)). As emphasized in the literature, the great advantage of this method lies in the fact that there is no need to mimic some important features (such as dependence or endogeneity structure) in the data generating process in order to justify its asymptotic validity.

Nevertheless, as Sun (2006) observes, the commonly used wild bootstrap fails in the quantile regression (QR) framework where the quantile error terms do not satisfy the zero mean assumption. This motivates her to propose a modified version of the wild bootstrap procedure for the QR framework without endogeneity. More recently, Feng et al. (2011) propose a modification of the wild bootstrap that admits a broader class of weight distributions for quantile regressions. Here we follow the latter paper and propose to generate the bootstrap version of \hat{T}_n as follows:

1. Obtain the sieve IVQR estimates $\hat{\alpha}_\tau(U_i)$ and $\hat{\beta}_\tau(U_i)$, and calculate the unrestricted residuals $\hat{\varepsilon}_i = Y_i - \hat{\alpha}_\tau(U_i)' D_i - \hat{\beta}_\tau(U_i)' X_i$.
2. For $i = 1, \dots, n$, generate the wild bootstrap residuals $\varepsilon_i^* = |\hat{\varepsilon}_i| e_i$, where e_i 's are independent of $\mathcal{D}_n \equiv \{Y_j, U_j, D_j, W_j\}_{j=1}^n$ and are IID taking values $2(1 - \tau)$ and -2τ with probabilities $1 - \tau$ and τ , respectively.
3. For $i = 1, \dots, n$, generate $Y_i^* = \hat{\alpha}'_\tau D_i + \hat{\beta}'_\tau X_i + \varepsilon_i^*$ where $\hat{\alpha}_\tau \equiv n^{-1} \sum_{i=1}^n \hat{\alpha}_\tau(U_i)$ and $\hat{\beta}_\tau \equiv n^{-1} \sum_{i=1}^n \hat{\beta}_\tau(U_i)$ are the restricted IVQR estimates under the null hypothesis $\mathbb{H}_{0s} : \delta_\tau(U_i) = \delta_\tau$ a.s. for some parameter $\delta_\tau \in \mathbb{R}^{k_1+k_2}$.
4. Redo the sieve IVQR estimation and compute the bootstrap test statistic \hat{T}_n^* in the same way as \hat{T}_n by using $\{Y_i^*, U_i, D_i, W_i\}_{i=1}^n$.
5. Repeat Steps 1-4 B times to obtain B bootstrap test statistic $\{\hat{T}_{nj}^*\}_{j=1}^B$. Calculate the bootstrap p -values $p^* \equiv B^{-1} \sum_{j=1}^B 1\{\hat{T}_{nj}^* \geq \hat{T}_n\}$ and reject the null hypothesis $\mathbb{H}_0 : \delta_{1\tau}(U_i) = \delta_{1\tau}$ a.s. if p^* is smaller than the prescribed nominal level of significance.

We make several remarks regarding the above bootstrap procedure. First, in sharp contrast with the original wild bootstrap method that uses the residuals $\hat{\varepsilon}_i$, we use the absolute residuals in Step 2. By construction, the τ th quantile of e_i is zero, which ensures that τ th conditional quantile of ε_i^* is zero given the data \mathcal{D}_n . One can replace the two-point distribution of e_i by some other distribution that has the τ th conditional quantile given by 0. Second, note that in Step 3 we impose the null hypothesis $\mathbb{H}_{0s} : \delta_\tau(U_i) = \delta_\tau$ a.s., which is stronger than $\mathbb{H}_0 : \delta_{1\tau}(U_i) = \delta_{1\tau}$ a.s. unless $\mathbb{S} = I_{k_1+k_2}$ (i.e., $r = k_1 + k_2$). It turns out that this will greatly facilitate the justification of the asymptotic validity of the above bootstrap procedure. In addition, it saves in computation when we try to test many subvectors of $\delta_\tau(\cdot)$ are constant or not because we can generate the same bootstrap dependent variable once for all and the

computation burden is almost identical to the case of testing the constancy of a single subvector of $\delta_\tau(\cdot)$. Our simulations indicate that this procedure does not result in the loss of power in comparison with the alternative approach by generating Y_i^* through the imposition of the original null hypothesis \mathbb{H}_0 . But the justification for the validity of this latter approach would be much more involved as one cannot ensure that the estimated functional coefficients satisfy the required smoothness conditions.

To show that the bootstrap statistic \hat{T}_n^* can be used to approximate the asymptotic null distribution of \hat{T}_n , we follow Li et al. (2003) and Su et al. (2014) and rely on the notion of *convergence in distribution in probability* defined in Giné and Zinn (1990). The following theorem establishes the asymptotic validity of the above bootstrap procedure.

Theorem 4.5 *Suppose Assumptions A1-A5, A6**, and A7-A8 hold. Let z_α^* be the α -level bootstrap critical value based on $B \rightarrow \infty$ bootstrap resamples. Then (i) \hat{T}_n^* converges to $N(0, 1)$ in distribution in probability, (ii) $\lim_{n \rightarrow \infty} P(\hat{T}_n \geq z_\alpha^*) = \alpha$ under \mathbb{H}_0 , (iii) $\lim_{n \rightarrow \infty} P(\hat{T}_n \geq z_\alpha^*) \rightarrow 1 - \Phi(z_\alpha - \mu_A)$ under $\mathbb{H}_1(\sigma_n^{1/2}n^{-1/2})$, and (iv) $\lim_{n \rightarrow \infty} P(\hat{T}_n \geq z_\alpha^*) = 1$ under \mathbb{H}_1 , where z_α denotes the $100(1 - \alpha)$ th percentile of the standard normal distribution.*

Remark 11. Theorem 4.5 shows that the QR wild bootstrap provides an asymptotic valid approximation to the limit null distribution of \hat{T}_n because the null hypothesis is always satisfied in the bootstrap resamples. If the null hypothesis does not hold in the original sample \mathcal{D}_n , then \hat{T}_n explodes at the rate n/σ_n but \hat{T}_n^* is still well behaved. This intuitively explains the consistency of the bootstrap-based test \hat{T}_n^* .

Remark 12. As a referee kindly points out, we can improve the speed of the wild bootstrap by considering the score-based approach to wild bootstrap as advocated by Kline and Santos (2012, KS hereafter). KS proposes a generalization of the wild bootstrap based upon perturbing the scores of M-estimators and avoids recomputing the M-estimator in each bootstrap iteration, which saves in computation time greatly. They study test statistics J_n that are quadratic forms in a vector-valued underlying statistic $S_n : J_n = S_n' S_n$. Under the null hypothesis, S_n is required to be *asymptotically pivotal* and exhibit a linear expansion. But this is not the case for our test statistic \hat{T}_n . Despite this, we can follow the spirit of KS and the idea of weighted bootstrap in the statistics literature and propose a bootstrap procedure that does not require parameter estimation in each bootstrap iteration. In fact, under \mathbb{H}_0 we show in the proof of Theorem 4.3 that

$$\sigma_n^{-1}(T_n - \mathbb{B}_n) = \sigma_n^{-1} \frac{2}{n} \sum_{1 \leq j < k \leq n} \varphi_n(\zeta_j, \zeta_k) + o_P(1),$$

where $\varphi_n(\zeta_j, \zeta_k) = \psi_\tau(\varepsilon_j) \phi_{W_j}^K(U_j)' \bar{\Omega}_\tau \phi_{W_k}^K(U_k) \psi_\tau(\varepsilon_k)$ and $\zeta_i = (U_i', W_i', \varepsilon_i)'$. That is, the dominant term in T_n after bias correction is a second order degenerate U -statistic with kernel given by $\varphi_n(\cdot, \cdot)$. We can perturb a feasible version of the dominant term when constructing an alternative bootstrap statistic. Let $\hat{\zeta}_i = (U_i', W_i', \hat{\varepsilon}_i)'$. We consider the following bootstrap statistic:

$$\hat{T}_n^{**} = \sigma_n^{*-1} \frac{2}{n} \sum_{1 \leq j < k \leq n} \hat{\varphi}_n(\hat{\zeta}_j, \hat{\zeta}_k) w_j w_k,$$

where $\hat{\varphi}_n(\hat{\zeta}_j, \hat{\zeta}_k) = \psi_\tau(\hat{\varepsilon}_j) \phi_{W_j}^K(U_j)' \hat{\Omega}_\tau' \hat{\Upsilon} \hat{\Omega}_\tau \phi_{W_k}^K(U_k) \psi_\tau(\hat{\varepsilon}_k)$, $\sigma_n^{**2} = 2\tau^2(1 - \tau)^2 \text{tr}(\hat{\Omega}_\tau' \hat{\Upsilon} \hat{\Omega}_\tau \hat{\Psi}_K^{**} \hat{\Omega}_\tau' \hat{\Upsilon} \hat{\Omega}_\tau \hat{\Psi}_K^{**})$, $\hat{\Psi}_K^{**} \equiv \frac{1}{n\tau(1-\tau)} \sum_{j=1}^n \phi_{W_j}^K(U_j) \phi_{W_j}^K(U_j)' \psi_\tau^2(\hat{\varepsilon}_j)$, and $\{w_i, i = 1, \dots, n\}$ is an IID sequence that is independent of the data and has mean zero, variance one, and finite fourth moment μ_4 . We show in Appendix E

that \hat{T}_n^{**} converges to $N(0, 1)$ in distribution in probability and thus can be used to obtain the bootstrap p -value.

5 Monte Carlo simulations

5.1 Evaluation of the sieve IVQR estimates

In this subsection, we examine the finite sample performance of the sieve IVQR estimator. We consider four different data generating processes (DGPs) for Monte Carlo experiments.

DGP 1 corresponds to a location-scale model where the regression coefficients are independent of quantiles:

$$\text{DGP 1: } \begin{cases} Y = D\alpha(U) + X\beta(U) + \sigma(U)\Phi^{-1}(V) \\ D = [Z + \rho\Phi^{-1}(V)] / \sqrt{1 + \rho^2} \end{cases}$$

where $\alpha_\tau(U) \equiv \alpha(U) = 1 + \sin(1.5U)$, $\beta_\tau(U) \equiv \beta(U) = 2\Phi(U)$, $U \sim \text{Uniform}(-1, 1)$, $Z \sim N(2, 1)$, $X \sim N(0, 1)$, $V \sim \text{Uniform}(0, 1)$, $\sigma(U) = (1 + 0.5U^2)\exp(-U^2)$, V , U , X , and Z are mutually independent, and ρ is a parameter that controls the degree of endogeneity. Apparently, a larger value of $|\rho|$ indicates a stronger degree of endogeneity. For this DGP, the SQF is written as $s(u, d, x) = \sigma(u)\Phi^{-1}(\tau) + d\alpha(u) + x\beta(u)$.

DGP 2 is the same as DGP 1 except that $D = [0.2Z + \rho\Phi^{-1}(V)] / \sqrt{1 + \rho^2}$. In comparison with DGP 1, the instrument in DGP 2 is quite weak. Thus, we can check how our estimation performs in the presence of weak instrument. Alternatively, one can also consider situations where the correlation between the instrument and endogenous regressor decreases to zero as the sample size increases.

DGP 3 considers a random coefficient structural model where the regression coefficients vary with not only the exogenous variable U but also an unobserved uniform random variable V :

$$\text{DGP 3: } \begin{cases} Y = D\alpha(U, V) + X\beta(U, V) \\ D = [Z + \rho\Phi^{-1}(V)] / \sqrt{1 + \rho^2} \end{cases}$$

where $\alpha_\tau(U) \equiv \alpha(U, V)|_{V=\tau} = 1 + \sin(1.5U) + \ln(2 + U)\Phi^{-1}(\tau)$, $\beta_\tau(U) \equiv \beta(U, V)|_{V=\tau} = 2\Phi(U) + \exp(-U^2)\Phi^{-1}(\tau)$, $X \sim N(2, 1)$, and the remaining components are the same as those in DGP 1. It is possible to see this model as a location-scale model which is more general than DGP 1, in a sense that the scale function $\sigma(\cdot)$ takes a functional coefficient form $\sigma(U, D, X) = D \ln(2 + U) + X \exp(-U^2)$, which depends not only on U but also on (D, X) . The SQF of this DGP is simply $s(u, d, x) = d\alpha_\tau(u) + x\beta_\tau(u)$.

Finally, DGP 4 considers a location-scale model which is similar to DGP 1, but it differs in that $\alpha(\cdot)$ is a function of U_1 and $\beta(\cdot)$ is a function of U_2 , $U = (U_1, U_2)$, where $U_1, U_2 \sim \text{Uniform}(-1, 1)$ and are mutually independent, and $\sigma(U) = 0.2 + 0.5(\cos(U_1) + \exp(U_2/3))$. The remaining components including the shapes of the functions $\alpha(\cdot)$ and $\beta(\cdot)$ are the same as those in DGP 1. The SQF of this DGP can be written as $s(u, d, x) = \sigma(u)\Phi^{-1}(\tau) + d\alpha(u_1) + x\beta(u_2)$. To simplify the computation, we only consider the case where researchers know that the scale function $\sigma(u)$ is additively separable with respect to u_1 and u_2 (c.f., Horowitz and Lee (2005)). It should be stressed that, when one uses a kernel-based estimator, it is often computationally tedious to estimate functional coefficient models with different smoothing variables. On the other hand, our sieve estimator can be easily applied to the estimation of such models.

For each DGP, we consider three sample sizes: $n = 200, 400, \text{ and } 800$. As for the choice of the sieve space, we use the cubic B-spline basis functions (see, e.g., Schumaker (2007)). Let the number of internal

knots be $\lfloor cn^{1/5} \rfloor$, where $\lfloor a \rfloor$ denotes the integer part of a , and c is a positive number. Then the number of sieve approximation terms is given by $K = 4 + \lfloor cn^{1/5} \rfloor$. To evaluate how our IVQR estimator is sensitive to the choice of K , we consider three different values of c , namely, $c \in \{1, 1.5, 2\}$. The simulations are coded in **R**. The first stage quantile regression is implemented by the function **rq** in the package **quantreg**. For the degree of endogeneity, we consider three values for ρ , namely, $\rho \in \{0.2, 0.5, 0.8\}$. For τ , we try two quantile values, 0.5 (median) and 0.9. Thus we have 216 simulation set-ups in total (four DGPs, three sample sizes, three choices of K , three ρ 's, and two τ 's). The number of Monte Carlo repetitions for each scenario is set to be 1000. The estimated functional coefficients are evaluated by the mean absolute deviation (MAD) statistic:

$$\text{MAD}(\hat{\alpha}_\tau^{(r)}(\cdot)) = \frac{1}{n} \sum_{i=1}^n \left| \hat{\alpha}_\tau^{(r)}(U_i^{(r)}) - \alpha_\tau(U_i^{(r)}) \right|,$$

and the mean squared error (MSE) statistic:

$$\text{MSE}(\hat{\alpha}_\tau^{(r)}(\cdot)) = \frac{1}{n} \sum_{i=1}^n \left[\hat{\alpha}_\tau^{(r)}(U_i^{(r)}) - \alpha_\tau(U_i^{(r)}) \right]^2,$$

where $U_i^{(r)}$ means the i th draw of U in the r th replicated data set and $\hat{\alpha}_\tau^{(r)}(\cdot)$ is the IVQR estimate for $\alpha_\tau(\cdot)$ obtained from the r th replicated data set. $\text{MAD}(\hat{\beta}_\tau^{(r)}(\cdot))$ and $\text{MSE}(\hat{\beta}_\tau^{(r)}(\cdot))$ are computed in the same manner as above. In DGPs 1, 2 and 4, we need to estimate not only $\alpha_\tau(\cdot)$ and $\beta_\tau(\cdot)$ but also the functional *intercept* term $\sigma(\cdot)\Phi^{-1}(\tau)$, but its estimation results are omitted to save space.

Tables 1 and 2 report the average MAD values and the average MSE values over 1000 replications for $\tau = 0.5$ and $\tau = 0.9$, respectively. We summarize some important findings from Tables 1-2. First, in terms of the choice of K , we observe that estimates with $c = 1$ always outperform those with $c = 1.5$ and $c = 2$. Recall that the estimation bias becomes smaller and the variance becomes larger as c or equivalently K increases. Therefore, for the DGPs under investigation it seems that bias is of not a big concern and a small value of sieve approximating terms could do a good job in terms of bias reduction. Second, as expected, the estimation becomes more accurate so that MAD and MSE decrease quickly as the sample size increases. This is true for all DGPs and all choices of ρ , K and τ . Third, it becomes hard to estimate the functional coefficient $\alpha_\tau(\cdot)$ of the endogenous regressor as the degree of endogeneity increases. This phenomenon becomes even more transparent when the sample size is small and the instrument is weak as in DGP 2. These results support the common knowledge that the availability of a large data set and strong instruments is crucial for obtaining accurate estimates.⁴ On the other hand, the estimation of the functional coefficient $\beta_\tau(\cdot)$ of the exogenous regressor appears more or less independent of the degree of endogeneity. Fourth, the results for DGP 4 show that our sieve IVQR estimator works well for the cases where the functional coefficients have different smoothing variables. Finally, comparing the results for DGP 1, 2 and 4 in Table 1 with those in Table 2 suggests that the parameters in the conditional median regression can be estimated more precisely than those in the high-quantile ($\tau = 0.9$) regression. Note that we can make such a comparison here because in these DGPs the functional coefficients do not vary over the quantile index τ , which is not the case for DGP 3. This result is reasonable because the

⁴Chernozhukov and Hansen (2008) propose an inference procedure for an instrumental variable quantile regression which is robust to weak and partial identification. It seems possible to extend their approach to FCQR models, which is a topic for future research.

conditional density of the error term at $\tau = 0.9$ is lower than that at the median in our simulation set-ups so estimates at high quantiles are expected to have larger variance than those for the median regression.

5.2 Tests for the constancy of functional coefficients

We next examine the finite sample performance of the proposed test. Two DGPs for Monte Carlo experiments are considered, which are, respectively, modifications of DGPs 1 and 3 from the previous subsection:

$$\text{DGP } 1': \begin{cases} Y = D\alpha(U) + X\beta(U) + \sigma(U)\Phi^{-1}(V) \\ D = [Z + \rho\Phi^{-1}(V)] / \sqrt{1 + \rho^2}, \end{cases}$$

where $\alpha(U) = 1 + \sin(1.5\Delta_0 U)$, and $\beta(U) = 2\Phi(\Delta_0 U)$.

$$\text{DGP } 3': \begin{cases} Y = D\alpha(U, V) + X\beta(U, V) \\ D = [Z + \rho\Phi^{-1}(V)] / \sqrt{1 + \rho^2} \end{cases}$$

where $\alpha_\tau(U) = 1 + \sin(1.5\Delta_0 U) + \ln(2 + \Delta_0 U)\Phi^{-1}(\tau)$, and $\beta_\tau(U) = 2\Phi(\Delta_0 U) + \exp(-\Delta_0 U^2)\Phi^{-1}(\tau)$. In both DGPs V , U , X and Z are generated as before. We consider the following three null hypotheses:

$$\mathbb{H}_{0,\alpha} : \alpha_\tau(U) \text{ is constant with respect to } U,$$

$$\mathbb{H}_{0,\beta} : \beta_\tau(U) \text{ is constant with respect to } U,$$

$$\mathbb{H}_{0,\alpha\beta} : \alpha_\tau(U) \text{ and } \beta_\tau(U) \text{ are both constant with respect to } U.$$

When $\Delta_0 = 1$, DGPs 1' and 3' reduce to DGPs 1 and 3, respectively, so that neither functional coefficient is constant and we shall examine the power behavior of our test. When $\Delta_0 = 0$, both functional coefficients become a constant and we shall examine the size behavior of our test.

For each DGP, we consider three sample sizes: $n = 200, 400, \text{ and } 800$. For the choice of sieve space, we use the cubic B-spline basis functions with the number of internal knots being $\lfloor n^{1/5} \rfloor$. We consider three values for ρ (0.2, 0.5 and 0.8) and two values of Δ_0 (0 and 1), and fix τ to be 0.5. The number of Monte Carlo repetitions and bootstrap resamples for each set-up are set to be 500 and 200, respectively.

Table 3 report the results for our \hat{T}_n^* -based bootstrap test. We summarize some important findings from Table 3. First, when the sample size is 800, the size of our test is well controlled in both DGP 1' and DGP 3' despite some small variation, for all values of ρ , and all three null hypotheses under investigation. The degree of endogeneity has some effect on the size behavior. Second, our test tends to be oversized for small sample sizes (but the size distortion is quickly corrected as the sample size increases, as described just above). Third, in terms of power, our test has good power property in both DGPs. In particular, as the sample size increases, the empirical power also increases, as expected. Another noteworthy phenomenon is that the increase in the degree of endogeneity tends to decrease the power of the test in DGP 3' for $\mathbb{H}_{0,\alpha}$. We also implement the \hat{T}_n^{**} -based bootstrap test and find it is severely undersized for both DGPs under consideration but has power comparable to that of the \hat{T}_n^* -based bootstrap test. See Tables A.1 and A.2 in Supplementary Appendix F for details.

Table 1: Finite sample performance of our IVQR estimator ($\tau = 0.5$)

DGP	n	ρ	MAD						MSE					
			$c = 1$		$c = 1.5$		$c = 2$		$c = 1$		$c = 1.5$		$c = 2$	
			$\alpha_\tau(\cdot)$	$\beta_\tau(\cdot)$	$\alpha_\tau(\cdot)$	$\beta_\tau(\cdot)$	$\alpha_\tau(\cdot)$	$\beta_\tau(\cdot)$	$\alpha_\tau(\cdot)$	$\beta_\tau(\cdot)$	$\alpha_\tau(\cdot)$	$\beta_\tau(\cdot)$	$\alpha_\tau(\cdot)$	$\beta_\tau(\cdot)$
1	200	0.2	0.147	0.144	0.175	0.173	0.189	0.183	0.036	0.034	0.052	0.051	0.063	0.058
		0.5	0.166	0.148	0.200	0.175	0.218	0.189	0.049	0.037	0.077	0.054	0.100	0.064
		0.8	0.198	0.148	0.245	0.184	0.273	0.198	0.087	0.040	0.168	0.075	0.269	0.097
	400	0.2	0.111	0.108	0.119	0.118	0.135	0.134	0.020	0.019	0.023	0.023	0.030	0.029
		0.5	0.121	0.107	0.133	0.117	0.151	0.134	0.025	0.019	0.030	0.023	0.041	0.030
		0.8	0.144	0.108	0.156	0.117	0.179	0.135	0.039	0.020	0.048	0.023	0.069	0.033
	800	0.2	0.075	0.073	0.086	0.084	0.097	0.094	0.009	0.009	0.012	0.011	0.015	0.014
		0.5	0.083	0.074	0.095	0.085	0.106	0.095	0.011	0.009	0.015	0.012	0.019	0.015
		0.8	0.096	0.074	0.111	0.085	0.125	0.096	0.016	0.009	0.021	0.012	0.028	0.015
2	200	0.2	0.809	0.151	1.005	0.187	1.094	0.199	1.498	0.041	2.785	0.073	3.089	0.083
		0.5	1.031	0.167	1.194	0.191	1.245	0.205	2.546	0.066	3.268	0.096	3.330	0.117
		0.8	1.211	0.160	1.332	0.187	1.472	0.208	3.922	0.080	4.519	0.132	5.938	0.191
	400	0.2	0.586	0.111	0.637	0.121	0.735	0.140	0.651	0.021	0.795	0.025	1.117	0.034
		0.5	0.761	0.122	0.835	0.133	0.953	0.151	1.387	0.031	1.697	0.038	2.118	0.051
		0.8	0.924	0.117	0.984	0.127	1.070	0.144	1.960	0.036	2.151	0.043	2.510	0.062
	800	0.2	0.387	0.074	0.447	0.086	0.508	0.096	0.267	0.009	0.367	0.012	0.482	0.016
		0.5	0.479	0.079	0.574	0.094	0.654	0.105	0.520	0.011	0.822	0.017	1.037	0.022
		0.8	0.612	0.084	0.712	0.096	0.789	0.107	0.870	0.015	1.165	0.021	1.416	0.027
3	200	0.2	0.304	0.308	0.362	0.368	0.391	0.398	0.174	0.170	0.250	0.247	0.298	0.293
		0.5	0.317	0.300	0.382	0.357	0.415	0.385	0.199	0.162	0.307	0.246	0.360	0.286
		0.8	0.339	0.290	0.413	0.351	0.451	0.379	0.237	0.161	0.399	0.259	0.497	0.326
	400	0.2	0.227	0.230	0.246	0.251	0.282	0.288	0.095	0.092	0.111	0.110	0.148	0.146
		0.5	0.229	0.218	0.251	0.238	0.286	0.273	0.100	0.085	0.119	0.102	0.157	0.136
		0.8	0.247	0.210	0.269	0.229	0.311	0.261	0.124	0.081	0.146	0.097	0.200	0.130
	800	0.2	0.151	0.151	0.176	0.178	0.199	0.200	0.042	0.040	0.058	0.055	0.074	0.070
		0.5	0.156	0.149	0.181	0.172	0.202	0.192	0.045	0.038	0.061	0.052	0.076	0.065
		0.8	0.164	0.139	0.189	0.161	0.215	0.182	0.051	0.034	0.068	0.046	0.089	0.059
4	200	0.2	0.203	0.202	0.243	0.241	0.261	0.262	0.071	0.073	0.103	0.106	0.120	0.125
		0.5	0.226	0.207	0.275	0.242	0.295	0.260	0.095	0.076	0.152	0.107	0.186	0.127
		0.8	0.272	0.206	0.339	0.246	0.364	0.263	0.157	0.075	0.291	0.110	0.340	0.130
	400	0.2	0.149	0.148	0.161	0.160	0.184	0.181	0.038	0.039	0.044	0.046	0.057	0.058
		0.5	0.167	0.150	0.179	0.163	0.206	0.185	0.049	0.040	0.057	0.047	0.077	0.061
		0.8	0.195	0.150	0.213	0.162	0.244	0.183	0.074	0.040	0.092	0.047	0.124	0.060
	800	0.2	0.104	0.103	0.118	0.119	0.133	0.131	0.018	0.018	0.023	0.024	0.029	0.030
		0.5	0.114	0.103	0.130	0.118	0.147	0.132	0.022	0.019	0.029	0.024	0.037	0.030
		0.8	0.133	0.105	0.152	0.120	0.172	0.133	0.031	0.019	0.041	0.025	0.054	0.030

Table 2: Finite sample performance of our IVQR estimator ($\tau = 0.9$)

DGP	n	ρ	MAD						MSE					
			$c = 1$		$c = 1.5$		$c = 2$		$c = 1$		$c = 1.5$		$c = 2$	
			$\alpha_\tau(\cdot)$	$\beta_\tau(\cdot)$	$\alpha_\tau(\cdot)$	$\beta_\tau(\cdot)$	$\alpha_\tau(\cdot)$	$\beta_\tau(\cdot)$	$\alpha_\tau(\cdot)$	$\beta_\tau(\cdot)$	$\alpha_\tau(\cdot)$	$\beta_\tau(\cdot)$	$\alpha_\tau(\cdot)$	$\beta_\tau(\cdot)$
1	200	0.2	0.201	0.200	0.242	0.236	0.262	0.254	0.070	0.068	0.105	0.098	0.125	0.114
		0.5	0.235	0.204	0.290	0.249	0.314	0.268	0.113	0.074	0.201	0.120	0.340	0.146
		0.8	0.291	0.210	0.371	0.262	0.407	0.287	0.231	0.092	0.434	0.166	0.579	0.199
	400	0.2	0.152	0.149	0.164	0.161	0.187	0.183	0.038	0.036	0.045	0.043	0.058	0.056
		0.5	0.168	0.149	0.184	0.164	0.208	0.184	0.050	0.037	0.063	0.045	0.081	0.058
		0.8	0.204	0.151	0.225	0.166	0.258	0.190	0.088	0.041	0.115	0.051	0.151	0.068
	800	0.2	0.103	0.102	0.119	0.117	0.134	0.131	0.017	0.017	0.023	0.022	0.029	0.028
		0.5	0.114	0.102	0.133	0.117	0.149	0.131	0.022	0.017	0.030	0.023	0.038	0.028
		0.8	0.133	0.104	0.156	0.121	0.178	0.136	0.031	0.018	0.046	0.024	0.062	0.031
2	200	0.2	1.174	0.218	1.485	0.266	1.598	0.288	3.510	0.091	5.943	0.154	6.988	0.188
		0.5	1.400	0.222	1.588	0.262	1.625	0.273	4.593	0.156	5.298	0.233	5.555	0.250
		0.8	1.567	0.255	1.791	0.312	1.848	0.341	6.441	0.276	8.666	0.485	10.803	0.859
	400	0.2	0.856	0.157	0.925	0.172	1.079	0.198	1.830	0.045	2.004	0.054	2.738	0.075
		0.5	1.110	0.172	1.176	0.182	1.268	0.201	2.797	0.071	3.016	0.076	3.378	0.100
		0.8	1.277	0.174	1.356	0.190	1.478	0.219	3.326	0.097	3.959	0.121	4.798	0.181
	800	0.2	0.544	0.105	0.638	0.121	0.727	0.137	0.554	0.019	0.822	0.025	1.080	0.032
		0.5	0.735	0.115	0.860	0.132	0.964	0.147	1.281	0.027	1.745	0.036	2.079	0.046
		0.8	0.896	0.117	1.024	0.135	1.133	0.150	1.697	0.034	2.131	0.048	2.599	0.061
3	200	0.2	0.409	0.460	0.490	0.545	0.527	0.588	0.325	0.392	0.476	0.557	0.556	0.654
		0.5	0.431	0.501	0.523	0.602	0.564	0.648	0.371	0.487	0.580	0.752	0.689	0.876
		0.8	0.476	0.529	0.576	0.638	0.625	0.682	0.481	0.553	0.831	0.950	0.953	1.095
	400	0.2	0.306	0.342	0.330	0.370	0.382	0.425	0.179	0.211	0.208	0.250	0.281	0.332
		0.5	0.320	0.374	0.345	0.404	0.392	0.455	0.205	0.268	0.243	0.316	0.311	0.398
		0.8	0.346	0.387	0.381	0.420	0.428	0.475	0.239	0.273	0.294	0.328	0.383	0.430
	800	0.2	0.208	0.229	0.242	0.268	0.272	0.302	0.082	0.093	0.111	0.127	0.140	0.162
		0.5	0.214	0.250	0.246	0.289	0.277	0.324	0.087	0.110	0.114	0.147	0.145	0.187
		0.8	0.231	0.263	0.268	0.303	0.303	0.341	0.104	0.122	0.140	0.161	0.179	0.207
4	200	0.2	0.280	0.284	0.334	0.335	0.362	0.365	0.140	0.145	0.202	0.214	0.242	0.256
		0.5	0.320	0.282	0.388	0.337	0.424	0.373	0.209	0.145	0.333	0.216	0.429	0.263
		0.8	0.404	0.290	0.500	0.343	0.554	0.374	0.398	0.155	0.681	0.223	1.027	0.270
	400	0.2	0.209	0.210	0.226	0.227	0.256	0.258	0.075	0.078	0.089	0.091	0.117	0.117
		0.5	0.236	0.212	0.254	0.229	0.292	0.259	0.103	0.079	0.122	0.093	0.185	0.120
		0.8	0.282	0.208	0.306	0.226	0.360	0.259	0.187	0.077	0.215	0.091	0.314	0.121
	800	0.2	0.144	0.145	0.165	0.166	0.186	0.186	0.034	0.037	0.046	0.048	0.058	0.060
		0.5	0.158	0.142	0.184	0.162	0.207	0.182	0.043	0.035	0.058	0.046	0.076	0.057
		0.8	0.185	0.141	0.218	0.162	0.247	0.183	0.063	0.035	0.091	0.046	0.123	0.058

Table 3: Finite sample rejection frequency of \hat{T}_n^* -based bootstrap test

DGP	Δ_0	ρ	n	$\mathbb{H}_{0,\alpha}$			$\mathbb{H}_{0,\beta}$			$\mathbb{H}_{0,\alpha\beta}$		
				1%	5%	10%	1%	5%	10%	1%	5%	10%
1'	0	0.2	200	0.020	0.072	0.138	0.030	0.064	0.116	0.028	0.066	0.122
			400	0.018	0.068	0.136	0.016	0.084	0.144	0.024	0.078	0.152
			800	0.010	0.056	0.102	0.014	0.050	0.118	0.014	0.068	0.118
		0.5	200	0.020	0.084	0.150	0.042	0.104	0.172	0.038	0.106	0.190
			400	0.008	0.060	0.106	0.010	0.064	0.128	0.006	0.064	0.130
			800	0.010	0.048	0.118	0.028	0.064	0.132	0.020	0.062	0.134
		0.8	200	0.014	0.080	0.136	0.040	0.086	0.164	0.032	0.086	0.152
			400	0.014	0.050	0.108	0.006	0.072	0.144	0.002	0.070	0.150
			800	0.014	0.048	0.108	0.012	0.048	0.110	0.018	0.060	0.104
	1	0.2	200	0.998	0.998	0.998	0.776	0.898	0.948	0.980	0.992	0.996
			400	1.000	1.000	1.000	0.974	0.986	0.992	1.000	1.000	1.000
			800	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.5	200	1.000	1.000	1.000	0.810	0.936	0.970	0.984	0.998	1.000
			400	1.000	1.000	1.000	0.972	0.992	1.000	1.000	1.000	1.000
			800	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.8	200	0.998	0.998	0.998	0.786	0.918	0.956	0.980	0.990	0.994
			400	1.000	1.000	1.000	0.976	0.994	0.996	1.000	1.000	1.000
			800	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3'	0	0.2	200	0.020	0.082	0.150	0.022	0.088	0.154	0.024	0.100	0.146
			400	0.010	0.046	0.108	0.024	0.084	0.134	0.026	0.060	0.114
			800	0.006	0.056	0.112	0.014	0.062	0.116	0.006	0.062	0.110
		0.5	200	0.026	0.078	0.142	0.022	0.066	0.132	0.024	0.068	0.118
			400	0.020	0.066	0.116	0.012	0.064	0.124	0.020	0.060	0.120
			800	0.012	0.054	0.112	0.010	0.060	0.112	0.014	0.060	0.114
		0.8	200	0.016	0.072	0.128	0.018	0.086	0.146	0.014	0.074	0.136
			400	0.008	0.064	0.120	0.018	0.060	0.132	0.006	0.072	0.132
			800	0.004	0.044	0.098	0.016	0.054	0.096	0.010	0.046	0.106
	1	0.2	200	0.460	0.676	0.770	0.266	0.332	0.488	0.370	0.624	0.746
			400	0.726	0.888	0.926	0.324	0.594	0.748	0.674	0.866	0.934
			800	0.968	0.986	0.992	0.820	0.942	0.968	0.968	0.990	0.992
		0.5	200	0.424	0.616	0.738	0.286	0.420	0.552	0.374	0.646	0.780
			400	0.704	0.852	0.904	0.408	0.658	0.786	0.694	0.864	0.910
			800	0.940	0.986	0.992	0.826	0.952	0.978	0.952	0.990	0.998
		0.8	200	0.358	0.556	0.664	0.262	0.406	0.552	0.312	0.578	0.710
			400	0.688	0.834	0.886	0.422	0.728	0.832	0.686	0.856	0.928
			800	0.940	0.976	0.990	0.904	0.976	0.988	0.954	0.988	0.994

6 An empirical application

In this section, we present an empirical application of estimating quantile Engel curves for food. The analysis is performed on a pooled U.K. Family Expenditure Survey (FES) data from 1994 to 1996. There are a number of studies that estimate Engel curves using the FES data; see Banks et al. (1997), Blundell et al. (1998), Blundell et al. (2003), Blundell et al. (2007), and Chen and Pouzo (2009, 2012), among others. Although most of these studies have used a (parametric or non/semi-parametric) mean regression approach, by applying a quantile regression approach to the estimation of Engel curves, we can account for unobserved taste heterogeneity in households’ consumption as in Chen and Pouzo (2009, 2012). It is also important in empirical Engel curve analysis to account for observable household demographics in a way consistent with consumer optimization theory; see, e.g., Blundell et al. (1998), Blundell et al. (2003) and Blundell et al. (2007) for details. For categorical demographics, a straightforward approach is to split the data into subsamples according to the categories, and estimate Engel curves within each subsample. Similarly, we may use a “localization” approach in the case of continuous demographic variables such as the age of household children.⁵ Then, in order to preserve a degree of demographic homogeneity, we select from the FES data during 1994 to 1996 a subsample of coupled households with one child who own a car. As in Blundell et al. (2003), the selection of households with cars has a role to include motoring expenditures and petrol as commodity consumption. In addition, in order to lower the risk of misreporting bias, we exclude observations with the share of food expenditure being zero, and with the household income being less than 100 GBP. Then, after excluding these observations, the analysis is performed on a sample of 1672 households. Using this dataset, we estimate the following model by the proposed IVQR estimator:

$$\text{Food Share} = \alpha_{\tau}(\text{Child's Age}) \ln(\text{Total Expenditure}) + \beta_{\tau}(\text{Child's Age}) + \varepsilon.$$

Following the literature, we treat $\ln(\text{Total Expenditure})$ as an endogenous variable and employ $\ln(\text{Household Income})$ as an instrument for it.⁶ The descriptive statistics for each variable are summarized in Table 4.

Table 4: Descriptive statistics for the FES dataset ($n = 1672$)

Variable	Mean	Median	Std. Dev.	Min	Max
Food Share	0.266	0.263	0.088	0.015	0.710
Child’s Age	7.583	6.500	6.253	0	18
Total Expenditure (GBP)	293.969	255.126	174.236	57.211	2380.785
Household Income (GBP)	435.096	389.629	261.561	101.070	5875.380

⁵When the dimension of demographic variables is not small, using the extended partially linear model introduced by Blundell et al. (1998) can be more attractive than the localization approach, in the sense of alleviating the notorious “curse of dimensionality” associated with a pure nonparametric model.

⁶It should be noted that, as pointed out by the aforementioned authors, the assumption of linear Engel curves with respect to log-expenditure is very restrictive in general. However, at least for the Engel curve for food, it is empirically known that it is well approximated by a function linear in log-expenditure (see, e.g., Banks et al. (1997) and Blundell et al. (1998)). Thus, for illustrative purposes, we employ this simple linear specification. When estimating Engel curves for the other categories of goods, it is desirable to include additional higher order total expenditure variables in the estimation. Of course, the IVQR estimator is available for estimating such models. Another important aspect of this analysis is that, since we use a pooled data over 1994-96, the effects from changes in relative prices between different time periods are averaged over the three years.

For the model estimation, we use the cubic B-spline basis functions with the number of internal knots being 4 ($\lfloor 1672^{1/5} \rfloor = 4$). The sieve IVQR estimates of $\alpha_\tau(\cdot)$ at $\tau = 0.1, 0.3, 0.5, 0.7$ and 0.9 are reported in Figure 1. The 95% confidence intervals are calculated by using the wild bootstrap with 500 resamples based on Feng et al. (2011). For comparison, the parametric IVQR estimate of α_τ is also reported in the figure, based on the assumption that neither $\alpha_\tau(\cdot)$ nor $\beta_\tau(\cdot)$ varies with child's age but $\ln(\text{Total Expenditure})$ is endogenous. The parametric IVQR estimations are implemented by using the method of Chernozhukov and Hansen (2006), and their results for α_τ and β_τ are also summarized in Table 5. From Figure 1, we can confirm that the effects of total expenditure amount on the food share vary over both the proportion of food expenditure and the age of household child. As expected, the sign of the total expenditure term is negative for the all quantiles and the whole domain of child's age (except for a very small region where $\tau = 0.9$ and Child's Age = 0). Note, however, that the estimates may not be statistically significantly different from zero for some part of the domain. In particular, the estimate of $\alpha_\tau(\cdot)$ at $\tau = 0.1$ has relatively wide confidence interval compared to the other quantiles. On average, we observe that the magnitude of $\alpha_\tau(\cdot)$ becomes larger as τ increases. Therefore, we can conclude that on the whole the amount of total expenditure becomes important on the food share for those households who allocate a large proportion of their budget to foods. As shown in Table 5, the parametric IVQR estimates also show the same decreasing tendency of α_τ in τ . Based on these estimation results, we can estimate the quantile Engel curve for food at each τ . However, note that the monotonicity of conditional quantile function with respect to τ is not automatically satisfied with the IVQR estimation procedure. Thus, we have used the rearrangement method proposed by Chernozhukov et al. (2010) at each (Child's Age, $\ln(\text{Total Expenditure})$) based on 499 quantile indices: $\{0.002, 0.004, \dots, 0.998\}$. Figure 2 presents the estimated quantile Engel curve at $\tau = 0.1, 0.3, 0.5, 0.7$ and 0.9 . As a consequence of the rearrangement, the estimated Engel curves are not necessary continuous. From Figure 2, we can confirm that when the age of household child is one, the share of total expenditure spent on food is relatively low to the other stages of age for the all quantiles. When the age of household child becomes five and ten, the two estimated Engel curves take similar form.

Table 5: Parametric instrumental variable quantile regression estimates

	$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$
Estimate of α_τ	-0.0428	-0.0534	-0.0540	-0.0670	-0.0819
(t -statistic)	(-5.5214)	(-5.7578)	(-6.3160)	(-10.3387)	(-7.8687)
Estimate of β_τ	0.4058	0.5186	0.5645	0.6777	0.8184
(t -statistic)	(9.2330)	(9.8398)	(11.7411)	(18.8748)	(14.3012)

Now we consider the test of constancy of the functional coefficients. We consider the following three null hypotheses:

$\mathbb{H}_{0\alpha} : \alpha_\tau(\cdot)$ is constant with respect to Child's Age

$\mathbb{H}_{0\beta} : \beta_\tau(\cdot)$ is constant with respect to Child's Age

$\mathbb{H}_{0\alpha\beta} : \alpha_\tau(\cdot)$ and $\beta_\tau(\cdot)$ are both constant with respect to Child's Age.

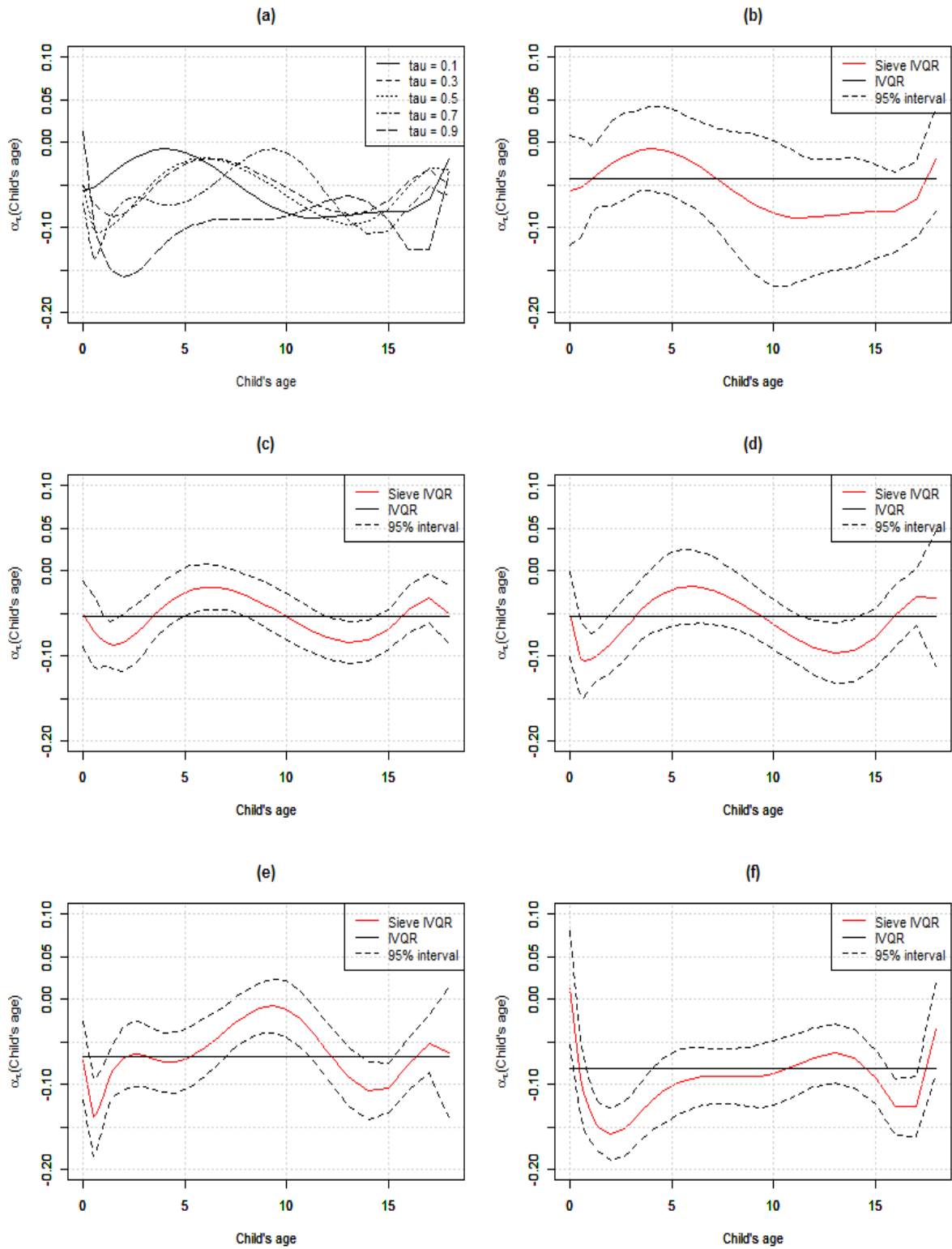


Figure 1: Plots of estimated $\alpha_\tau(\text{Child's Age})$: (a) estimated $\alpha_\tau(\text{Child's Age})$ at $\tau = 0.1, 0.3, 0.5, 0.7,$ and 0.9 ; (b) $\tau = 0.1$; (c) $\tau = 0.3$; (d) $\tau = 0.5$; (e) $\tau = 0.7$; (f) $\tau = 0.9$.

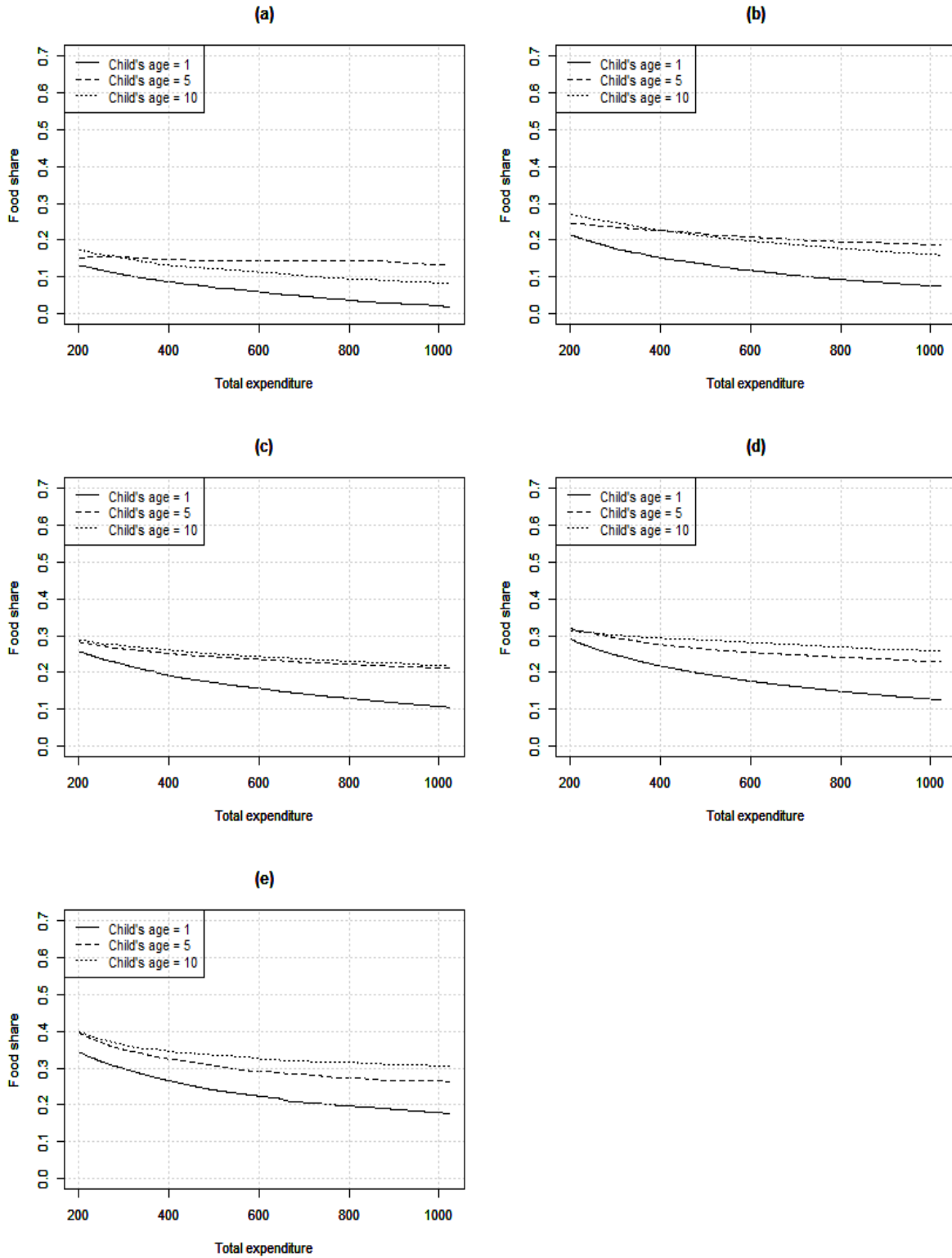


Figure 2: Estimated Engel curves: (a) $\tau = 0.1$; (b) $\tau = 0.3$; (c) $\tau = 0.5$; (d) $\tau = 0.7$; (e) $\tau = 0.9$.

We implement the test by following the same test procedure used in the simulations. In particular, we set the weighting function $a(\text{Child's Age}) = 1$ uniformly in Child's Age. Table 6 reports the bootstrap p -values for the above three null hypotheses where the number of bootstrap resamples is 500. We summarize some interesting findings from Table 6. First, the results clearly indicate that the effects of total expenditure on the food share are significantly heterogeneous with respect to Child's Age at the middle and higher quantiles. When $\tau = 0.7$ and 0.9 , all of the three null hypotheses are rejected at 5% significance level. When $\tau = 0.5$, we can reject $\mathbb{H}_{0\alpha}$ at 5% significance level but not $\mathbb{H}_{0\beta}$ and $\mathbb{H}_{0\alpha\beta}$. Second, we fail to reject the constancy of $\alpha_\tau(\cdot)$ (and similarly $\beta_\tau(\cdot)$) at $\tau = 0.1$ and 0.3 for any reasonable level of significance. These results indicate that, for those (probably rich) households who only need to spend a small proportion of their budget on foods, the growth of their child does not much affect the share of food expenditure. For $\tau = 0.1$, the high p -value could be partially due to the large variance of the estimate as suggested in Figure 1.

Table 6: p -values for our nonparametric tests

Null hypotheses	$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$
$\mathbb{H}_{0\alpha}$: $\alpha_\tau(\cdot)$ is constant	0.468	0.114	0.024	0.024	0.004
$\mathbb{H}_{0\beta}$: $\beta_\tau(\cdot)$ is constant	0.434	0.196	0.484	0.020	0.002
$\mathbb{H}_{0\alpha\beta}$: $\alpha_\tau(\cdot)$ and $\beta_\tau(\cdot)$ are both constant	0.428	0.194	0.462	0.020	0.002

7 Conclusion

In this paper we consider sieve IVQR estimation of functional coefficient models with endogeneity. We establish the uniform consistency and asymptotic normality of the estimators, based on which we also propose a nonparametric specification test for the constancy of the functional coefficients and establish its asymptotic properties. Monte Carlo simulations indicate that our estimator and test perform well in finite samples. An application to the estimation of quantile Engel curves indicates the usefulness of our model, estimator, and test.

Several extensions are possible. First, as an alternative, one may consider kernel estimation of functional coefficient models with endogeneity. Even though kernel method is based on local approximation of unknown functions and is computationally expensive, it is interesting to study the asymptotic properties of kernel estimates for our model. Second, one may propose a different test for the constancy of functional coefficients. For example, one may consider an LM type of test by estimating the model under the null hypothesis and basing a test statistic on the estimated score function. This approach surely has its advantage for nonparametric sieve estimation, but it may result in much greater computational burden if one uses kernel estimation unless one wants to test the constancy of all functional coefficients. Third, endogeneity may be present in other types of nonparametric or semiparametric quantile regression models. It is also interesting to broaden the research scope of the current paper to a more general structural model. We leave these as future research topics.

Appendix

A Proof of the results in section 3

Proof of Theorem 3.1. By Assumptions A1(ii) and A2(i)-(ii), $\lim_{K \rightarrow \infty} P(Y \leq P_K(D, U)'A_\tau + P_K(X, U)'B_\tau | U, X, Z) = \tau$ a.s. Then by Theorem 3 of Chernozhukov and Hansen (2006), (A_τ, B_τ, C_τ) is identified for $K \rightarrow \infty$ under our Assumptions A1-A4, which leads to the identification of $(\alpha_\tau(u), \beta_\tau(u), \gamma_\tau(u))$ by Assumption A2(ii). ■

To prove Theorem 3.2, we first state some technical lemmas whose proofs are provided in the supplementary appendix.

Lemma A.1 *Suppose the conditions of Theorem 3.2 hold. Then*

- (i) $\sup_{A \in \mathcal{A}_K} \|\Phi_{n,K}(A) - \Phi_K(A)\| = O_P(\zeta_K K^{1/2} (\ln n/n)^{1/2}) = o_P(1)$.
- (ii) $\underline{c}_\Phi/2 \leq \inf_{A \in \mathcal{A}_K} \lambda_{\min}(\Phi_{n,K}(A)) \leq \sup_{A \in \mathcal{A}_K} \lambda_{\max}(\Phi_{n,K}(A)) \leq 2\bar{c}_\Phi$ w.p.a.1.
- (iii) $\|\hat{\Psi}_K - \Psi_K\| = O_P(\zeta_K K^{1/2}/n^{1/2}) = o_P(1)$.
- (iv) $\underline{c}_\Psi/2 \leq \lambda_{\min}(\hat{\Psi}_K) \leq \lambda_{\max}(\hat{\Psi}_K) \leq 2\bar{c}_\Psi$ w.p.a.1.

Lemma A.2 *Suppose the conditions of Theorem 3.2 hold. Then $\sup_{A \in \mathcal{A}_K} \left\| n^{-1} \sum_{i=1}^n g_i(A, \hat{\Theta}_\tau(A)) \right\| = o_P(n^{-1/2})$.*

Lemma A.3 *Suppose the conditions of Theorem 3.2 hold. Then $\sup_{A \in \mathcal{A}_K} \left\| n^{-1} \sum_{i=1}^n g_i(A, \Theta_\tau(A)) \right\| = O_P[(K \ln n/n)^{1/2}]$.*

Lemma A.4 *Suppose the conditions of Theorem 3.2 hold. Then for any constant $L > 0$,*

$$\sup_{\|\mathbf{c}\|=1} \sup_{A \in \mathcal{A}_K} \sup_{\|\Theta - \Theta_\tau(A)\| \leq L(K \ln n/n)^{1/2}} \left| n^{-1} \mathbf{c}' \sum_{i=1}^n E[g_i(A, \Theta) - g_i(A, \Theta_\tau(A))] + \mathbf{c}' \Phi_K(A) [\Theta - \Theta_\tau(A)] \right| = o_P(n^{-1/2}).$$

Lemma A.5 *Let $\eta_i(A; \Theta_1, \Theta_2) \equiv g_i(A, \Theta_1) - g_i(A, \Theta_2) - E[g_i(A, \Theta_1)] + E[g_i(A, \Theta_2)]$. Suppose the conditions of Theorem 3.2 hold. Then*

- (i) $E[\sup_{A \in \mathcal{A}_K} \sup_{\|\Theta_1 - \Theta_2\| \leq \Delta} \|\eta_i(A; \Theta_1, \Theta_2)\|^2] \leq n^{1/2} \Delta$ for sufficiently large n ;
- (ii) for any constant $L > 0$ and $\mathbf{c} \in \mathbb{R}^{(k_2+k_3)K}$ with $\|\mathbf{c}\| = 1$, $E[\sup_{A \in \mathcal{A}_K} \sup_{\|\Theta - \Theta_\tau(A)\| \leq L(K \ln n/n)^{1/2}} \sum_{i=1}^n |\mathbf{c}' \eta_i(A; \Theta, \Theta_\tau(A))|^2] = O(V(n, K))$ where $V(n, K) = n \zeta_K K (K \ln n/n)^{1/2}$.

Lemma A.6 *Suppose the conditions of Theorem 3.2 hold. Then for any $\mathbf{c} \in \mathbb{R}^{(k_2+k_3)K}$ with $\|\mathbf{c}\| \leq 1$ and $r > 0$, we have*

$$\sup_{\|(A, \Theta_1)\| \leq n^r, \|(A, \Theta_2)\| \leq n^r} \frac{\sum_{i=1}^n \|\mathbf{c}' \eta_i(A; \Theta_1, \Theta_2)\|}{t_{1n}(A; \Theta_1, \Theta_2) + t_{2n}(A; \Theta_1, \Theta_2) + n^{-2}} = O_P\left((K \ln n)^{1/2}\right),$$

where $t_{1n}(A; \Theta_1, \Theta_2) = \left\{ \sum_{i=1}^n E \|\mathbf{c}' \eta_i(A; \Theta_1, \Theta_2)\|^2 \right\}^{1/2}$ and $t_{2n}(A; \Theta_1, \Theta_2) = \left\{ \sum_{i=1}^n \|\mathbf{c}' \eta_i(A; \Theta_1, \Theta_2)\|^2 \right\}^{1/2}$.

Proof of Theorem 3.2. To prove (i), we extend the proof of Theorem 2.1 in He and Shao (2000) from a pointwise result to a uniform one. By the convexity of the objective function it suffices to show that for any $\epsilon > 0$, there exists a large constant $L \equiv L(\epsilon)$ which does not depend on $A \in \mathcal{A}_K$ such that $P(\inf_{\|\mathbf{c}\|=1} - \sum_{i=1}^n \mathbf{c}' g_i(A, \Theta_{\tau, \mathbf{c}}(A)) > 0 \text{ for all } A \in \mathcal{A}_K) > 1 - \epsilon$ for sufficiently large n , where $\Theta_{\tau, \mathbf{c}}(A) =$

$\Theta_\tau(A) + L(K \ln n/n)^{1/2} \mathbf{c}$.⁷ By Lemma A.6, $\sum_{i=1}^n \mathbf{c}' \eta_i(A; \Theta_{\tau, \mathbf{c}}(A), \Theta_\tau(A)) = O_P[(K \ln n)^{1/2}] \{T_{1n}(A) + T_{2n}(A) + n^{-2}\}$ uniformly in $A \in \mathcal{A}_K$, where $T_{1n}(A) = \{\sum_{i=1}^n E[\mathbf{c}' \eta_i(A; \Theta_{\tau, \mathbf{c}}(A), \Theta_\tau(A))]^2\}^{1/2}$ and $T_{2n}(A) = \{\sum_{i=1}^n |\mathbf{c}' \eta_i(A; \Theta_{\tau, \mathbf{c}}(A), \Theta_\tau(A))|^2\}^{1/2}$. By Lemma A.5(ii), we have that uniformly in $A \in \mathcal{A}_K$, $T_{1n}(A) = O(V(n, K)^{1/2})$ and $T_{2n}(A) = O_P(V(n, K)^{1/2})$ by Jensen and Markov inequalities, respectively. In addition, $V(n, K)K \ln n = nK \ln n[\zeta_K K^{3/2}(\ln n/n)^{1/2}] = o(nK \ln n)$ under Assumption A6(i). These results, in conjunction with Lemma A.4 implies that uniformly in $A \in \mathcal{A}_K$

$$\begin{aligned} \sum_{i=1}^n \mathbf{c}' g_i(A, \Theta_{\tau, \mathbf{c}}(A)) &= \sum_{i=1}^n \mathbf{c}' g_i(A, \Theta_\tau(A)) + \sum_{i=1}^n \mathbf{c}' E[g_i(A, \Theta_{\tau, \mathbf{c}}(A)) - g_i(A, \Theta_\tau(A))] \\ &\quad + [T_{1n}(A) + T_{2n}(A) + n^{-2}] O_P[(K \ln n)^{1/2}] \\ &= \sum_{i=1}^n \mathbf{c}' g_i(A, \Theta_\tau(A)) - L(nK \ln n)^{1/2} \mathbf{c}' \Phi_K(A) \mathbf{c} + o_P((nK \ln n)^{1/2}). \end{aligned}$$

Consequently, for sufficiently large n we have

$$\begin{aligned} &P \left(\inf_{\|\mathbf{c}\|=1} - \sum_{i=1}^n \mathbf{c}' g_i(A, \Theta_{\tau, \mathbf{c}}(A)) > 0 \text{ for all } A \in \mathcal{A}_K \right) \\ &\geq P \left((nK \ln n)^{-1/2} \inf_{\|\mathbf{c}\|=1} - \sum_{i=1}^n \mathbf{c}' g_i(A, \Theta_\tau(A)) > -\frac{L}{2} \mathbf{c}' \Phi_K(A) \mathbf{c} \text{ for all } A \in \mathcal{A}_K \right) \\ &\geq P \left((nK \ln n)^{-1/2} \inf_{\|\mathbf{c}\|=1} - \sum_{i=1}^n \mathbf{c}' g_i(A, \Theta_\tau(A)) > -\frac{L}{2} \underline{c}_\Phi \text{ for all } A \in \mathcal{A}_K \right) \\ &= P \left((nK \ln n)^{-1/2} \sup_{A \in \mathcal{A}_K} \sup_{\|\mathbf{c}\|=1} \sum_{i=1}^n \mathbf{c}' g_i(A, \Theta_\tau(A)) < \frac{L}{2} \underline{c}_\Phi \right) \rightarrow 1 \text{ as } L, n \rightarrow \infty, \end{aligned}$$

where the last line follows by Lemma A.3.

For part (ii), we apply Lemma A.6 with $(\Theta_1, \Theta_2) = (\hat{\Theta}_\tau(A), \Theta_\tau(A))$ to obtain $\sum_{i=1}^n \mathbf{c}' \eta_i(A; \hat{\Theta}_\tau(A), \Theta_\tau(A)) = (\hat{t}_n(A) + n^{-2}) O_P[(K \ln n)^{1/2}]$, where $\hat{t}_n(A) = O_P(V(n, K)^{1/2})$ by (i) and Lemma A.5(ii). Then by Lemmas A.2 and A.4, we have

$$\mathbf{c}' \Phi_K(A) [\hat{\Theta}_\tau(A) - \Theta_\tau(A)] = n^{-1} \mathbf{c}' \sum_{i=1}^n g_i(A, \Theta_\tau(A)) + O_P \left(n^{-1} [V(n, K)K \ln n]^{1/2} \right) + o_P \left(n^{-1/2} \right)$$

uniformly in $A \in \mathcal{A}_K$ and \mathbf{c} with $\|\mathbf{c}\| = 1$. It follows that $\hat{\Theta}_\tau(A) - \Theta_\tau(A) = n^{-1} \Phi_K(A)^{-1} \sum_{i=1}^n g_i(A, \Theta_\tau(A)) + o_P(n^{-1/2}) + r_n$, where $\|r_n\| = O_P \left(n^{-1} [V(n, K)K \ln n]^{1/2} \right) = O_P(\zeta_K^{1/2} K^{5/4} n^{-3/4} \ln n)$. ■

Proof of Theorem 3.3. We first prove part (i). Recall $\mathbf{Q}_K(A) = C_\tau(A)' M_K C_\tau(A)$. Let $\hat{\mathbf{Q}}_n(A) \equiv \hat{C}_\tau(A)' M_K \hat{C}_\tau(A)$. By the triangle inequality,

$$\begin{aligned} &\sup_{A \in \mathcal{A}_K} \left| \hat{C}_\tau(A)' M_K \hat{C}_\tau(A) - C_\tau(A)' M_K C_\tau(A) \right| \\ &\leq \sup_{A \in \mathcal{A}_K} \left| [\hat{C}_\tau(A) - C_\tau(A)]' M_K [\hat{C}_\tau(A) - C_\tau(A)] \right| + 2 \sup_{A \in \mathcal{A}_K} \left| C_\tau(A)' M_K [\hat{C}_\tau(A) - C_\tau(A)] \right| \\ &\equiv D_{1n} + D_{2n}, \text{ say.} \end{aligned}$$

⁷Note that $-\sum_{i=1}^n g_i(A, \Theta_{\tau, \mathbf{c}}(A))$ corresponds to $\sum_{i=1}^n \psi(x_i, \cdot)$ in He and Shao (2000) which is the directional derivative of the objective function in the direction \mathbf{c} defined in Koenker (2005, p.33).

$D_{1n} \leq \lambda_{\max}(M_K) \sup_{A \in \mathcal{A}_K} \|\hat{C}_\tau(A) - C_\tau(A)\|^2 = O_P(K \ln n/n) = o_P(1)$ by Theorem 3.2(i) and Assumptions A5(i) and A6. By the matrix Cauchy-Schwarz inequality and the same assumptions,

$$\begin{aligned} D_{2n} &\leq \sup_{A \in \mathcal{A}_K} \{C_\tau(A)' M_K C_\tau(A)\}^{1/2} \{D_{1n}\}^{1/2} \leq \{\lambda_{\max}(M_K)\}^{1/2} \sup_{A \in \mathcal{A}_K} \|C_\tau(A)\| \{D_{1n}\}^{1/2} \\ &= O_P(K^{1/2}) O_P((K \ln n/n)^{1/2}) = O_P(K(\ln n/n)^{1/2}) = o_P(1). \end{aligned}$$

It follows that

$$\sup_{A \in \mathcal{A}_K} \left| \hat{\mathbf{Q}}_n(A) - \mathbf{Q}_K(A) \right| = o_P(1). \quad (\text{A.1})$$

In view of the fact that the dimension of A is increasing with K , we cannot conclude that $\|\hat{A}_\tau - A_\tau\| = o_P(1)$ directly from (A.1) by referring to the usual consistency theorem that works for the estimation of finite dimensional parameter. Instead, we extend the consistency proof of White (1994, Theorem 3.4; see also Gallant and White (1988, Theorem 3.3)) to allow for diverging number of parameters. Because A_τ is identifiably unique on \mathcal{A}_K by Assumption A4, for any $\epsilon > 0$, there exists $n_0(\epsilon) < \infty$ such that $\inf_{n \geq n_0(\epsilon)} \left[\min_{A \in \mathcal{A}_K \cap \mathcal{N}_{K,\epsilon}^c} \mathbf{Q}_K(A) - \mathbf{Q}_K(A_\tau) \right] \equiv \eta(\epsilon) > 0$. Clearly, $\eta(\epsilon)$ is nondecreasing in ϵ and it cannot increase when ϵ decreases. By the definition of \hat{A}_τ , we have $\hat{\mathbf{Q}}_n(\hat{A}_\tau) \leq \hat{\mathbf{Q}}_n(A_\tau) + \eta(\epsilon)/3$. [Note that we allow for approximate minimization.] By (A.1), we have $\mathbf{Q}_K(\hat{A}_\tau) \leq \hat{\mathbf{Q}}_n(\hat{A}_\tau) + \eta(\epsilon)/3$ and $\hat{\mathbf{Q}}_n(A_\tau) < \mathbf{Q}_K(A_\tau) + \eta(\epsilon)/3$ w.a.p.1. It follows that $\mathbf{Q}_K(\hat{A}_\tau) \leq \hat{\mathbf{Q}}_n(\hat{A}_\tau) + \eta(\epsilon)/3 < \hat{\mathbf{Q}}_n(A_\tau) + 2\eta(\epsilon)/3 \leq \mathbf{Q}_K(A_\tau) + \eta(\epsilon)$ w.p.a.1. That is, $\mathbf{Q}_K(\hat{A}_\tau) - \mathbf{Q}_K(A_\tau) < \eta(\epsilon)$ w.p.a.1. It follows that $\hat{A}_\tau \in \mathcal{N}_{K,\epsilon}$. Since ϵ is arbitrary, we conclude that $\|\hat{A}_\tau - A_\tau\| = o_P(1)$.

Since $\Theta_\tau(A)$ is continuous in A by the maximum theorem, we have

$$\begin{aligned} \left\| \hat{\Theta}_\tau - \Theta_\tau \right\| &= \left\| \hat{\Theta}_\tau(\hat{A}_\tau) - \Theta_\tau(\hat{A}_\tau) + \Theta_\tau(\hat{A}_\tau) - \Theta_\tau(A_\tau) \right\| \\ &\leq \left\| \hat{\Theta}_\tau(\hat{A}_\tau) - \Theta_\tau(\hat{A}_\tau) \right\| + \left\| \Theta_\tau(\hat{A}_\tau) - \Theta_\tau(A_\tau) \right\| = O_P((K \ln n/n)^{1/2}) + o_P(1) = o_P(1) \end{aligned}$$

by Theorem 3.2(i) and Slutsky theorem. This completes the proof of part (i).

Now, we prove part (ii). Let $\tilde{\eta}_i(A, \Theta) \equiv g_i(A, \Theta) - g_i(A_\tau, \Theta_\tau) - E[g_i(A, \Theta)] + E[g_i(A_\tau, \Theta_\tau)]$. Using the arguments similar to those used in the proof of (A.6) in Lu and Su (2015) or Lemma A.6, we can prove that

$$\sup_{\|(A, \Theta) - (A_\tau, \Theta_\tau)\| \leq L\sqrt{K \ln n/n}} \left| n^{-1} \sum_{i=1}^n \mathbf{c}' \tilde{\eta}_i(A, \Theta) \right| = o_P(n^{-1/2}) \quad (\text{A.2})$$

for any $L > 0$ and $\mathbf{c} \in \mathbb{R}^{(k_2+k_3)K}$ with $\|\mathbf{c}\| = 1$. The problem is that we have only established the convergence of $(\hat{A}_\tau, \hat{\Theta}_\tau)$ in Frobenius norm but not its convergence rate. Let $e_{1n} = (K \ln n/n)^{1/2}$. In the following we first demonstrate that $\|\hat{A}_\tau - A_\tau\| = O_P(e_{1n})$ and $\|\hat{\Theta}_\tau - \Theta_\tau\| = O_P(e_{1n})$ based on the fact that

$$\sup_{\|(A, \Theta) - (A_\tau, \Theta_\tau)\| \leq L\sqrt{K \ln n/n}} \left| n^{-1} \sum_{i=1}^n \mathbf{c}' \tilde{\eta}_i(A, \Theta) \right| = O_P(e_{1n}). \quad (\text{A.3})$$

Then we apply (A.2) to refine the Bahadur representations for $\hat{A}_\tau - A_\tau$ and $\hat{\Theta}_\tau - \Theta_\tau$.

By Lemma A.2, (A.3) and Theorem 3.2(i),

$$\begin{aligned} \mathbf{o}_P(n^{-1/2}) &= n^{-1} \sum_{i=1}^n g_i(A, \hat{\Theta}_\tau(A)) \\ &= n^{-1} \sum_{i=1}^n g_i(A_\tau, \Theta_\tau) + \left\{ E[g_i(A, \hat{\Theta}_\tau(A))] - E[g_i(A_\tau, \Theta_\tau)] \right\} + n^{-1} \sum_{i=1}^n \tilde{\eta}_i(A, \hat{\Theta}_\tau(A)) \\ &= n^{-1} \sum_{i=1}^n g_i(A_\tau, \Theta_\tau) + \left\{ E[g_i(A, \hat{\Theta}_\tau(A))] - E[g_i(A_\tau, \Theta_\tau)] \right\} + \mathbf{O}_P(e_{1n}), \end{aligned} \quad (\text{A.4})$$

where $E[g_i(A, \hat{\Theta}_\tau(A))]$ denotes $E[g_i(A, \Theta)]|_{\Theta=\hat{\Theta}_\tau(A)}$ by the convention in empirical process theory, and hereafter $\mathbf{O}_P(\cdot)$ or $\mathbf{o}_P(\cdot)$ denotes the probability order of the Frobenius norm of the corresponding term. By Taylor expansion and Assumption A1(iii), for any $\|A - A_\tau\| = O(e_{2n})$ with $e_{2n} = o(1)$ we have

$$E[g_i(A, \hat{\Theta}_\tau(A))] - E[g_i(A, \Theta_\tau)] = [J_{K,A} + \mathbf{O}_P(e_{2n})](A - A_\tau) + [J_{K,\Theta} + \mathbf{O}_P(e_{2n})][\hat{\Theta}_\tau(A) - \Theta_\tau], \quad (\text{A.5})$$

where

$$\begin{aligned} J_{K,A} &\equiv \left. \frac{\partial E[g_i(A, \Theta)]}{\partial A'} \right|_{(A,\Theta)=(A_\tau,\Theta_\tau)} = -E \left\{ f_\varepsilon(v_i(A_\tau, \Theta_\tau) | \chi_i) \phi_{W_i}^K(U_i) P_K(D_i, U_i)' \right\} \text{ and} \\ J_{K,\Theta} &\equiv \left. \frac{\partial E[g_i(A, \Theta)]}{\partial \Theta'} \right|_{(A,\Theta)=(A_\tau,\Theta_\tau)} = -E \left\{ f_\varepsilon(v_i(A_\tau, \Theta_\tau) | \chi_i) \phi_{W_i}^K(U_i) \phi_{W_i}^K(U_i)' \right\} = -\Phi_K(A_\tau). \end{aligned}$$

Combining (A.4) and (A.5) yields

$$\hat{\Theta}_\tau(A) - \Theta_\tau = [1 + \mathbf{O}_P(e_{2n})] \Phi_K(A_\tau)^{-1} \left[n^{-1} \sum_{i=1}^n g_i(A_\tau, \Theta_\tau) + J_{K,A}(A - A_\tau) \right] + \mathbf{O}_P(e_{1n}). \quad (\text{A.6})$$

The last $k_3 K \times 1$ elements of $\hat{\Theta}_\tau(A) - \Theta_\tau$ are given by

$$\hat{C}_\tau(A) - 0 = [1 + \mathbf{O}_P(e_{2n})] \bar{\Phi}_{K,C} \left[n^{-1} \sum_{i=1}^n g_i(A_\tau, \Theta_\tau) + J_{K,A}(A - A_\tau) \right] + \mathbf{O}_P(e_{1n}), \quad (\text{A.7})$$

where recall $\Phi_K(A_\tau)^{-1} = [\bar{\Phi}'_{K,B} \quad \bar{\Phi}'_{K,C}]'$.

The second stage minimization problem implies that w.p.a.1, $\hat{A}_\tau = \arg \min_{A \in \mathcal{N}_{K,\epsilon_n}} \|\hat{C}_\tau(A)\|_{M_K}$, where $\epsilon_n = o(1)$ and

$$\|\hat{C}_\tau(A)\|_{M_K} = \left\| \bar{\Phi}_{K,C} n^{-1} \sum_{i=1}^n g_i(A_\tau, \Theta_\tau) + \bar{\Phi}_{K,C} J_{K,A}(A - A_\tau) \right\|_{M_K} [1 + O_P(e_{2n})] + \mathbf{O}_P(e_{1n}).$$

Noting that $\bar{\Phi}_{K,C} J_{K,A}$ has full rank under Assumptions A5(ii)-(iv), the solution \hat{A}_τ satisfies

$$\begin{aligned} \hat{A}_\tau - A_\tau &= - (J'_{K,A} \bar{\Phi}'_{K,C} M_K \bar{\Phi}_{K,C} J_{K,A})^{-1} J'_{K,A} \bar{\Phi}'_{K,C} M_K \bar{\Phi}_{K,C} n^{-1} \sum_{i=1}^n g_i(A_\tau, \Theta_\tau) [1 + O_P(e_{2n})] \\ &\quad + \mathbf{O}_P(e_{1n}). \\ &= \Omega_{A_\tau} n^{-1} \sum_{i=1}^n g_i(A_\tau, \Theta_\tau) [1 + O_P(e_{2n})] + \mathbf{O}_P(e_{1n}). \end{aligned} \quad (\text{A.8})$$

Next, in view of the fact that $\|\Omega_{A_\tau} n^{-1} \sum_{i=1}^n g_i(A_\tau, \Theta_\tau)\| = O_P((K/n)^{1/2} + K^{-\lambda/d})$ by moment calculations and Chebyshev inequality, we can obtain a rough probability bound: $\|\hat{A}_\tau - A_\tau\| = O_P[(K/n)^{1/2} + K^{-\lambda/d} + e_{1n}] = O_P(e_{1n})$ by Assumption A6(i). With this and (A.6), we can obtain a rough probability bound for $\hat{\Theta}_\tau - \Theta_\tau$ too: $\|\hat{\Theta}_\tau - \Theta_\tau\| = O_P(e_{1n})$.

Now, given these convergence rates for \hat{A}_τ and $\hat{\Theta}_\tau$, we can apply (A.2) and refine the results in the above procedure by replacing the term $\mathbf{O}_P(e_{1n})$ by $\mathbf{o}_P(n^{-1/2})$ in (A.4) and (A.6)-(A.8). With this replacement, we obtain an improved convergence rate for \hat{A}_τ : $\|\hat{A}_\tau - A_\tau\| = O_P(e_{2n})$ where $e_{2n} = (K/n)^{1/2} + K^{-\lambda/d}$. With this choice of e_{2n} , we obtain

$$\hat{A}_\tau - A_\tau = \Omega_{A_\tau} n^{-1} \sum_{i=1}^n g_i(A_\tau, \Theta_\tau) + \mathbf{o}_P(n^{-1/2}), \quad (\text{A.9})$$

as $K/n + K^{-2\lambda/d} = o(n^{-1/2})$. This proves (ii). Combining (A.6) and (A.8) with $\mathbf{O}_P(e_{1n})$ replaced by $\mathbf{O}_P(n^{-1/2})$ and choosing $e_{2n} = (K/n)^{1/2} + K^{-\lambda/d}$, we have

$$\begin{aligned}
\hat{\Theta}_\tau - \Theta_\tau &= \hat{\Theta}_\tau(\hat{A}_\tau) - \Theta_\tau \\
&= [1 + \mathbf{O}_P(e_{2n})] \Phi_K(A_\tau)^{-1} \left[I_{(k_2+k_3)K} - J_{K,A} (J'_{K,A} \bar{\Phi}'_{K,C} M_K \bar{\Phi}_{K,C} J_{K,A})^{-1} J'_{K,A} \bar{\Phi}'_{K,C} M_K \bar{\Phi}_{K,C} \right] \\
&\quad \times n^{-1} \sum_{i=1}^n g_i(A_\tau, \Theta_\tau) + \mathbf{O}_P(n^{-1/2}) \\
&= \Omega_{\Theta_\tau} n^{-1} \sum_{i=1}^n g_i(A_\tau, \Theta_\tau) + \mathbf{O}_P(n^{-1/2}), \tag{A.10}
\end{aligned}$$

where $\Omega_{\Theta_\tau} = \Phi_K(A_\tau)^{-1} [I_{(k_2+k_3)K} + \Omega_{A_\tau}]$. Thus (iii) follows. ■

To prove Theorems 3.4 and 3.5, we first state a lemma whose proof is given in Supplementary Appendix C.

Lemma A.7 *Suppose Assumptions A5(i)-(iv) hold. Then there exist \underline{c}_Ω and \bar{c}_Ω such that $0 < \underline{c}_\Omega \leq \lambda_{\min}(\Omega_\tau \Omega'_\tau) \leq \lambda_{\max}(\Omega_\tau \Omega'_\tau) \leq \bar{c}_\Omega < \infty$ uniformly in K .*

Proof of Theorem 3.4. We only show (i) as the proof of (ii) is analogous. Using the notation defined above (3.1), Minkowski inequality and Assumption A2, we have $\sup_{u \in \mathcal{U}} \|\hat{\alpha}_\tau(u) - \alpha_\tau(u)\| \leq \sup_{u \in \mathcal{U}} \left\| \Pi_\alpha(u)(\hat{A}_\tau - A_\tau) \right\| + \sup_{u \in \mathcal{U}} \|\Pi_\alpha(u)A_\tau - \alpha_\tau(u)\| = O(\zeta_K) \left\| \hat{A}_\tau - A_\tau \right\| + O_P(K^{-\lambda/d})$. By Theorem 3.3(ii), $\hat{A}_\tau - A_\tau = S_{1n} + S_{2n} + \mathbf{O}_P(n^{-1/2})$, where $S_{1n} = \Omega_{A_\tau} n^{-1} \sum_{i=1}^n \phi_{W_i}^K(U_i) \psi_\tau(\varepsilon_i)$ and $S_{2n} = \Omega_{A_\tau} n^{-1} \sum_{i=1}^n \phi_{W_i}^K(U_i) [\psi_\tau(\varepsilon_i - v_i(A_\tau, \Theta_\tau)) - \psi_\tau(\varepsilon_i)]$. First, one can readily show that $E \|S_{1n}\|^2 = n^{-1} \tau (1 - \tau) \text{tr}(\Omega_{A_\tau} \Psi_K \Omega'_{A_\tau}) \leq n^{-1} \lambda_{\max}(\Psi_K) \|\Omega_{A_\tau}\|^2 = O(K/n)$. Next, we show that

$$\|S_{2n}\|^2 = O_P(K^{-2\lambda/d}) \tag{A.11}$$

To see this, let $\mathbf{F}_{1n} = (\psi_\tau(\varepsilon_1 - v_1(A_\tau, \Theta_\tau)) - \psi_\tau(\varepsilon_1), \dots, \psi_\tau(\varepsilon_n - v_n(A_\tau, \Theta_\tau)) - \psi_\tau(\varepsilon_n))'$, and $\mathbf{F}_{2n} = (\phi_{W_1}^K(U_1), \dots, \phi_{W_n}^K(U_n))'$. By Assumptions A1(i) and (iii) and A2(i)-(ii), Taylor expansion, and Markov inequality, we can readily show that $\|\mathbf{F}_{1n}\|^2 = O_P(nK^{-2\lambda/d})$. Noting that $\mathbf{F}'_{2n} \mathbf{F}_{2n} / n = \hat{\Psi}_K$ and that $\mathbf{F}_{2n} (\mathbf{F}'_{2n} \mathbf{F}_{2n})^{-1} \mathbf{F}'_{2n}$ is a projection matrix with maximum eigenvalue 1, we have by Lemmas A.1(iv) and A.7,

$$\begin{aligned}
\|S_{2n}\|^2 &= n^{-2} \|\Omega_{A_\tau} \mathbf{F}'_{2n} \mathbf{F}_{1n}\|^2 = n^{-2} \text{tr}(\mathbf{F}_{1n}' \mathbf{F}_{2n} \mathbf{F}'_{2n} \mathbf{F}_{1n} \Omega'_{A_\tau} \Omega_{A_\tau}) \\
&\leq \lambda_{\max}(\Omega'_{A_\tau} \Omega_{A_\tau}) n^{-2} \text{tr}(\mathbf{F}_{1n}' \mathbf{F}_{2n} \mathbf{F}'_{2n} \mathbf{F}_{1n}) \\
&= \bar{c}_\Omega n^{-2} \text{tr}(\hat{\Psi}_K \hat{\Psi}_K^{-1/2} \mathbf{F}'_{2n} \mathbf{F}_{1n} \mathbf{F}'_{1n} \mathbf{F}_{2n} \hat{\Psi}_K^{-1/2}) \\
&\leq \bar{c}_\Omega \lambda_{\max}(\hat{\Psi}_K) n^{-2} \text{tr}(\hat{\Psi}_K^{-1/2} \mathbf{F}'_{2n} \mathbf{F}_{1n} \mathbf{F}'_{1n} \mathbf{F}_{2n} \hat{\Psi}_K^{-1/2}) \\
&\leq \bar{c}_\Omega [\bar{c}_\Psi + o_P(1)] n^{-1} \text{tr}(\mathbf{F}'_{1n} \mathbf{F}_{2n} (\mathbf{F}'_{2n} \mathbf{F}_{2n})^{-1} \mathbf{F}'_{2n} \mathbf{F}_{1n}) \\
&\leq \bar{c}_\Omega [\bar{c}_\Psi + o_P(1)] n^{-1} \|\mathbf{F}_{1n}\|^2 = O_P(K^{-2\lambda/d})
\end{aligned}$$

where the first and second inequalities follow from the fact that $\text{tr}(AB) \leq \lambda_{\max}(A) \text{tr}(B)$ for any symmetric matrix A and p.s.d. matrix B (see, e.g., Bernstein (2005, Proposition 8.4.13)), and the third equality follows from the fact that $\lambda_{\max}(\Omega'_{A_\tau} \Omega_{A_\tau}) = \lambda_{\max}(\Omega_{A_\tau} \Omega'_{A_\tau}) \leq \lambda_{\max}(\Omega_\tau \Omega'_\tau) \leq \bar{c}_\Omega$ by Lemma A.7. It follows that $\|\hat{A}_\tau - A_\tau\| = O_P((K/n)^{1/2} + K^{-\lambda/d})$ and $\sup_{u \in \mathcal{U}} \|\hat{\alpha}_\tau(u) - \alpha_\tau(u)\| = O_P[\zeta_K((K/n)^{1/2} + K^{-\lambda/d})]$. ■

Proof of Theorem 3.5. Let

$$\mathbf{V}_\tau = \begin{pmatrix} \hat{A}_\tau - A_\tau \\ \hat{B}_\tau - B_\tau \end{pmatrix} \text{ and } \mathbf{B}_\tau(u) = \begin{pmatrix} \alpha_\tau(u) - \Pi_\alpha(u) A_\tau \\ \beta_\tau(u) - \Pi_\beta(u) B_\tau \end{pmatrix}.$$

Noting that $\hat{\delta}_\tau(u) = \begin{pmatrix} \hat{\alpha}_\tau(u) \\ \hat{\beta}_\tau(u) \end{pmatrix} = \begin{pmatrix} \Pi_\alpha(u) \hat{A}_\tau \\ \Pi_\beta(u) \hat{B}_\tau \end{pmatrix} = \Pi(u) \begin{pmatrix} \hat{A}_\tau \\ \hat{B}_\tau \end{pmatrix}$, we have

$$\hat{\delta}_\tau(u) - \delta_\tau(u) = \Pi(u) \begin{pmatrix} \hat{A}_\tau - A_\tau \\ \hat{B}_\tau - B_\tau \end{pmatrix} - \begin{pmatrix} \alpha_\tau(u) - \Pi_\alpha(u) A_\tau \\ \beta_\tau(u) - \Pi_\beta(u) B_\tau \end{pmatrix} = \Pi(u) \mathbf{V}_\tau - \mathbf{B}_\tau(u).$$

Let $\Sigma_\tau(u) \equiv \tau(1-\tau)\Pi(u)\Omega_\tau\Psi_K\Omega'_\tau\Pi(u)'$ and $\Lambda \equiv \Lambda_\tau(u) = \Sigma_\tau(u)^{-1/2}\Pi(u)\Omega_\tau$. By Theorem 3.4,

$$\begin{aligned} \Sigma_\tau(u)^{-1/2}\sqrt{n}\left[\hat{\delta}_\tau(u) - \delta_\tau(u)\right] &= \Lambda n^{-1/2}\sum_{i=1}^n\phi_{W_i}^K(U_i)\psi_\tau(\varepsilon_i) \\ &\quad + \Lambda n^{-1/2}\sum_{i=1}^n\phi_{W_i}^K(U_i)\left[\psi_\tau(\varepsilon_i - v_i(A_\tau, \Theta_\tau)) - \psi_\tau(\varepsilon_i)\right] \\ &\quad + \sqrt{n}\Sigma_\tau(u)^{-1/2}\Pi(u)r_n - \sqrt{n}\Sigma_\tau(u)^{-1/2}\mathbf{B}_\tau(u) \\ &\equiv D_{1n} + D_{2n} + D_{3n} + D_{4n}, \text{ say,} \end{aligned}$$

where $\|r_n\| = o_P(n^{-1/2})$. We prove the theorem by showing that (i) $D_{1n} \xrightarrow{d} N(0, I_{k_1+k_2})$ and (ii) $D_{sn} = o_P(1)$ for $s = 2, 3, 4$.

First, we prove (i). Let $\omega \in \mathbb{R}^{k_1+k_2}$ such that $\|\omega\| = 1$. Let $\xi_{in} = n^{-1/2}\omega'\Lambda\phi_{W_i}^K(U_i)\psi_\tau(\varepsilon_i)$. Then $\omega'D_{1n} = \sum_{i=1}^n\xi_{in}$. By construction, $E(\xi_{in}) = 0$ and $E(\xi_{in}^2) = n^{-1}$. In addition,

$$\begin{aligned} E(\xi_{in}^4) &\leq \frac{1}{n^2}E\left[\omega'\Lambda\phi_{W_i}^K(U_i)\phi_{W_i}^K(U_i)'\Lambda'\omega\omega'\Lambda\phi_{W_i}^K(U_i)\phi_{W_i}^K(U_i)'\Lambda'\omega\right] \\ &= \frac{1}{n^2}\text{tr}\left\{E\left[\phi_{W_i}^K(U_i)\phi_{W_i}^K(U_i)'\Lambda'\omega\omega'\Lambda\phi_{W_i}^K(U_i)\phi_{W_i}^K(U_i)'\Lambda'\omega\omega'\Lambda\right]\right\} \\ &\leq \frac{1}{n^2}\lambda_{\max}(\Lambda'\omega\omega'\Lambda)\text{tr}\left\{E\left[\phi_{W_i}^K(U_i)\phi_{W_i}^K(U_i)'\Lambda'\omega\omega'\Lambda\phi_{W_i}^K(U_i)\phi_{W_i}^K(U_i)'\right]\right\} \\ &= \frac{1}{n^2}\lambda_{\max}(\Lambda'\omega\omega'\Lambda)\text{tr}\left\{\Lambda'\omega\omega'\Lambda E\left[\phi_{W_i}^K(U_i)\phi_{W_i}^K(U_i)'\phi_{W_i}^K(U_i)\phi_{W_i}^K(U_i)'\right]\right\} \\ &\leq \frac{1}{n^2}\left[\lambda_{\max}(\Lambda'\omega\omega'\Lambda)\right]^2\text{tr}\left\{E\left[\phi_{W_i}^K(U_i)\phi_{W_i}^K(U_i)'\phi_{W_i}^K(U_i)\phi_{W_i}^K(U_i)'\right]\right\} \\ &\leq \frac{1}{n^2}\left[\lambda_{\max}(\Lambda'\Lambda)\right]^2O(\zeta_K^2)E\left\|\phi_{W_i}^K(U_i)\right\|^2 = O(\zeta_K^2K^3/n^2) = o(n^{-1}), \end{aligned}$$

where we use the fact that $\lambda_{\max}(\omega\omega') \leq \text{tr}(\omega\omega') = 1$, and that

$$\begin{aligned} \lambda_{\max}(\Lambda'\omega\omega'\Lambda) &\leq \lambda_{\max}(\Lambda'\Lambda) \leq [\tau(1-\tau)]^{-1}\text{tr}\left(\Omega'_\tau\Pi(u)'\left[\Pi(u)\Omega_\tau\Psi_K\Omega'_\tau\Pi(u)'\right]^{-1}\Pi(u)\Omega_\tau\right) \\ &\leq \underline{c}_\Psi^{-1}[\tau(1-\tau)]^{-1}\text{tr}\left(\Omega'_\tau\Pi(u)'\left[\Pi(u)\Omega_\tau\Omega'_\tau\Pi(u)'\right]^{-1}\Pi(u)\Omega_\tau\right) = \underline{c}_\Psi^{-1}[\tau(1-\tau)]^{-1}K. \end{aligned}$$

It follows that $\omega'D_{1n} \xrightarrow{d} N(0, 1)$ by the Liapounov central limit theorem for triangular array independent sequences (see, e.g., Davidson (1994, Theorem 23.11)). Thus (i) follows.

Now we prove (ii). By straightforward moment calculations and Assumption A6(ii), we can show that

$$\|E(D_{2n})\| = O\left(\zeta_{K,u}^{-1}n^{1/2}K^{-\lambda/d}\right) = o(1) \text{ and } \|\text{Var}(D_{2n})\| = O\left(\zeta_{K,u}^{-2}K^{1-2\lambda/d}\right) = o(1).$$

Then $D_{2n} = o_P(1)$ by Chebyshev inequality. For D_{3n} , we have

$$\begin{aligned}\|D_{3n}\|^2 &= ntr \left(r_n r_n' \Pi(u)' \Sigma_\tau(u)^{-1} \Pi(u) \right) \leq \lambda_{\max} \left(\Pi(u)' \Sigma_\tau(u)^{-1} \Pi(u) \right) \left\{ n \|r_n\|^2 \right\} \\ &\leq [\lambda_{\min}(\Sigma_\tau(u))]^{-1} \|\Pi(u)\|^2 \left\{ n \|r_n\|^2 \right\} \leq [\tau(1-\tau) \underline{c}_\Psi \underline{c}_\Omega]^{-1} \left\{ n \|r_n\|^2 \right\} \\ &= O(1) o_P(1) = o_P(1).\end{aligned}$$

where we use the fact that $\lambda_{\min}(\Sigma_\tau(u)) \geq \tau(1-\tau) \lambda_{\min}(\Omega_\tau \Psi_K \Omega_\tau') \|\Pi(u)\|^2 \geq \tau(1-\tau) \underline{c}_\Psi \underline{c}_\Omega \|\Pi(u)\|^2$ by Assumption A5(ii) and Lemma A.7. Similarly, we can show that $\|D_{4n}\|^2 \leq [\lambda_{\min}(\Sigma_\tau(u))]^{-1} n \|\mathbf{B}_\tau(u)\|^2 = O(\zeta_u^{-2} n K^{-2\lambda/d}) = o(1)$ by Assumption A6(ii). This completes the proof. ■

B Proof of the results in section 4

Proof of Theorem 4.1. Using the notation defined in the proof of Theorem 3.5, we have $\hat{\delta}_{1\tau}(u) - \delta_{1\tau}(u) = \mathbb{S} \left[\hat{\delta}_\tau(u) - \delta_\tau(u) \right] = \mathbb{S} \Pi(u) \mathbf{V}_\tau - \mathbb{S} \mathbf{B}_\tau(u)$. It follows that

$$\begin{aligned}\sqrt{n} \left(\hat{\delta}_{1\tau} - \delta_{1\tau} \right) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[\hat{\delta}_{1\tau}(U_i) - \delta_{1\tau}(U_i) \right] + \frac{1}{\sqrt{n}} \sum_{i=1}^n [\delta_{1\tau}(U_i) - \delta_{1\tau}] \\ &= \frac{1}{\sqrt{n}} \mathbb{S} \sum_{i=1}^n \Pi(U_i) \mathbf{V}_\tau + \frac{1}{\sqrt{n}} \mathbb{S} \sum_{i=1}^n \mathbf{B}_\tau(U_i) + \frac{1}{\sqrt{n}} \sum_{i=1}^n [\delta_{1\tau}(U_i) - \delta_{1\tau}] \\ &\equiv R_{1n} + R_{2n} + R_{3n}, \text{ say.}\end{aligned}$$

Observe that $R_{3n} = 0$ under \mathbb{H}_0 and $\|R_{2n}\| \leq n^{1/2} \|\mathbb{S}\| \sup_u \|\mathbf{B}_\tau(u)\| \leq O(n^{1/2} K^{-\lambda/d}) = o(1)$ by Assumptions A2 and A6*. It suffices to prove the theorem by showing that $R_{1n} \xrightarrow{d} N(0, \Sigma_{\delta_{1\tau}})$.

By Theorem 3.3, $\mathbf{V}_\tau = \Omega_\tau \frac{1}{n} \sum_{i=1}^n \phi_{W_i}^K(U_i) \psi_\tau(\varepsilon_i - v_i(A_\tau, \Theta_\tau)) + R_\tau$ with $\|R_\tau\| = o_P(n^{-1/2})$. We decompose R_{1n} as follows:

$$\begin{aligned}R_{1n} &= \frac{1}{\sqrt{n}} \mathbb{S} \sum_{i=1}^n \Pi(U_i) \mathbf{V}_\tau \\ &= \frac{1}{n^{3/2}} \mathbb{S} \sum_{j=1}^n \Pi(U_j) \Omega_\tau \sum_{i=1}^n \phi_{W_i}^K(U_i) \psi_\tau(\varepsilon_i - v_i(A_\tau, \Theta_\tau)) + \frac{1}{n^{1/2}} \mathbb{S} \sum_{j=1}^n \Pi(U_j) R_\tau \\ &= \frac{1}{n^{1/2}} \mathbb{S} \bar{\Pi} \Omega_\tau \sum_{i=1}^n \phi_{W_i}^K(U_i) \psi_\tau(\varepsilon_i) + \frac{1}{n^{1/2}} \mathbb{S} \frac{1}{n} \sum_{j=1}^n [\Pi(U_j) - \bar{\Pi}] \Omega_\tau \sum_{i=1}^n \phi_{W_i}^K(U_i) \psi_\tau(\varepsilon_i) \\ &\quad + \frac{1}{n^{3/2}} \mathbb{S} \sum_{j=1}^n \Pi(U_j) \Omega_\tau \sum_{i=1}^n \phi_{W_i}^K(U_i) \eta_{1\tau i} + \frac{1}{n^{1/2}} \mathbb{S} \sum_{j=1}^n \Pi(U_j) R_\tau \\ &\equiv R_{1n,1} + R_{1n,2} + R_{1n,3} + R_{1n,4}, \text{ say,}\end{aligned}$$

where $\eta_{1\tau i} = \psi_\tau(\varepsilon_i - v_i(A_\tau, \Theta_\tau)) - \psi_\tau(\varepsilon_i)$ and $\bar{\Pi} = E[\Pi(U_j)]$.

Let $\omega \in \mathbb{R}^r$ such that $\|\omega\| = 1$. Let $\varsigma_{in} = n^{-1/2} \omega' \mathbb{S} \bar{\Pi} \Omega_\tau \phi_{W_i}^K(U_i) \psi_\tau(\varepsilon_i)$. Then $\omega' R_{1n,1} = \frac{1}{\sqrt{n}} \omega' \mathbb{S} \bar{\Pi} \Omega_\tau \times \sum_{i=1}^n \phi_{W_i}^K(U_i) \psi_\tau(\varepsilon_i) = \sum_{i=1}^n \varsigma_{in}$. Note that $E(\varsigma_{in}^2) = \tau(1-\tau) n^{-1} \omega' \mathbb{S} \bar{\Pi} \Omega_\tau \Psi_K \Omega_\tau' \bar{\Pi}' \mathbb{S}' \omega = n^{-1} \omega' \Sigma_{\delta_{1\tau}}^K \omega$, and $E(\varsigma_{in}^4) = O(\zeta_K^2 K^3/n^2) = o(n^{-1})$ following the same arguments as used in the proof of Theorem 3.5. It follows that $R_{1n,1} \xrightarrow{d} N(0, \Sigma_{\delta_{1\tau}})$ by the Liapounov central limit theorem for triangular array independent sequences.

Since $E\|R_{1n,2}\|^2 = O(K/n)$, $R_{1n,2} = o_P((K/n)^{1/2}) = o_P(1)$ by Chebyshev inequality. By the same inequality, one can readily show that $\|\sum_{i=1}^n \phi_{W_i}^K(U_i) \eta_{1\tau i}\|^2 = o_P(n^2 K^{-2\lambda/d})$. Using the arguments as used in the proof of Lemma A.1(ii), one can show that $\lambda_{\max}(n^{-2} \sum_{j=1}^n \Pi(U_j) \sum_{i=1}^n \Pi(U_i)')$

$= \lambda_{\max}(\bar{\Pi}\bar{\Pi}') + O_P(K/n^{1/2}) = O_P(1)$. In conjunction with Lemma A.7, these imply that

$$\begin{aligned}
\|R_{1n,3}\|^2 &= \frac{1}{n^3} \text{tr} \left\{ \Omega_\tau \sum_{i=1}^n \phi_{W_i}^K(U_i) \eta_{1\tau i} \sum_{k=1}^n \phi_{W_k}^K(U_k)' \eta_{1\tau k} \Omega_\tau' \sum_{l=1}^n \Pi(U_l)' \mathbb{S}' \mathbb{S} \sum_{j=1}^n \Pi(U_j) \right\} \\
&\leq \frac{1}{n^3} \text{tr} \left\{ \Omega_\tau \sum_{i=1}^n \phi_{W_i}^K(U_i) \eta_{1\tau i} \sum_{k=1}^n \phi_{W_k}^K(U_k)' \eta_{1\tau k} \Omega_\tau' \right\} \text{tr} \left\{ \sum_{l=1}^n \Pi(U_l)' \mathbb{S}' \mathbb{S} \sum_{j=1}^n \Pi(U_j) \right\} \\
&= \frac{1}{n^3} \text{tr} \left\{ \sum_{i=1}^n \phi_{W_i}^K(U_i) \eta_{1\tau i} \sum_{k=1}^n \phi_{W_k}^K(U_k)' \eta_{1\tau k} \Omega_\tau' \Omega_\tau \right\} \text{tr} \left\{ \mathbb{S}' \mathbb{S} \sum_{j=1}^n \Pi(U_j) \sum_{l=1}^n \Pi(U_l)' \right\} \\
&\leq \frac{1}{n} \lambda_{\max}(\Omega_\tau' \Omega_\tau) \lambda_{\max} \left(\frac{1}{n^2} \sum_{j=1}^n \Pi(U_j) \sum_{l=1}^n \Pi(U_l)' \right) \left\| \sum_{i=1}^n \phi_{W_i}^K(U_i) \eta_{1\tau i} \right\|^2 \|\mathbb{S}\|^2 \\
&= \frac{1}{n} O(1) O_P(1) O_P(1) \left(n^2 K^{-2\lambda/d} \right) O(1) = O_P(n K^{-2\lambda/d}) = o_P(1),
\end{aligned}$$

where the first inequality follows from the fact that $\text{tr}(AB) \leq \text{tr}(A)\text{tr}(B)$ for any two conformable p.s.d. matrices A and B , and the second inequality follows from the fact that $\text{tr}(AB) \leq \lambda_{\max}(A)\text{tr}(B)$ for real symmetric matrix A and p.s.d. matrix B . Similarly,

$$\begin{aligned}
\|R_{1n,4}\|^2 &= \frac{1}{n} \text{tr} \left(R_\tau R_\tau' \sum_{j=1}^n \Pi(U_j)' \mathbb{S}' \mathbb{S} \sum_{i=1}^n \Pi(U_i) \right) \leq \lambda_{\max} \left(\frac{1}{n^2} \sum_{j=1}^n \Pi(U_j)' \mathbb{S}' \mathbb{S} \sum_{i=1}^n \Pi(U_i) \right) [n \text{tr}(R_\tau R_\tau')] \\
&\leq \lambda_{\max}(\mathbb{S}' \mathbb{S}) \lambda_{\max} \left(\frac{1}{n^2} \sum_{j=1}^n \sum_{i=1}^n \Pi(U_j)' \Pi(U_i) \right) (n \|R_\tau\|^2) \\
&= O(1) O_P(1) o_P(1) = o_P(1).
\end{aligned}$$

It follows that $R_{1n} \xrightarrow{d} N(0, \Sigma_{\delta_{1\tau}})$. This completes the proof. ■

Proof of Theorems 4.2 and 4.3. We only prove Theorem 4.3, as the proof of Theorem 4.2 is a special case. Let $a_i \equiv a(U_i)$ and $\bar{a} \equiv n^{-1} \sum_{i=1}^n a_i$. Decompose T_n as follows

$$T_n = \sum_{i=1}^n \left[\hat{\delta}_{1\tau}(U_i) - \bar{\delta}_{1\tau} + \bar{\delta}_{1\tau} - \bar{\delta}_{1\tau} \right]' \left[\hat{\delta}_{1\tau}(U_i) - \bar{\delta}_{1\tau} + \bar{\delta}_{1\tau} - \bar{\delta}_{1\tau} \right] a_i = T_{n1} + T_{n2} - 2T_{n3}, \quad (\text{B.1})$$

where $\bar{\delta}_{1\tau} = n^{-1} \sum_{i=1}^n \delta_{1\tau}(U_i)$, $T_{n1} = \sum_{i=1}^n [\hat{\delta}_{1\tau}(U_i) - \bar{\delta}_{1\tau}] [\hat{\delta}_{1\tau}(U_i) - \bar{\delta}_{1\tau}] a_i$, $T_{n2} = n\bar{a} [\bar{\delta}_{1\tau} - \bar{\delta}_{1\tau}]' [\bar{\delta}_{1\tau} - \bar{\delta}_{1\tau}]$, and $T_{n3} = [\bar{\delta}_{1\tau} - \bar{\delta}_{1\tau}]' \sum_{i=1}^n [\hat{\delta}_{1\tau}(U_i) - \bar{\delta}_{1\tau}] a_i$. We further decompose T_{n1} as follows:

$$\begin{aligned}
T_{n1} &= \sum_{i=1}^n \left[\hat{\delta}_{1\tau}(U_i) - \delta_{1\tau}(U_i) \right]' \left[\hat{\delta}_{1\tau}(U_i) - \delta_{1\tau}(U_i) \right] a_i + \sum_{i=1}^n \left[\delta_{1\tau}(U_i) - \bar{\delta}_{1\tau} \right]' \left[\delta_{1\tau}(U_i) - \bar{\delta}_{1\tau} \right] a_i \\
&\quad + 2 \sum_{i=1}^n \left[\hat{\delta}_{1\tau}(U_i) - \delta_{1\tau}(U_i) \right]' \left[\delta_{1\tau}(U_i) - \bar{\delta}_{1\tau} \right] a_i \\
&\equiv T_{n11} + T_{n12} + 2T_{n13}, \text{ say.} \tag{B.2}
\end{aligned}$$

We complete the proof of the theorem by showing that $\mathbb{H}_1(\sigma_n^{1/2} n^{-1/2})$, (i) $\sigma_n^{-1}(T_{n11} - \mathbb{B}_n) \xrightarrow{d} N(0, 1)$, (ii) $\sigma_n^{-1} T_{n12} = \mu_0 + o_P(1)$, (iii) $\sigma_n^{-1} T_{n13} = o_P(1)$, (iv) $\sigma_n^{-1} T_{n2} = o_P(1)$, (v) $\sigma_n^{-1} T_{n3} = o_P(1)$, (vi) $\sigma_n^{-1}(\mathbb{B}_n - \mathbb{B}_n) = o_P(1)$, and (vii) $(\hat{\sigma}_n - \sigma_n)/\sigma_n = o_P(1)$. These are respectively proven in Propositions B.1-B.7 below.

Proposition B.1 *Suppose that the conditions in Theorem 4.3 hold. Then $\sigma_n^{-1}(T_{n11} - \mathbb{B}_n) \xrightarrow{d} N(0, 1)$.*

Proof. Let $r_{1i} = \mathbb{S}\Pi(U_i)\Omega_\tau \frac{1}{n} \sum_{j=1}^n \phi_{W_j}^K(U_j)\psi_\tau(\varepsilon_j)$, $r_{2i} = \mathbb{S}\Pi(U_i)\Omega_\tau \frac{1}{n} \sum_{j=1}^n \phi_{W_j}^K(U_j)[\psi_\tau(\varepsilon_j - v_j(A_\tau, \Theta_\tau)) - \psi_\tau(\varepsilon_j)]$, and $r_{3i} = \mathbb{S}[\Pi(U_i)R_\tau + \mathbf{B}_\tau(U_i)]$. Noting that $\hat{\delta}_{1\tau}(U_i) - \delta_{1\tau}(U_i) = \mathbb{S}\Pi(U_i)\mathbf{V}_\tau + \mathbb{S}\mathbf{B}_\tau(U_i) = r_{1i} + r_{2i} + r_{3i}$, we decompose T_{n11} as follows:

$$\begin{aligned} T_{n11} &= \sum_{i=1}^n \left[\hat{\delta}_{1\tau}(U_i) - \delta_{1\tau}(U_i) \right]' \left[\hat{\delta}_{1\tau}(U_i) - \delta_{1\tau}(U_i) \right] a_i \\ &= \sum_{i=1}^n \{r'_{1i}r_{1i} + r'_{2i}r_{2i} + r'_{3i}r_{3i} + 2r'_{1i}r_{2i} + 2r'_{1i}r_{3i} + 2r'_{2i}r_{3i}\} a_i \\ &\equiv T_{n11,1} + T_{n11,2} + T_{n11,3} + 2T_{n11,4} + 2T_{n11,5} + 2T_{n11,6}, \end{aligned}$$

where, e.g., $T_{n11,1} = \sum_{i=1}^n r'_{1i}r_{1i}a_i$. We prove the lemma by showing that (i) $\sigma_n^{-1}(T_{n11,1} - \mathbb{B}_n) \xrightarrow{d} N(0, 1)$, and (ii) $\sigma_n^{-1}T_{n11,s} = o_P(1)$ for $s = 2, 3, \dots, 6$.

We first prove (i). Recall $\Upsilon_i = \Pi(U_i)' \mathbb{S}' \mathbb{S} \Pi(U_i) a_i$ and $\bar{\Omega}_\tau = \Omega_\tau' E(\Upsilon_1) \Omega_\tau$. Let $\zeta_i = (U_i', W_i', \varepsilon_i)'$ and $\varphi_n(\zeta_j, \zeta_k) = \psi_\tau(\varepsilon_j) \phi_{W_j}^K(U_j)' \bar{\Omega}_\tau \phi_{W_k}^K(U_k) \psi_\tau(\varepsilon_k)$. Let $\tilde{\Upsilon}_i = \Upsilon_i - E(\Upsilon_i)$. Then

$$\begin{aligned} T_{n11,1} &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \psi_\tau(\varepsilon_j) \phi_{W_j}^K(U_j)' \Omega_\tau' \Upsilon_i \Omega_\tau \phi_{W_k}^K(U_k) \psi_\tau(\varepsilon_k) \\ &= \frac{2}{n} \sum_{1 \leq j < k \leq n} \varphi_n(\zeta_j, \zeta_k) + \frac{1}{n} \sum_{j=1}^n \varphi_n(\zeta_j, \zeta_j) + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \phi_{W_j}^K(U_j) \psi_\tau(\varepsilon_j) \Omega_\tau' \tilde{\Upsilon}_i \Omega_\tau \phi_{W_k}^K(U_k) \psi_\tau(\varepsilon_k) \\ &\equiv T_{n11,1a} + T_{n11,1b} + T_{n11,1c}, \text{ say.} \end{aligned}$$

Noting that

$$\begin{aligned} \|E(\Upsilon_1)\|^2 &= \text{tr}(E[\Pi(U_i)' \mathbb{S}' \mathbb{S} \Pi(U_i) a_i] E[\Pi(U_i)' \mathbb{S}' \mathbb{S} \Pi(U_i) a_i]) \\ &\geq [\lambda_{\min}(\mathbb{S}' \mathbb{S})]^2 \text{tr}(E[\Pi(U_i)' \Pi(U_i) a_i] E[\Pi(U_i)' \Pi(U_i) a_i]) \\ &\geq \underline{c}_{\mathbb{S}}^2 \underline{c}_{\Pi}^2 \text{tr}(I_{(k_1+k_2)K}) = (k_1 + k_2) \underline{c}_{\mathbb{S}}^2 \underline{c}_{\Pi}^2 K \text{ by Assumption 7} \end{aligned}$$

where $\underline{c}_{\mathbb{S}} = \lambda_{\min}(\mathbb{S}' \mathbb{S})$, we have by Lemma A.7 and Assumption A5

$$\begin{aligned} \sigma_n^2 &= 2\tau^2(1-\tau)^2 \text{tr}(\bar{\Omega}_\tau \Psi_K \bar{\Omega}_\tau \Psi_K) \geq 2\tau^2(1-\tau)^2 \underline{c}_{\Psi}^2 \text{tr}(\bar{\Omega}_\tau \bar{\Omega}_\tau) \\ &= 2\tau^2(1-\tau)^2 \underline{c}_{\Psi}^2 \text{tr}\{E(\Upsilon_1) \Omega_\tau \Omega_\tau' E(\Upsilon_1) \Omega_\tau \Omega_\tau'\} \\ &\geq 2\tau^2(1-\tau)^2 \underline{c}_{\Psi}^2 \underline{c}_{\Omega}^2 \|E(\Upsilon_1)\|^2 \geq 2\tau^2(1-\tau)^2 \underline{c}_{\Psi}^2 \underline{c}_{\Omega}^2 (k_1 + k_2) \underline{c}_{\mathbb{S}}^2 \underline{c}_{\Pi}^2 K. \end{aligned} \quad (\text{B.3})$$

Let $\mathbf{H}_n(\zeta_j, \zeta_k) = \frac{2}{n\sigma_n} \varphi_n(\zeta_j, \zeta_k)$. Noting that $\sigma_n^{-1}T_{n11,1a} = \sum_{1 \leq j < k \leq n} \mathbf{H}_n(\zeta_j, \zeta_k)$ is a second order degenerate U -statistic with kernel function, we prove (i) by verifying all the conditions of Theorem 1 in Hall (1984) are satisfied. By construction, $\mathbf{H}_n(\cdot, \cdot)$ is symmetric, $E[\mathbf{H}_n(\zeta_1, \zeta_2) | \zeta_2] = 0$, and

$$\begin{aligned} E[\mathbf{H}_n^2(\zeta_1, \zeta_2)] &= \frac{4}{n^2 \sigma_n^2} E \left\{ \left[\psi_\tau(\varepsilon_1) \phi_{W_1}^K(U_1)' \bar{\Omega}_\tau \phi_{W_2}^K(U_2) \psi_\tau(\varepsilon_2) \right]^2 \right\} \\ &= \frac{4}{n^2 \sigma_n^2} E \text{tr} \left\{ \bar{\Omega}_\tau \phi_{W_2}^K(U_2) \psi_\tau(\varepsilon_2)^2 \phi_{W_2}^K(U_2)' \bar{\Omega}_\tau \phi_{W_1}^K(U_1) \psi_\tau(\varepsilon_1)^2 \phi_{W_1}^K(U_1)' \right\} \\ &= \frac{4}{n^2 \sigma_n^2} \text{tr} \left\{ \bar{\Omega}_\tau E \left[\phi_{W_2}^K(U_2) \phi_{W_2}^K(U_2)' \psi_\tau(\varepsilon_2)^2 \right] \bar{\Omega}_\tau E \left[\phi_{W_1}^K(U_1) \phi_{W_1}^K(U_1)' \psi_\tau(\varepsilon_1)^2 \right] \right\} \\ &= \frac{4\tau^2(1-\tau)^2}{n^2 \sigma_n^2} \text{tr}(\bar{\Omega}_\tau \Psi_K \bar{\Omega}_\tau \Psi_K) = \frac{2}{n^2} < \infty. \end{aligned}$$

Using the fact that $|\psi_\tau(\varepsilon_2)| \leq 1$ and that $b'Bb \leq \lambda_{\max}(B)b'b$ for conformable vector b and symmetric matrix B , $\text{tr}(AB) \leq \lambda_{\max}(A)\text{tr}(B)$ for any symmetric matrix A and p.s.d. matrix B , we have

$$\begin{aligned}
& E [\mathbf{H}_n^4(\zeta_1, \zeta_2)] \\
& \leq \frac{16}{n^4\sigma_n^4} E \left\{ \phi_{W_1}^K(U_1)' \bar{\Omega}_\tau \phi_{W_2}^K(U_2) \phi_{W_2}^K(U_2)' \bar{\Omega}_\tau \phi_{W_1}^K(U_1) \phi_{W_1}^K(U_1)' \bar{\Omega}_\tau \phi_{W_2}^K(U_2) \phi_{W_2}^K(U_2)' \bar{\Omega}_\tau \phi_{W_1}^K(U_1) \right\} \\
& \leq c_\phi^2 \bar{c}_\Omega^2 \frac{16}{n^4\sigma_n^4} E \left\{ \phi_{W_1}^K(U_1)' \bar{\Omega}_\tau \phi_{W_2}^K(U_2) \phi_{W_2}^K(U_2)' \phi_{W_2}^K(U_2) \phi_{W_2}^K(U_2)' \bar{\Omega}_\tau \phi_{W_1}^K(U_1) \right\} \\
& = \frac{16c_\phi^2 \bar{c}_\Omega^2}{n^4\sigma_n^4} \text{tr} \left\{ \bar{\Omega}_\tau E \left[\phi_{W_2}^K(U_2) \phi_{W_2}^K(U_2)' \phi_{W_2}^K(U_2) \phi_{W_2}^K(U_2)' \right] \bar{\Omega}_\tau E \left[\phi_{W_1}^K(U_1) \phi_{W_1}^K(U_1)' \right] \right\} \\
& \leq \frac{16c_\phi^2 \bar{c}_\Omega^2 \bar{c}_\Psi}{n^4\sigma_n^4} \text{tr} \left\{ \bar{\Omega}_\tau E \left[\phi_{W_2}^K(U_2) \phi_{W_2}^K(U_2)' \phi_{W_2}^K(U_2) \phi_{W_2}^K(U_2)' \right] \bar{\Omega}_\tau \right\} \\
& \leq \frac{16c_\phi^2 \bar{c}_\Omega^4 \bar{c}_\Psi}{n^4\sigma_n^4} E \left\| \phi_{W_2}^K(U_2) \right\|^4 = O \left(\frac{\zeta_K^2 K^2}{n^4\sigma_n^4} \right)
\end{aligned}$$

where $c_\phi = \sup_{(w,u)} \left\| \phi_w^K(u) \right\| = O(\zeta_K)$, and $\bar{c}_\Omega = \lambda_{\max}(\bar{\Omega}_\tau) = \lambda_{\max}(\Omega_\tau' E(\Upsilon_1) \Omega_\tau) \leq \lambda_{\max}(E(\Upsilon_1)) \bar{c}_\Omega \leq \|\mathbb{S}\|^2 \bar{c}_\Pi \bar{c}_\Omega = O(1)$ by Assumptions A5 and A7(ii) and Lemma A.7.

Let $\mathbf{G}_n(v, w) = E[\mathbf{H}_n(\zeta_1, v) \mathbf{H}_n(\zeta_2, w)]$. Then

$$\begin{aligned}
E[\mathbf{G}_n^2(\zeta_1, \zeta_2)] & = \frac{16}{n^4\sigma_n^4} E[\varphi_n(\zeta_1, \zeta_3) \varphi_n(\zeta_2, \zeta_4)]^2 \\
& \leq \frac{16}{n^4\sigma_n^4} E \left[\phi_{W_1}^K(U_1)' \bar{\Omega}_\tau \phi_{W_3}^K(U_3) \phi_{W_2}^K(U_2)' \bar{\Omega}_\tau \phi_{W_4}^K(U_4) \right]^2 \\
& = \frac{16}{n^4\sigma_n^4} E \left\{ \phi_{W_1}^K(U_1)' \bar{\Omega}_\tau \phi_{W_3}^K(U_3) \phi_{W_2}^K(U_2)' \bar{\Omega}_\tau \Psi_K \bar{\Omega}_\tau \phi_{W_2}^K(U_2) \phi_{W_3}^K(U_3)' \bar{\Omega}_\tau \phi_{W_1}^K(U_1) \right\} \\
& \leq \frac{16c_K}{n^4\sigma_n^4} E \left\{ \phi_{W_1}^K(U_1)' \bar{\Omega}_\tau \phi_{W_3}^K(U_3) \phi_{W_2}^K(U_2)' \phi_{W_2}^K(U_2) \phi_{W_3}^K(U_3)' \bar{\Omega}_\tau \phi_{W_1}^K(U_1) \right\} \\
& = \frac{16c_K}{n^4\sigma_n^4} E \left\{ \phi_{W_2}^K(U_2)' \phi_{W_2}^K(U_2) \phi_{W_1}^K(U_1)' \bar{\Omega}_\tau \phi_{W_3}^K(U_3) \phi_{W_3}^K(U_3)' \bar{\Omega}_\tau \phi_{W_1}^K(U_1) \right\} \\
& = \frac{16c_K}{n^4\sigma_n^4} E \left\{ \phi_{W_2}^K(U_2)' \phi_{W_2}^K(U_2) \phi_{W_1}^K(U_1)' \bar{\Omega}_\tau \Psi_K \bar{\Omega}_\tau \phi_{W_1}^K(U_1) \right\} \\
& = \frac{16c_K}{n^4\sigma_n^4} \text{tr} \left\{ E \left[\phi_{W_1}^K(U_1) \phi_{W_2}^K(U_2)' \phi_{W_2}^K(U_2) \phi_{W_1}^K(U_1)' \right] \bar{\Omega}_\tau \Psi_K \bar{\Omega}_\tau \right\} \\
& \leq \frac{16c_K^2}{n^4\sigma_n^4} \text{tr} \left\{ E \left[\phi_{W_1}^K(U_1) E \left[\phi_{W_2}^K(U_2)' \phi_{W_2}^K(U_2) \right] \phi_{W_1}^K(U_1)' \right] \right\} \\
& \leq \frac{16c_K^2 \bar{c}_\Psi}{n^4\sigma_n^4} \text{tr}(\Psi_K) = O \left(\frac{K}{n^4\sigma_n^4} \right),
\end{aligned}$$

where $c_K \equiv \lambda_{\max}(\bar{\Omega}_\tau \Psi_K \bar{\Omega}_\tau) \leq \bar{c}_\Psi \lambda_{\max}(\bar{\Omega}_\tau)^2 \leq \bar{c}_\Psi \|\mathbb{S}\|^4 [\lambda_{\max}(\bar{\Pi}' \bar{\Pi})]^2 = O(1)$ by Assumptions A5 and A7(i), the first inequality follows from the fact that $|\psi_\tau(\varepsilon_2)| \leq 1$, the second inequality from the fact that $a'Ba \leq \lambda_{\max}(B)a'a$ for any conformable vector a and symmetric matrix B , and the third inequality from the fact that $\text{tr}(AB) \leq \lambda_{\max}(A)\text{tr}(B)$ for any symmetric matrix A and p.s.d. matrix B . It follows that

$$\frac{E[\mathbf{G}_n^2(\zeta_1, \zeta_2)] + n^{-1} E[\mathbf{H}_n^4(\zeta_1, \zeta_2)]}{\{E[\mathbf{H}_n^2(\zeta_1, \zeta_2)]\}^2} = O \left(\frac{K}{\sigma_n^4} + \frac{\zeta_K^2 K^2}{n\sigma_n^4} \right) = O \left(\frac{1}{K} + \frac{\zeta_K^2}{n} \right) = o(1)$$

by Assumption A6**. Consequently, all conditions in Theorem 1 of Hall (1984) are satisfied and we can

conclude $\sigma_n^{-1}T_{n11,1a} \xrightarrow{d} N(0, 1)$. Next, write

$$D_n \equiv \sigma_n^{-1}(T_{n11,1b} - \mathbb{B}_n) = \frac{\sigma_n^{-1}}{n} \sum_{i=1}^n \phi_{W_i}^K(U_i)' \Omega'_\tau E(\Upsilon_1) \Omega_\tau \phi_{W_i}^K(U_i) [\psi_\tau(\varepsilon_i)^2 - \tau(1 - \tau)].$$

By straightforward moment calculations, $E(D_n) = 0$ and $\text{Var}(D_n^2) = \sigma_n^{-2}O(\zeta_K^2 K/n) = o(1)$. Hence $D_n = o_P(1)$.

Now we write $\sigma_n^{-1}T_{n11,1c}$ as follows

$$\begin{aligned} \sigma_n^{-1}T_{n11,1c} &= \frac{\sigma_n^{-1}}{n^2} \sum_{1 \leq i \neq j \neq k \leq n} \phi_{W_j}^K(U_j) \psi_\tau(\varepsilon_j) \Omega'_\tau \tilde{\Upsilon}_i \Omega_\tau \phi_{W_k}^K(U_k) \psi_\tau(\varepsilon_k) \\ &\quad + \frac{2\sigma_n^{-1}}{n^2} \sum_{1 \leq j \neq k \leq n} \phi_{W_j}^K(U_j) \psi_\tau(\varepsilon_j) \Omega'_\tau \tilde{\Upsilon}_j \Omega_\tau \phi_{W_k}^K(U_k) \psi_\tau(\varepsilon_k) \\ &\quad + \frac{\sigma_n^{-1}}{n^2} \sum_{1 \leq i \neq j \leq n} \phi_{W_j}^K(U_j) \Omega'_\tau \tilde{\Upsilon}_i \Omega_\tau \phi_{W_j}^K(U_j) \psi_\tau(\varepsilon_j)^2 + \frac{\sigma_n^{-1}}{n^2} \sum_{j=1}^n \phi_{W_j}^K(U_j) \Omega'_\tau \tilde{\Upsilon}_j \Omega_\tau \phi_{W_j}^K(U_j) \psi_\tau(\varepsilon_j)^2 \\ &\equiv \mathbb{R}_{n1} + \mathbb{R}_{n2} + \mathbb{R}_{n3} + \mathbb{R}_{n4}, \text{ say.} \end{aligned}$$

By straightforward moment calculations, $E(\mathbb{R}_{n1}^2) = O(K/(\sigma_n^2 n))$, $E(\mathbb{R}_{n2}^2) = O(K^2/(\sigma_n^2 n^2))$, $E(\mathbb{R}_{n3}^2) = O(K/(\sigma_n^2 n))$, and $E(\mathbb{R}_{n4}^2) = O(K^2/(\sigma_n^2 n^2))$. It follows that $\sigma_n^{-1}T_{n11,1c} = O((K^{1/2}/n)^{1/2}) = o_P(1)$ by Chebyshev's inequality. Consequently, we have proved that $\sigma_n^{-1}(T_{n11,1} - \mathbb{B}_n) \xrightarrow{d} N(0, 1)$.

We now prove (ii). Using arguments as used in the proof of (A.11), we can readily show that $\sigma_n^{-1}T_{n11,2} = \sigma_n^{-1} \sum_{i=1}^n r'_{2i} r_{2i} a_i = \sigma_n^{-1} O_P(nK^{-2\lambda/d}) = O_P(nK^{-(1/2+2\lambda/d)}) = o_P(1)$. By Cauchy-Schwarz inequality,

$$\begin{aligned} |\sigma_n^{-1}T_{n11,3}| &\leq 2\sigma_n^{-1} \left\{ R'_\tau \sum_{i=1}^n \Pi(U_i)' \mathbb{S}' \mathbb{S} \Pi(U_i) R_\tau a_i + \sum_{i=1}^n \mathbf{B}_\tau(U_i)' \mathbb{S}' \mathbb{S} \mathbf{B}_\tau(U_i) a_i \right\} \\ &\leq 2\sigma_n^{-1} \left\{ \lambda_{\max} \left(n^{-1} \sum_{i=1}^n \Pi(U_i)' \mathbb{S}' \mathbb{S} \Pi(U_i) a_i \right) (n \|R_\tau\|^2) + \|\mathbb{S}\|^2 \sum_{i=1}^n \|\mathbf{B}_\tau(U_i)\|^2 a_i \right\} \\ &= \sigma_n^{-1} \left\{ O_P(1) o_P(1) + O_P(nK^{-2\lambda/d}) \right\} = o_P(\sigma_n^{-1}) + O_P(nK^{-(1/2+2\lambda/d)}) = o_P(1). \end{aligned}$$

By Cauchy-Schwarz inequality, $\sigma_n^{-1}T_{n11,6} \leq \{\sigma_n^{-1}T_{n11,2}\}^{1/2} \{\sigma_n^{-1}T_{n11,3}\}^{1/2} = o_P(1) o_P(1) = o_P(1)$. If one also assumes that $nK^{-2\lambda/d} = o(1)$, one can use the same inequality to demonstrate that $\sigma_n^{-1}T_{n11,s} = o(1)$ for $s = 4, 5$ because $\sigma_n^{-1}T_{n11,1} = \sigma_n^{-1}(T_{n11,1} - \mathbb{B}_n) + \sigma_n^{-1}\mathbb{B}_n = O_P(1) + O_P(K^{1/2}) = O_P(\sigma_n)$ by noting that $\mathbb{B}_n = O_P(K)$ and $\sigma_n^{-1} = O(K^{-1/2})$ (see (B.3)). But we only assume that $nK^{-(1/2+2\lambda/d)} = o(1)$ and needs to prove $\sigma_n^{-1}T_{n11,s} = o(1)$ for $s = 4, 5$ via another method. Fortunately, we can prove these claims by straightforward moment calculations and Chebyshev inequality under Assumption A6**. This completes the proof of (ii). ■

Proposition B.2 *Suppose that the conditions in Theorem 4.3 hold. Then $\sigma_n^{-1}T_{n12} = \mu_0 + o_P(1)$.*

Proof. Under $\mathbb{H}_1(\sigma_n^{1/2}n^{-1/2})$, $\bar{\delta}_{1\tau} = n^{-1} \sum_{i=1}^n \delta_{1\tau}(U_i) = \delta_{1\tau} + \sigma_n^{1/2}n^{-1/2} \bar{\Delta}_n$, where $\bar{\Delta}_n = n^{-1} \sum_{i=1}^n \Delta_{\tau n}(U_i) = E[\Delta_{\tau n}(U_i)] + O_P(n^{-1/2})$. It follows that $\sigma_n^{-1}T_{n12} = n^{-1} \sum_{i=1}^n \|\Delta_{\tau n}(U_i) - \bar{\Delta}_n\|^2 \xrightarrow{P} \lim_{n \rightarrow \infty} E[\|\Delta_{\tau n}(U_i) - E[\Delta_{\tau n}(U_i)]\|^2] = \mu_0$. ■

Proposition B.3 *Suppose that the conditions in Theorem 4.3 hold. Then $\sigma_n^{-1}T_{n13} = o_P(1)$.*

Proof. Under $\mathbb{H}_1(\sigma_n^{1/2}n^{-1/2})$, we have

$$\begin{aligned}\sigma_n^{-1}T_{n13} &= \sigma_n^{-1/2}n^{-1/2}\sum_{i=1}^n[\mathbb{S}\Pi(U_i)\mathbf{V}_\tau + \mathbb{S}\mathbf{B}_\tau(U_i)]'[\Delta_{\tau n}(U_i) - \bar{\Delta}_{\tau n}]a_i \\ &= \sigma_n^{-1/2}n^{-1/2}\sum_{i=1}^n\mathbf{V}'_\tau\Pi(U_i)'\mathbb{S}[\Delta_{\tau n}(U_i) - \bar{\Delta}_{\tau n}]a_i + \sigma_n^{-1/2}n^{-1/2}\sum_{i=1}^n\mathbf{B}'_\tau(U_i)'\mathbb{S}'[\Delta_{\tau n}(U_i) - \bar{\Delta}_{\tau n}]a_i \\ &\equiv T_{n13,a} + T_{n13,b}, \text{ say.}\end{aligned}$$

We can bound $T_{n13,b}$ directly: $|T_{n13,b}| \leq \sigma_n^{-1/2}n^{-1/2}\sup_u\|\mathbf{B}_\tau(u)\|\|\mathbb{S}\|\sum_{i=1}^n\|\Delta_{\tau n}(U_i) - \bar{\Delta}_{\tau n}\|a_i = O_P(n^{1/2}K^{-(\lambda/d+1/4)}) = o_P(1)$ by Assumption A6**. For $T_{n13,a}$, we have

$$\begin{aligned}T_{n13,a} &= \sigma_n^{-1/2}n^{-3/2}\sum_{j=1}^n\sum_{i=1}^n\phi_{W_j}^K(U_j)'\psi_\tau(\varepsilon_j)\Omega'_\tau\Pi(U_i)'\mathbb{S}[\Delta_{\tau n}(U_i) - \bar{\Delta}_{\tau n}]a_i \\ &\quad + \sigma_n^{-1/2}n^{-3/2}\sum_{j=1}^n\sum_{i=1}^n\phi_{W_j}^K(U_j)'\psi_\tau(\varepsilon_j - v_j(A_\tau, \Theta_\tau)) - \psi_\tau(\varepsilon_j)]\Omega'_\tau\Pi(U_i)'\mathbb{S}[\Delta_{\tau n}(U_i) - \bar{\Delta}_{\tau n}]a_i \\ &\quad + \sigma_n^{-1/2}n^{-1/2}R'_\tau\sum_{i=1}^n\Pi(U_i)'\mathbb{S}[\Delta_{\tau n}(U_i) - \bar{\Delta}_{\tau n}]a_i \\ &\equiv T_{n13,a1} + T_{n13,a2} + T_{n13,a3}, \text{ say.}\end{aligned}$$

Noting that $E[T_{n13,a1}]^2 = O(\sigma_n^{-1}) = O(K^{-1/2})$, we have $T_{n13,a1} = o_P(1)$ by Chebyshev inequality. Using similar arguments to those used in the proof of (A.11), we can readily show that $T_{n13,a2} = o_P(1)$ and $T_{n13,a3} = o_P(1)$. Consequently we have $\sigma_n^{-1}T_{n13} = o_P(1)$. ■

Proposition B.4 *Suppose that the conditions in Theorem 4.3 hold. Then $\sigma_n^{-1}T_{n2} = o_P(1)$.*

Proof. Note that

$$\begin{aligned}\sqrt{n}(\bar{\delta}_{1\tau} - \delta_{1\tau}) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n[\hat{\delta}_{1\tau}(U_i) - \delta_{1\tau}(U_i)] \\ &= \frac{1}{\sqrt{n}}\mathbb{S}\sum_{i=1}^n\Pi(U_i)\mathbf{V}_\tau + \frac{1}{\sqrt{n}}\mathbb{S}\sum_{i=1}^n\mathbf{B}_\tau(U_i) \equiv S_{1n} + S_{2n}, \text{ say.}\end{aligned}$$

Using arguments similar to those used in the proof of Theorem 4.1, we can show $S_{1n} = O_P(1)$ and $S_{2n} = O_P(n^{1/2}K^{-\lambda/d})$. Noting that $\sigma_n^{-1} = O(K^{-1/2})$ by (B.3), we then have $\sigma_n^{-1}T_{n2} = \sigma_n^{-1}n\bar{a}\|\bar{\delta}_{1\tau} - \delta_{1\tau}\|^2 \leq 2\sigma_n^{-1}\bar{a}(S_{1n}^2 + S_{2n}^2) = O(K^{-1/2})[O_P(1) + O_P(nK^{-2\lambda/d})] = o_P(1)$ by Assumption A6**. ■

Proposition B.5 *Suppose that the conditions in Theorem 4.3 hold. Then $\sigma_n^{-1}T_{n3} = o_P(1)$.*

Proof. By Minkowski inequality,

$$\left\|\sum_{i=1}^n[\hat{\delta}_{1\tau}(U_i) - \bar{\delta}_{1\tau}]a_i\right\| \leq \left\|\sum_{i=1}^n[\hat{\delta}_{1\tau}(U_i) - \delta_{1\tau}(U_i)]a_i\right\| + \frac{1}{n}\left\|\sum_{i=1}^n\sum_{j=1}^n[\delta_{1\tau}(U_i) - \delta_{1\tau}(U_j)]a_i\right\|.$$

As in the proof of Proposition B.4, one can readily show the first term on the right hand side of the last expression is $O_P(n^{1/2} + nK^{-\lambda/d})$. The second term is $\mathbb{H}_1(\sigma_n^{1/2}n^{1/2})$ under $\mathbb{H}_1(\sigma_n^{1/2}n^{-1/2})$. It follows that

$$\begin{aligned}\sigma_n^{-1}|T_{n3}| &= \sigma_n^{-1}\left\|\sqrt{n}(\bar{\delta}_{1\tau} - \delta_{1\tau})\right\|\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n[\hat{\delta}_{1\tau}(U_i) - \bar{\delta}_{1\tau}]a_i\right\| \\ &\leq \sigma_n^{-1}O_P\left(1 + n^{1/2}K^{-\lambda/d}\right)O_P(1 + n^{1/2}K^{-\lambda/d} + \sigma_n^{1/2}) = o_P(1)\end{aligned}$$

as $\sigma_n^{-1} = O(K^{-1/2})$ by (B.3) and $nK^{-(2\lambda/d+1/2)} = o(1)$ by Assumption A6**. ■

Proposition B.6 *Suppose that the conditions in Theorem 4.3 hold. Then $\sigma_n^{-1}(\hat{\mathbb{B}}_n - \mathbb{B}_n) = o_P(1)$.*

Proof. Observe that $\hat{\mathbb{B}}_n - \mathbb{B}_n = \tau(1 - \tau)\text{tr}\{[\hat{\Omega}'_\tau \hat{\Upsilon} \hat{\Omega}_\tau - \Omega'_\tau E(\Upsilon_1)\Omega_\tau] \hat{\Psi}_K\}$. Following and strengthening the proof of Theorem 3 in Powell (1991), we can show that both $\|\hat{\Phi}_K - \Phi_K\|$ and $\|\hat{J}_{K,A} - J_{K,A}\|$ are $O_P[K/(nh)^{1/2}]$. Then

$$\begin{aligned} \left\| \hat{\Phi}_K^{-1} - \Phi_K^{-1} \right\| &= \left\| \hat{\Phi}_K^{-1} (\hat{\Phi}_K - \Phi_K) \Phi_K^{-1} \right\| = \left\{ \text{tr} \left[\hat{\Phi}_K^{-1} (\hat{\Phi}_K - \Phi_K) \Phi_K^{-1} \Phi_K^{-1} (\hat{\Phi}_K - \Phi_K) \hat{\Phi}_K^{-1} \right] \right\}^{1/2} \\ &\leq [\lambda_{\min}(\Phi_K)]^{-1} [\lambda_{\min}(\hat{\Phi}_K)]^{-1} \left\| \hat{\Phi}_K - \Phi_K \right\| \\ &= O(1) O_P(1) O_P[K/(nh)^{1/2}] = O_P[K/(nh)^{1/2}] \end{aligned}$$

as $\lambda_{\min}(\hat{\Phi}_K) \geq \lambda_{\min}(\Phi_K)/2 + o_P(1)$ w.p.a.1. using similar arguments to those in the the proof of Lemma A.1(ii). With this, it is easy to show that $\|\hat{\Omega}_\tau - \Omega_\tau\| = O_P[K/(nh)^{1/2}]$. In addition, $\|\hat{\Upsilon} - E(\Upsilon_1)\| = O_P(K/n^{1/2})$ by Chebyshev inequality. It follows that

$$\left\| \hat{\Omega}'_\tau \hat{\Upsilon} \hat{\Omega}_\tau - \Omega'_\tau E(\Upsilon_1)\Omega_\tau \right\| = O_P[K/(nh)^{1/2}], \quad (\text{B.4})$$

and $|\hat{\mathbb{B}}_n - \mathbb{B}_n| \leq 2 \left\| \hat{\Omega}'_\tau \hat{\Upsilon} \hat{\Omega}_\tau - \Omega'_\tau E(\Upsilon_1)\Omega_\tau \right\| \left\| \hat{\Psi}_K \right\| = O_P[K^{3/2}/(nh)^{1/2}]$. Finally, in view of the fact that $\sigma_n^{-1} = O(K^{-1/2})$, we have $\sigma_n^{-1}(\hat{\mathbb{B}}_n - \mathbb{B}_n) = O_P[K/(nh)^{1/2}] = o_P(1)$ by Assumption A8. ■

Proposition B.7 *Suppose that the conditions in Theorem 4.3 hold. Then $(\hat{\sigma}_n - \sigma_n)/\sigma_n = o_P(1)$.*

Proof. Recall $\bar{\Omega}_\tau = \Omega'_\tau E(\Upsilon_1)\Omega_\tau$. Let $\hat{\bar{\Omega}}_\tau \equiv \hat{\Omega}'_\tau \hat{\Upsilon} \hat{\Omega}_\tau$. Then by the triangle inequality,

$$\begin{aligned} |\hat{\sigma}_n^2 - \sigma_n^2|/[2\tau^2(1 - \tau)^2] &= \left| \text{tr} \left(\hat{\bar{\Omega}}_\tau \hat{\Psi}_K \hat{\bar{\Omega}}_\tau \hat{\Psi}_K - \bar{\Omega}_\tau \Psi_K \bar{\Omega}_\tau \Psi_K \right) \right| \\ &\leq \left| \text{tr} \left((\hat{\bar{\Omega}}_\tau - \bar{\Omega}_\tau) \hat{\Psi}_K \hat{\bar{\Omega}}_\tau \hat{\Psi}_K \right) \right| + \left| \text{tr} \left(\bar{\Omega}_\tau (\hat{\Psi}_K - \Psi_K) \bar{\Omega}_\tau \Psi_K \right) \right| \\ &\quad + \left| \text{tr} \left(\bar{\Omega}_\tau \Psi_K (\hat{\bar{\Omega}}_\tau - \bar{\Omega}_\tau) \hat{\Psi}_K \right) \right| + \left| \text{tr} \left(\bar{\Omega}_\tau \Psi_K \bar{\Omega}_\tau (\hat{\Psi}_K - \Psi_K) \right) \right| \\ &\equiv \sigma_{n1}^2 + \sigma_{n2}^2 + \sigma_{n3}^2 + \sigma_{n4}^2, \text{ say.} \end{aligned}$$

Noting that $\left\| \hat{\Psi}_K \hat{\bar{\Omega}}_\tau \hat{\Psi}_K \right\|^2 = \text{tr}(\hat{\Phi}_\tau \hat{\bar{\Omega}}_\tau \hat{\Phi}_\tau \hat{\bar{\Omega}}_\tau \hat{\Phi}_\tau) \leq [\lambda_{\max}(\hat{\Psi}_K)]^4 \left\| \hat{\bar{\Omega}}_\tau \right\|^2 = O_P(K)$, using the matrix Cauchy-Schwarz inequality and (B.4) yields: $\sigma_{n1}^2 \leq \|\hat{\bar{\Omega}}_\tau - \bar{\Omega}_\tau\| \|\hat{\Psi}_K \hat{\bar{\Omega}}_\tau \hat{\Psi}_K\| = O_P[K/(nh)^{1/2}] O_P(K^{1/2}) = O_P[K^{3/2}/(nh)^{1/2}]$. Similarly, we can show that $\sigma_{n3}^2 = O_P[K^{3/2}/(nh)^{1/2}]$ and $\sigma_{ns}^2 = K^{1/2} O_P(K/n^{1/2})$ for $s = 2$ and 4 as $\|\hat{\Psi}_K - \Psi_K\| = O_P(K/n^{1/2})$ by Lemma A.1(iii). It follows that $\hat{\sigma}_n^2 - \sigma_n^2 = O_P[K^{3/2}/(nh)^{1/2}]$. This result, in conjunction with the fact that $\sigma_n^{-2} = O(1/K)$ by (B.3) and Assumption A8, implies that $\frac{\hat{\sigma}_n - \sigma_n}{\sigma_n} = \frac{\hat{\sigma}_n^2 - \sigma_n^2}{\sigma_n(\hat{\sigma}_n + \sigma_n)} \leq \frac{\hat{\sigma}_n^2 - \sigma_n^2}{\sigma_n^2} = O_P[K^{1/2}/(nh)^{1/2}] = o_P(1)$. ■

Proof of Theorem 4.4. Using the notation defined in the proof of Theorem 4.3, we have $n^{-1}T_n = n^{-1}(T_{n1} + T_{n2} - 2T_{n3})$. Under \mathbb{H}_1 , it is easy to show that $n^{-1}T_{n2} = o_P(1)$, $n^{-1}T_{n3} = o_P(1)$, and

$$\begin{aligned} n^{-1}T_{n1} &= n^{-1}T_{n12} + o_P(1) = n^{-1} \sum_{i=1}^n [\delta_{1\tau}(U_i) - \bar{\delta}_{1\tau}]' [\delta_{1\tau}(U_i) - \bar{\delta}_{1\tau}] a_i + o_P(1) \\ &= E[\|\delta_{1\tau}(U_i) - E[\delta_{1\tau}(U_i)]\|^2 a_i] + o_P(1) = \mu_A + o_P(1). \end{aligned}$$

On the other hand, $n^{-1}\hat{\mathbb{B}}_n = O_P(K/n) = o_P(1)$ and $\hat{\sigma}_n^2 = \sigma_0^2 + o_P(1)$. It follows that $n^{-1}\hat{\sigma}_n \hat{T}_n = n^{-1}T_n - n^{-1}\hat{\mathbb{B}}_n \xrightarrow{P} \mu_A$. The conclusion follows as $(\hat{\sigma}_n - \sigma_n)/\sigma_n = o_P(1)$. ■

Proof of Theorem 4.5. Let P^* denote the probability measure induced by the wild bootstrap conditional on the original sample \mathcal{D}_n . Let $O_{P^*}(\cdot)$ and $o_{P^*}(\cdot)$ denote the probability order under P^* ; e.g., $b_n = o_{P^*}(1)$ if for any $\epsilon > 0$, $P^*(\|b_n\| > \epsilon) = o_P(1)$. Let $\hat{\delta}_{1\tau}^*(\cdot)$, $\bar{\delta}_{1\tau}^*$, T_n^* , Ω_τ^* , \mathbb{B}_n^* , σ_n^* , $\hat{\mathbb{B}}_n^*$ and $\hat{\sigma}_n^{*2}$ denote the bootstrap analogues of $\hat{\delta}_{1\tau}(\cdot)$, $\bar{\delta}_{1\tau}$, T_n , Ω_τ , \mathbb{B}_n , σ_n , $\hat{\mathbb{B}}_n$ and $\hat{\sigma}_n^2$, respectively. Their definitions are self-evident. For example,

$$\mathbb{B}_n^* \equiv \frac{1}{n} \sum_{i=1}^n \varphi_n^*(\zeta_i^*, \zeta_i^*), \text{ and } \sigma_n^{*2} \equiv 2\tau^2 (1 - \tau)^2 \text{tr} \left\{ \bar{\Omega}_\tau^* \hat{\Psi}_K \bar{\Omega}_\tau^* \hat{\Psi}_K \right\},$$

where $\bar{\Omega}_\tau^* \equiv \Omega_\tau^{*'} \hat{\Upsilon} \Omega_\tau^*$, $\zeta_i^* \equiv (U_i', W_i', \varepsilon_i^*)'$, and $\varphi_n^*(\zeta_j^*, \zeta_k^*) \equiv \psi_\tau(\varepsilon_j^*) \phi_{W_j}^K(U_j)' \Omega_\tau^{*'} \hat{\Upsilon} \Omega_\tau^* \phi_{W_k}^K(U_k) \psi_\tau(\varepsilon_k^*)$. The proof follows closely from that of Theorem 4.3.

Note that $\hat{\delta}_{1\tau} = n^{-1} \sum_{i=1}^n \hat{\delta}_{1\tau}(U_i)$ in the bootstrap world plays the role of $\delta_1(\cdot)$ in the real data world. Let $a_i^* \equiv a(U_i^*)$ and $\bar{a}^* \equiv n^{-1} \sum_{i=1}^n a_i^*$. The decomposition of T_n in (B.1) continues to hold for T_n^* in the bootstrap world:

$$T_n^* = \sum_{i=1}^n [\hat{\delta}_{1\tau}^*(U_i^*) - \hat{\delta}_{1\tau} + \hat{\delta}_{1\tau} - \bar{\delta}_{1\tau}^*]' [\hat{\delta}_{1\tau}^*(U_i^*) - \hat{\delta}_{1\tau} + \hat{\delta}_{1\tau} - \bar{\delta}_{1\tau}^*] a_i^* = T_{n1}^* + T_{n2}^* - 2T_{n3}^*,$$

where $T_{n1}^* = \sum_{i=1}^n [\hat{\delta}_{1\tau}^*(U_i^*) - \hat{\delta}_{1\tau}]' [\hat{\delta}_{1\tau}^*(U_i^*) - \hat{\delta}_{1\tau}] a_i^*$, $T_{n2}^* = n\bar{a}^* (\hat{\delta}_{1\tau} - \bar{\delta}_{1\tau}^*)' (\hat{\delta}_{1\tau} - \bar{\delta}_{1\tau}^*)$, and $T_{n3}^* = (\hat{\delta}_{1\tau} - \bar{\delta}_{1\tau}^*)' \sum_{i=1}^n [\hat{\delta}_{1\tau}^*(U_i^*) - \hat{\delta}_{1\tau}] a_i^*$. We prove the first part of the theorem by showing that (i) $(T_{n1}^* - \mathbb{B}_n^*) / \sigma_n^* \rightarrow N(0, 1)$ in distribution in probability, (ii) $T_{n2}^* / \sigma_n^* = o_{P^*}(1)$, (iii) $T_{n3}^* / \sigma_n^* = o_{P^*}(1)$, (iv) $(\hat{\mathbb{B}}_n^* - \mathbb{B}_n^*) / \sigma_n^* = o_{P^*}(1)$, and (v) $(\hat{\sigma}_n^* - \sigma_n^*) / \sigma_n^* = o_{P^*}(1)$. The proofs of (ii)-(v) parallel those of Propositions B.4-B.7, respectively, and thus omitted. We only sketch the proof of (i).

In view of the fact that $Y_i^* = \hat{\alpha}_\tau' D_i + \hat{\beta}_\tau' X_i + \varepsilon_i^*$ and the first element of the vector of basis functions $(p^K(\cdot))$ is 1, the bootstrap analogues of $\mathbf{B}_\tau(U_i)$ and $v_i(A_\tau, \Theta_\tau) = P_K(D_i, U_i)' A_\tau + \phi_{W_i}^K(U_i)' \Theta_\tau - \alpha_\tau(U_i)' D_i - \beta_\tau(U_i)' X_i$ are both 0. This implies that that the bootstrap analogue of r_{2i} defined in the proof of Lemma B.1 is 0 and that of r_{3i} can be simplified. Following the proof of Lemmas B.1 and B.3, we can show that $T_{n1}^* = \bar{T}_{n1}^* + o_{P^*}(\sigma_n^*)$, where $\bar{T}_{n1}^* = \frac{2}{n} \sum_{1 \leq i < j \leq n} \varphi_n^*(\zeta_i^*, \zeta_j^*)$. As $\bar{T}_{n1}^* / \sigma_n^*$ is a second order degenerate U -statistic with *independent but non-identically distributed (inid)*, we can apply the CLT for second order degenerate U -statistic for *inid* observations (e.g., de Jong, 1987) and conclude that $\bar{T}_{n1}^* / \sigma_n^* \rightarrow N(0, 1)$ in distribution in probability. Then (i) follows. This completes the first part of the theorem.

Parts (ii)-(iv) of Theorem 4.5 follow from the first part and Theorems 4.2-4.4, respectively. ■

References

- Banks, J., Blundell, R., Lewbel, A., 1997. Quadratic Engel curves and consumer demand. *Review of Economics and Statistics* 79, 527-539.
- Bernstein, D. S., 2005. *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory*. Princeton University Press, Princeton.
- Blundell, R., Duncan, A., Pendakur, K., 1998. Semiparametric estimation and consumer demand. *Journal of Applied Econometrics* 13, 435-461.
- Blundell, R., Browning, M., Crawford, I. A., 2003. Nonparametric Engel curves and revealed preference. *Econometrica* 71, 205-240.
- Blundell, R., Chen, X., Kristensen, D., 2007. Semi-nonparametric IV estimation of shape-invariant Engel curves. *Econometrica* 75, 1613-1669.
- Cai, Z., Das, M., Xiong, H., Wu, X., 2006. Functional coefficient instrumental variables models. *Journal of Econometrics* 133, 207-241.

- Cai, Z., Fan, J., Yao, Q., 2000. Functional-coefficient regression models for nonlinear time series. *Journal of American Statistical Association* 95, 941–956.
- Cai, Z., Fabf, Y., Lin, M., Su, J., 2010. Semiparametric varying-coefficient instrumental variables models. Working paper, University of North Carolina at Charlotte.
- Cai, Z., Li, Q., 2008. Nonparametric estimation of varying coefficient dynamic panel data models. *Econometric Theory* 24, 1321-1342.
- Cai, Z., Xiao, Z., 2012. Semiparametric quantile regression estimation in dynamic models with partially varying coefficients. *Journal of Econometrics* 167, 413-425.
- Cai, Z., Xiong, H., 2012. Partially varying coefficient instrumental variables models. *Statistica Neerlandica* 66, 85-110.
- Cai, Z., Xu, X., 2008. Nonparametric quantile estimations for dynamic smooth coefficient models. *Journal of the American Statistical Association* 103, 1596-1908.
- Card, D., 2001. Estimating the return to schooling: progress on some persistent econometric problems. *Econometrica* 69, 1127-1160.
- Chen, R., Tsay, R. S., 1993. Functional-coefficient autoregressive models. *Journal of the American Statistical Association* 88, 298-308.
- Chen, X., 2007. Large sample sieve estimation of semi-nonparametric models. In: Heckman, J. J., Leamer, E. (Eds.), *Handbook of Econometrics*, 6B (Chapter 76), 5549-5632. New York: Elsevier Science.
- Chen, X., Pouzo, D., 2009. Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals. *Journal of Econometrics* 152, 46-60.
- Chen, X., Pouzo, D., 2012. Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals. *Econometrica* 80, 277-321.
- Chernozhukov, V., Fernández-Val, I., Galichon, A., 2010. Quantile and probability curves without crossing. *Econometrica* 78, 1093-1125.
- Chernozhukov, V., Fernández-Val, I., Kowalski, A. E., 2015. Quantile regression with censoring and endogeneity. *Journal of Econometrics*, forthcoming.
- Chernozhukov, V., Hansen, C., 2005. An IV model of quantile treatment effects. *Econometrica*, 73, 245-261.
- Chernozhukov, V., Hansen, C., 2006. Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics*, 127, 491-525.
- Chernozhukov, V., Hansen, C., 2008. Instrumental variable quantile regression: a robust inference approach. *Journal of Econometrics* 142, 379-398.
- Chernozhukov, V., Hansen, C., Jansson, M., 2009. Finite sample inference for quantile regression models. *Journal of Econometrics* 152, 93-103.
- Chernozhukov, V., Hong, H., 2002. Three-step quantile regression and extramarital affairs. *Journal of the American Statistical Association* 97, 872-882.
- Das, M., 2005. Instrumental variable estimators for nonparametric models with discrete endogenous regressors. *Journal of Econometrics* 124, 335-361
- Davidson, J., 1994. *Stochastic Limit Theory*. Oxford University Press, Oxford.
- de Jong, P., 1987. A central limit theorem for generalized quadratic forms. *Probability Theory and Related Fields* 75, 261-277.
- Fan, J., Huang, T., 2005. Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli* 11, 1031-1057.
- Fan, J., Zhang, W., 1999. Statistical estimation in varying coefficient models. *Annals of Statistics* 27, 1491–1518.
- Fan, J., Zhang, C., Zhang, J., 2001. Generalized likelihood ratio statistics and Wilks phenomenon. *Annals of Statistics* 29, 153-193.
- Feng, X., He, X., Hu, J., 2011. Wild bootstrap for quantile regression. *Biometrika* 98, 995-999.

- Gallant, A. R., White, H., 1988. *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Basil Blackwell, New York.
- Galvao, A. F., Montes-Rojas, G. V., 2010. Penalized quantile regression for dynamic panel data. *Journal of Statistical Planning and Inference* 140, 3476-3497.
- Galvao, A. F., 2011. Quantile regression for dynamic panel data with fixed effects. *Journal of Econometrics* 164, 142-157.
- Giné, E., Zinn, J., 1990. Bootstrapping general empirical measures. *Annals of Probability* 18, 851-869.
- Hall, P., 1984. Central limit theorem for integrated square error properties of multivariate nonparametric density estimators. *Journal of Multivariate Analysis* 14, 1-16.
- Hansen, B. E., 2000. Testing for structural change in conditional models. *Journal of Econometrics* 97, 93-115.
- Harding, M., Lamarche, C., 2009. A quantile regression approach for estimating panel data models using instrumental variables. *Economics Letters* 104, 133-135.
- Harding, M., Lamarche, C., 2012. Quantile regression estimation of panel duration models with censored data. *Advances in Econometrics* 29, 237-267.
- Harding, M., Lamarche, C., 2014. Estimating and testing a quantile regression model with interactive effects. *Journal of Econometrics* 178, 101-113.
- Härdle, W., Mammen, E., 1993. Comparing nonparametric versus parametric regression fits. *Annals of Statistics* 21, 1926-1947.
- Hastie, T. J., Tibshirani, R. J., 1993. Varying-coefficient models (with Discussion). *Journal of the Royal Statistical Society, Series B*. 55, 757-796.
- He, X., Shao, Q., 2000. On parameters of increasing dimensions. *Journal of Multivariate Analysis* 73, 120-135.
- Hogg, R. V., Craig, A. T., 1995. *Introduction to Mathematical Statistics*. 5th ed., Prentice Hall, New Jersey.
- Honda, T., 2004. Quantile regression in varying coefficient models. *Journal of Statistical Planning and Inferences* 121, 113-125.
- Honoré, B. E., Hu, L., 2004. On the performance of some robust instrumental variable estimators. *Journal of Business & Economic Statistics* 22, 30-39.
- Horowitz, J., Lee, S., 2005. Nonparametric estimation of an additive quantile regression model. *Journal of the American Statistical Association*, 100, 1238-1249.
- Huang, J. Z., 2003. Local asymptotic for polynomial spline regression. *Annals of Statistics* 31, 1600-1635.
- Kaplan, D., Sun, Y., 2012. Smoothed estimating equations for instrumental variable quantile regression. Working paper, Department of Economics, UCSD.
- Kato, K., Galvao, A. F., Montes-Rojas, G. V., 2012. Asymptotics for panel quantile regression models with individual effects. *Journal of Econometrics* 10, 76-91.
- Kim, M.-O., 2007. Quantile regression with varying coefficients. *Annals of Statistics* 35, 92-108.
- Kline, P., Santos, A., 2012. A score based approach to wild bootstrap inference. *Journal of Econometric Methods* 1, 23-41.
- Koenker, R., 2004. Quantile regression for longitudinal data. *Journal of Multivariate Analysis* 91, 74-89.
- Koenker, R., 2005. *Quantile Regressions*. Cambridge University Press, Cambridge.
- Koenker, R., Basset, G., 1978. Regression quantiles. *Econometrica* 46, 33-50.
- Kottaridi, C., Stengos, T., 2010. Foreign direct investment, human capital and nonlinearities in economic growth. *Journal of Macroeconomics* 32, 858-871.
- Li, Q., Hsiao, C., Zinn, J., 2003. Consistent specification tests for semiparametric/nonparametric models based on series estimation method. *Journal of Econometrics* 112, 295-325.

- Li, Q., Racine, J., 2007. *Nonparametric Econometrics: Theory and Practice*. Princeton University Press, Princeton and Oxford.
- Lu, X., Su, L., 2015. Jackknife model averaging for quantile regressions. *Journal of Econometrics*, forthcoming.
- Newey, W. K., 1997. Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics* 79, 147-168.
- Pakes, A., Pollard, D., 1989. Simulation and the asymptotics of optimization estimators. *Econometrica* 57, 1027-1057.
- Powell, D., 2013. A new framework for estimation of quantile treatment effects. Working paper, RAND Labor & Population.
- Powell, J. L., 1991. Estimation of monotonic regression models under quantile restrictions. In: Barnett, W., Powell, J. L., Tauchen, G. (Eds), *Nonparametric and Semiparametric Methods in Econometrics*, 357-384. Cambridge University Press, Cambridge.
- Schultz, T.P., 2003. Human capital, schooling and health. *Economics and Human Biology* 1, 207-221.
- Schumaker, L. L., 2007. *Spline Functions: Basic Theory*, 3rd ed. Cambridge University Press, Cambridge.
- Su, L., Chen, Y., Ullah, A., 2009. Functional coefficient estimation with both categorical and continuous Data. *Advances in Econometrics* 25, 131-167.
- Su, L., Murtazashvili, I. and Ullah, A., 2014. Local linear GMM estimation of functional coefficient IV models with application to the estimation of rate of return to schooling. *Journal of Business & Economic Statistics* 31, 184-207.
- Su, L., White, H., 2010. Testing structural change in partially linear models. *Econometric Theory* 26, 1761-1806.
- Su, L., Yang, Z., 2012. Instrumental variable quantile estimation of spatial autoregressive models. Working Paper, Singapore Management University.
- Sun, Y., 2006. A consistent nonparametric equality test of conditional quantile functions. *Econometric Theory* 22, 614-632.
- Tran, K. C., and E. G. Tsionas, 2010. Local GMM estimation of semiparametric panel data with smooth coefficient models. *Econometric Reviews* 29, 39-61.
- Wang, H. J., Zhu, Z., Zhou, J., 2009. Quantile regression in partially linear varying coefficient models. *Annals of Statistics* 37, 3841-3866.
- White, H., 1994. *Estimation, Inference and Specification Analysis*. Cambridge University Press, Cambridge.

Supplementary Material for “Sieve Instrumental Variable Quantile Regression Estimation of Functional Coefficient Models”

(Not for publication)

Liangjun Su^a, Tadao Hoshino^b

^a *School of Economics, Singapore Management University*

^b *Global Education Center, Waseda University*

This supplementary material provides proofs for the technical lemmas in the above paper. We also present the asymptotic analysis for the panel data models in Section 3.4 and the alternative bootstrap test in Remark 12. More simulation results are reported in Appendix F.

C Proofs of the technical lemmas

Proof of Lemma A.1. (i) For fixed $A \in \mathcal{A}_K$, we can readily follow Newey (1997) and show that $\|\Phi_{n,K}(A) - \Phi_K(A)\| = O_P(\zeta_K K^{1/2}/n^{1/2})$ by Chebyshev inequality. To obtain the uniform result, one can cover the compact set \mathcal{A}_K by a finite number of cubes and apply Boole’s and Bernstein’s inequalities to show the claim.

(ii) Using the arguments as used in the proof of Lemma A.1 in Su and Jin (2012), we have by (i) and Assumption A5(iii),

$$\begin{aligned} \inf_{A \in \mathcal{A}_K} \lambda_{\min}(\Phi_{n,K}(A)) &= \inf_{A \in \mathcal{A}_K} \min_{\|\varkappa\|=1} \{\varkappa' \Phi_K(A) \varkappa + \varkappa' (\Phi_{n,K}(A) - \Phi_K(A)) \varkappa\} \\ &\geq \inf_{A \in \mathcal{A}_K} \lambda_{\min}(\Phi_K(A)) - \sup_{A \in \mathcal{A}_K} \|\Phi_{n,K}(A) - \Phi_K(A)\| \\ &\geq \underline{c}_\Phi - O_P(\zeta_K K^{1/2}(\ln n/n)^{1/2}) \geq \underline{c}_\Phi/2 \text{ w.p.a.1.} \end{aligned}$$

and

$$\begin{aligned} \sup_{A \in \mathcal{A}_K} \lambda_{\max}(\Phi_{n,K}(A)) &= \sup_{A \in \mathcal{A}_K} \max_{\|\varkappa\|=1} \{\varkappa' \Phi_K(A) \varkappa + \varkappa' (\Phi_{n,K}(A) - \Phi_K(A)) \varkappa\} \\ &\leq \sup_{A \in \mathcal{A}_K} \lambda_{\max}(\Phi_K(A)) + \sup_{A \in \mathcal{A}_K} \|\Phi_{n,K}(A) - \Phi_K(A)\| \\ &\leq \bar{c}_\Phi + O_P(\zeta_K K^{1/2}(\ln n/n)^{1/2}) \leq 2\bar{c}_\Phi \text{ w.p.a.1.} \end{aligned}$$

Analogously, we can prove (iii) and (iv). ■

Proof of Lemma A.2. Let $\mathcal{N} = \{1, 2, \dots, n\}$ and \mathcal{H}_{k_2, k_3} denote the collection of all $(k_2 + k_3)K$ -element subsets of \mathcal{N} . Also, let $\mathbf{p}_W(\mathbf{h})$ denote a $(k_2 + k_3)K \times (k_2 + k_3)K$ matrix whose rows are the vectors $\phi_{W_i}^K(U_i)'$ such that $i \in \mathbf{h} \in \mathcal{H}_{k_2, k_3}$, and let $\mathbf{Y}(\mathbf{h}, A)$ denote a $(k_2 + k_3)K \times 1$ vector whose elements are $Y_i - P_K(D_i, U_i)'A$ such that $i \in \mathbf{h}$. By Theorem 3.3 of Koenker and Bassett (1978) (see also Lemma A.2 of Horowitz and Lee (2005)), there uniquely exists $\mathbf{h}^*(A) \in \mathcal{H}_{k_2, k_3}$ such that $\hat{\Theta}_\tau(A) = \mathbf{p}_W(\mathbf{h}^*(A))^{-1} \mathbf{Y}(\mathbf{h}^*(A), A)$ for each $A \in \mathcal{A}_K$, and $\mathbf{H}_n(A, \hat{\Theta}_\tau(A)) \in (\tau - 1, \tau)^{(k_2 + k_3)K}$ (i.e., each

element of $\mathbf{H}_n(A, \hat{\Theta}_\tau(A))$ lies strictly between $\tau - 1$ and τ), where $\mathbf{H}_n(A, \hat{\Theta}_\tau(A)) = \sum_{i \in (\mathbf{h}^*(A))^c}^n \psi_\tau[\varepsilon_i - v_i(A, \hat{\Theta}_\tau(A))] \phi_{W_i}^K(U_i)' \mathbf{p}_W(\mathbf{h}^*(A))^{-1}$ and $(\mathbf{h}^*(A))^c = \mathcal{N} \setminus \mathbf{h}^*(A)$.

Write $n^{-1} \sum_{i=1}^n g_i(A, \hat{\Theta}_\tau(A)) = G_{1n}(A, \hat{\Theta}_\tau(A)) + G_{2n}(A, \hat{\Theta}_\tau(A))$, where $G_{1n}(A, \hat{\Theta}_\tau(A)) = n^{-1} \sum_{i \in \mathbf{h}^*(A)} \phi_{W_i}^K(U_i) \psi_\tau[\varepsilon_i - v_i(A, \hat{\Theta}_\tau(A))]$ and $G_{2n}(A, \hat{\Theta}_\tau(A)) = n^{-1} \sum_{i \in (\mathbf{h}^*(A))^c} \phi_{W_i}^K(U_i) \psi_\tau[\varepsilon_i - v_i(A, \hat{\Theta}_\tau(A))]$. Under Assumptions A1(i) and A6(i), we have

$$\sup_{A \in \mathcal{A}_K} \left\| G_{1n}(A, \hat{\Theta}_\tau(A)) \right\| \leq n^{-1} (k_2 + k_3) K \sup_{w \in \mathcal{W}} \sup_{u \in \mathcal{U}} \left\| \phi_w^K(u) \right\| = O_P(\zeta_K K/n) = o_P(n^{-1/2}),$$

and

$$\begin{aligned} \sup_{A \in \mathcal{A}_K} \left\| G_{2n}(A, \hat{\Theta}_\tau(A)) \right\| &= n^{-1} \sup_{A \in \mathcal{A}_K} \left\| \mathbf{H}_n(A, \hat{\Theta}_\tau(A)) \mathbf{p}_W(\mathbf{h}^*(A)) \right\| \\ &\leq n^{-1} \sup_{A \in \mathcal{A}_K} \left\| \mathbf{H}_n(A, \hat{\Theta}_\tau(A)) \right\| \sup_{A \in \mathcal{A}_K} \left\| \mathbf{p}_W(\mathbf{h}^*(A)) \right\| \\ &\leq n^{-1} ((k_2 + k_3) K)^{1/2} \cdot O_P(\zeta_K K^{1/2}) = O_P(\zeta_K K/n) = o_P(n^{-1/2}), \end{aligned}$$

where the second inequality follows from the fact that

$$\begin{aligned} \sup_{A \in \mathcal{A}_K} \left\| \mathbf{p}_W(\mathbf{h}^*(A)) \right\|^2 &= \sup_{A \in \mathcal{A}_K} \text{tr}[\mathbf{p}_W(\mathbf{h}^*(A)) \mathbf{p}_W(\mathbf{h}^*(A))'] = \sup_{A \in \mathcal{A}_K} \sum_{i \in \mathbf{h}^*(A)} \left\| \phi_{W_i}^K(U_i) \right\|^2 \\ &\leq (k_2 + k_3) K \sup_{w \in \mathcal{W}} \sup_{u \in \mathcal{U}} \left\| \phi_w^K(u) \right\|^2 = O_P(\zeta_K^2 K). \end{aligned}$$

It follows that $\sup_{A \in \mathcal{A}_K} \left\| n^{-1} \sum_{i=1}^n g_i(A, \hat{\Theta}_\tau(A)) \right\| = o_P(n^{-1/2})$ by triangle inequality. ■

Proof of Lemma A.3. Let $G_n(A) \equiv n^{-1} \sum_{i=1}^n g_i(A, \Theta_\tau(A))$. Noting that $E[g_i(A, \Theta_\tau(A))] = 0$ by the first order condition for the minimization problem in (2.7), we have $E[G_n(A)] = 0$ and $\text{Var}(G_n(A)) = O(K/n)$. It follows that $G_n(A) = O_P[(K/n)^{1/2}]$ for each $A \in \mathcal{A}_K$. The uniform result then follows from a standard application of Boole's and Bernstein's inequalities. ■

Proof of Lemma A.4. Let $\chi_i = (U_i', D_i', W_i')'$ and $\ell_n = L(K \ln n/n)^{1/2}$. By the law of iterated expectations, second order Taylor expansions, the fact that $v_i(A, \Theta) - v_i(A, \Theta_\tau(A)) = \phi_{W_i}^K(U_i)' [\Theta - \Theta_\tau(A)]$, and Assumptions A1(i) and (iii), A5(iii), and A6(i),

$$\begin{aligned} &\sup_{\|\mathbf{c}\|=1} \sup_{A \in \mathcal{A}_K} \sup_{\|\Theta - \Theta_\tau(A)\| \leq \ell_n} |\mathbf{c}' E[g_i(A, \Theta) - g_i(A, \Theta_\tau(A))] + \mathbf{c}' \Phi_K(A) [\Theta - \Theta_\tau(A)]| \\ &\leq \sup_{A \in \mathcal{A}_K} \sup_{\|\Theta - \Theta_\tau(A)\| \leq \ell_n} \left\| E \left[\mathbf{1} \{ \varepsilon_i \leq v_i(A, \Theta) \} - \mathbf{1} \{ \varepsilon_i \leq v_i(A, \Theta_\tau(A)) \} \phi_{W_i}^K(U_i) \right] - \Phi_K(A) [\Theta - \Theta_\tau(A)] \right\| \\ &= \sup_{A \in \mathcal{A}_K} \sup_{\|\Theta - \Theta_\tau(A)\| \leq \ell_n} \left\| E \left\{ [F_\varepsilon(v_i(A, \Theta) | \chi_i) - F_\varepsilon(v_i(A, \Theta_\tau(A)) | \chi_i)] \phi_{W_i}^K(U_i) \right\} - \Phi_K(A) [\Theta - \Theta_\tau(A)] \right\| \\ &\leq \frac{1}{2} \bar{c}_{f_\varepsilon} \sup_{A \in \mathcal{A}_K} \sup_{\|\Theta - \Theta_\tau(A)\| \leq \ell_n} E \left\| \left\{ \left[\phi_{W_i}^K(U_i)' [\Theta - \Theta_\tau(A)] \right]^2 \phi_{W_i}^K(U_i) \right\} \right\| \\ &\leq \frac{1}{2} \bar{c}_{f_\varepsilon} \sup_{w \in \mathcal{W}} \sup_{u \in \mathcal{U}} \left\| \phi_w^K(u) \right\| \sup_{A \in \mathcal{A}_K} \sup_{\|\Theta - \Theta_\tau(A)\| \leq \ell_n} [\Theta - \Theta_\tau(A)]' E \left[\phi_{W_i}^K(U_i) \phi_{W_i}^K(U_i)' \right] [\Theta - \Theta_\tau(A)] \\ &\leq \frac{1}{2} \bar{c}_\Phi \bar{c}_{f_\varepsilon} \sup_{w \in \mathcal{W}} \sup_{u \in \mathcal{U}} \left\| \phi_w^K(u) \right\| \ell_n^2 = O_P(\zeta_K K \ln n/n) = o_P(n^{-1/2}). \end{aligned}$$

The result then follows from the IID assumption in Assumption A.1(i). ■

Proof of Lemma A.5. We only prove (i) as the proof of (ii) is similar. Noting that $\sup_{\|\Theta_1 - \Theta_2\| \leq \Delta} |v_i(A, \Theta_1) - v_i(A, \Theta_2)| \leq \left\| \phi_{W_i}^K(U_i) \right\| \Delta$, we have

$$\begin{aligned}
& E \left[\sup_{A \in \mathcal{A}_K} \sup_{\|\Theta_1 - \Theta_2\| \leq \Delta} \|\eta_i(A; \Theta_1, \Theta_2)\|^2 \right] \\
= & E \left\{ \left\| \phi_{W_i}^K(U_i) \right\|^2 \sup_{A \in \mathcal{A}_K} \sup_{\|\Theta_1 - \Theta_2\| \leq \Delta} \left| \mathbf{1} \{ \varepsilon_i \leq v_i(A, \Theta_1) \} - \mathbf{1} \{ \varepsilon_i \leq v_i(A, \Theta_2) \} \right. \right. \\
& \left. \left. - F_\varepsilon(v_i(A, \Theta_1) | \chi_i) + F_\varepsilon(v_i(A, \Theta_2) | \chi_i) \right|^2 \right\} \\
\leq & 2E \left\{ \left\| \phi_{W_i}^K(U_i) \right\|^2 \sup_{A \in \mathcal{A}_K} \sup_{\Theta_2} \left| \mathbf{1} \{ \varepsilon_i \leq \left\| \phi_{W_i}^K(U_i) \right\| \Delta + v_i(A, \Theta_2) \} - \mathbf{1} \{ \varepsilon_i \leq - \left\| \phi_{W_i}^K(U_i) \right\| \Delta + v_i(A, \Theta_2) \} \right. \right. \\
& \left. \left. - F_\varepsilon \left(- \left\| \phi_{W_i}^K(U_i) \right\| \Delta + v_i(A, \Theta_2) | \chi_i \right) + F_\varepsilon \left(\left\| \phi_{W_i}^K(U_i) \right\| \Delta + v_i(A, \Theta_2) | \chi_i \right) \right\} \\
\leq & 2E \left\| \phi_{W_i}^K(U_i) \right\|^2 \sup_{A \in \mathcal{A}_K} \sup_{\Theta} \mathbf{1} \left\{ \|\varepsilon_i - v_i(A, \Theta)\| \leq \left\| \phi_{W_i}^K(U_i) \right\| \Delta \right\} + c\zeta_K E \left[\left\| \phi_{W_i}^K(U_i) \right\|^2 \right] \Delta \\
\leq & 2c\zeta_K E \left\| \phi_{W_i}^K(U_i) \right\|^2 \Delta \leq 2c^2\zeta_K K \Delta \leq n^{1/2} \Delta \text{ for sufficiently large } n,
\end{aligned}$$

where the first inequality follows from the monotonicity of the indicator function $\mathbf{1} \{ \varepsilon \leq \cdot \}$ and the conditional CDF $F_\varepsilon(\cdot | d, u, w)$ and the fact that $|\mathbf{1} \{ \varepsilon_i \leq \cdot \} - F_\varepsilon(\cdot | \chi_i)| \leq 1$, the second inequality follows from Minkowski inequality, Taylor expansion and Assumption A1(iii), and the third equality from the fact that by Assumption A1(iii): $E \left\{ \sup_{A \in \mathcal{A}_K} \sup_{\Theta} \mathbf{1} \left\{ \|\varepsilon_i - v_i(A, \Theta)\| \leq \left\| \phi_{W_i}^K(U_i) \right\| \Delta \right\} \middle| \chi_i \right\} \leq 2\bar{c}_f \left\| \phi_{W_i}^K(U_i) \right\| \Delta$. Here c is a generic large constant. ■

Proof of Lemma A.6. Let $\tilde{v}_i(A, \Theta) = \mathbf{c}' \{ g_i(A, \Theta) - E[g_i(A, \Theta)] \}$. $\tilde{v}_i(A, \Theta_1)$ and $\tilde{v}_i(A, \Theta_2)$ play the respective roles of $v_i(t)$ and $v_i(s)$ in Lemma 3.2 in He and Shao (2000). By Lemma A.5(i), condition (3.2) in that lemma is satisfied for $r_1 = \frac{1}{2}$ and $r_2 = 1$ and thus the conclusion follows. ■

Proof of Lemma A.7. Write

$$\Omega_\tau \Omega'_\tau = \begin{bmatrix} S_{1K} J'_{K,A} \bar{\Phi}'_{K,C} M_K \bar{\Phi}_{K,C} \bar{\Phi}'_{K,C} M_K \bar{\Phi}_{K,C} J_{K,A} S_{1K} & -S_{1K} J'_{K,A} \bar{\Phi}'_{K,C} M_K \bar{\Phi}_{K,C} S'_{2K} \bar{\Phi}'_{K,B} \\ -\bar{\Phi}_{K,B} S_{2K} \bar{\Phi}'_{K,C} M_K \bar{\Phi}_{K,C} J_{K,A} S_{1K} & \bar{\Phi}_{K,B} S_{2K} S'_{2K} \bar{\Phi}'_{K,B} \end{bmatrix}$$

where $S_{1K} = (J'_{K,A} \bar{\Phi}'_{K,C} M_K \bar{\Phi}_{K,C} J_{K,A})^{-1}$, and $S_{2K} = I_{(k_2+k_3)K} - J_{K,A} S_{1K} J'_{K,A} \bar{\Phi}'_{K,C} M_K \bar{\Phi}_{K,C}$. By the fact that $\lambda_{\max} \left(\begin{bmatrix} A & B \\ B' & C \end{bmatrix} \right) \leq \lambda_{\max}(A) + \lambda_{\max}(C)$ for any p.s.d. matrix $\begin{bmatrix} A & B \\ B' & C \end{bmatrix}$ and that $\lambda_{\max}(DED') \leq \lambda_{\max}(E) \lambda_{\max}(DD')$ for any symmetric p.s.d. matrix E , and using Assumptions A5(i)-(v), we have

$$\begin{aligned}
\lambda_{\max}(\Omega_\tau \Omega'_\tau) & \leq \lambda_{\max}(S_{1K} J'_{K,A} \bar{\Phi}'_{K,C} M_K \bar{\Phi}_{K,C} \bar{\Phi}'_{K,C} M_K \bar{\Phi}_{K,C} J_{K,A} S_{1K}) + \lambda_{\max}(\bar{\Phi}_{K,B} S_{2K} S'_{2K} \bar{\Phi}'_{K,B}) \\
& \leq \lambda_{\max}(M_K^{1/2} \bar{\Phi}_{K,C} \bar{\Phi}'_{K,C} M_K^{1/2}) \lambda_{\max}(S_{1K} J'_{K,A} \bar{\Phi}'_{K,C} M_K \bar{\Phi}_{K,C} J_{K,A} S_{1K}) \\
& \quad + \lambda_{\max}(S_{2K} S'_{2K}) \lambda_{\max}(\bar{\Phi}_{K,B} \bar{\Phi}'_{K,B}) \\
& \leq \bar{c}_M \lambda_{\max}(\bar{\Phi}_{K,C} \bar{\Phi}'_{K,C}) \lambda_{\max}(S_{1K}) + \lambda_{\max}(\bar{\Phi}_{K,B} \bar{\Phi}'_{K,B}) \\
& \leq \bar{c}_M \underline{c}_\Phi^{-2} + \underline{c}_\Phi^{-2} \equiv \bar{c}_\Omega < \infty,
\end{aligned}$$

where we have also used the fact that 1) $\lambda_{\max}(S_{2K}S'_{2K}) \leq \lambda_{\max}(S_{2K})^2 = 1$ as S_{2K} is idempotent, and 2) $\lambda_{\max}(\bar{\Phi}_{K,B}\bar{\Phi}'_{K,B}) \leq \lambda_{\max}([\Phi_K(A_\tau)]^{-1}[\Phi_K(A_\tau)]^{-1}) \leq \underline{c}_\Phi^{-2}$ and similarly $\lambda_{\max}(\bar{\Phi}_{K,C}\bar{\Phi}'_{K,C}) \leq \underline{c}_\Phi^{-2}$ (see, e.g., Lemma 8.4.4 in Bernstein (2005)).

By the fact that

$$\begin{bmatrix} A & B \\ B' & C \end{bmatrix} = \begin{bmatrix} A - BC^{-1}B' & B \\ 0 & C \end{bmatrix} \begin{bmatrix} I & 0 \\ C^{-1}B' & I \end{bmatrix}$$

for any square matrices A and C and conformable matrix B such that C is nonsingular (e.g., Fact 2.15.3 in Bernstein (2005)), and the fact that $\lambda_{\min}(DE) \geq \lambda_{\min}(D)\lambda_{\min}(E)$ for any two p.s.d. matrices D and E (see, e.g., Fact 8.14.20 in Bernstein (2005)), we have

$$\lambda_{\min}(\Omega_\tau\Omega'_\tau) \geq \min(\lambda_{\min}(A - BC^{-1}B'), \lambda_{\min}(C))$$

where $A = S_{1K}J'_{K,A}\bar{\Phi}'_{K,C}M_K\bar{\Phi}_{K,C}\bar{\Phi}'_{K,C}M_K\bar{\Phi}_{K,C}J_{K,A}S_{1K}$, $B = -S_{1K}J'_{K,A}\bar{\Phi}'_{K,C}M_K\bar{\Phi}_{K,C}S'_{2K}\bar{\Phi}'_{K,B}$, and $C = \bar{\Phi}_{K,B}S_{2K}S'_{2K}\bar{\Phi}'_{K,B}$. Observe that

$$A - BC^{-1}B' = S_{1K}J'_{K,A}\bar{\Phi}'_{K,C}M_K\bar{\Phi}_{K,C}M_{\bar{\Phi}_{K,B}S_{2K}}\bar{\Phi}'_{K,C}M_K\bar{\Phi}_{K,C}J_{K,A}S_{1K}$$

where $M_{\bar{\Phi}_{K,B}S_{2K}} = I_{(k_2+k_3)K} - S'_{2K}\bar{\Phi}'_{K,B}(\bar{\Phi}_{K,B}S_{2K}S'_{2K}\bar{\Phi}'_{K,B})^{-1}\bar{\Phi}_{K,B}S_{2K}$. Noting that $M_{\bar{\Phi}_{K,B}S_{2K}}$ is a projection matrix and $\text{tr}(M_{\bar{\Phi}_{K,B}S_{2K}}) = k_3K$, by the spectral theorem for symmetric matrices we can write $M_{\bar{\Phi}_{K,B}S_{2K}} = SDS'$ where $D = \text{diag}(1, \dots, 1, 0, \dots, 0)$ with k_3K ones and k_2K zeros on the main diagonal, and $S' = S^{-1}$ is a $(k_2 + k_3)K \times (k_2 + k_3)K$ unitary matrix (i.e., $S'S = I_{(k_2+k_3)K}$). Decompose $S = [S_1 \ S_2]$ where S_1 and S_2 are $(k_2 + k_3)K \times k_3K$ and $(k_2 + k_3)K \times k_2K$ matrices, respectively. Then $M_{\bar{\Phi}_{K,B}S_{2K}} = SDS' = S_1S'_1$ and

$$\begin{aligned} \lambda_{\min}(A - BC^{-1}B') &= \lambda_{\min}(S_{1K}J'_{K,A}\bar{\Phi}'_{K,C}M_K\bar{\Phi}_{K,C}S_1S'_1\bar{\Phi}'_{K,C}M_K\bar{\Phi}_{K,C}J_{K,A}S_{1K}) \\ &\geq \lambda_{\min}(\bar{\Phi}_{K,C}S_1S'_1\bar{\Phi}'_{K,C}) \lambda_{\min}(S_{1K}J'_{K,A}\bar{\Phi}'_{K,C}M_KM_K\bar{\Phi}_{K,C}J_{K,A}S_{1K}). \end{aligned}$$

Noting that FF' and $F'F$ have the same positive eigenvalues with the same algebraic multiplicities, we have $\lambda_{\min}(\bar{\Phi}_{K,C}S_1S'_1\bar{\Phi}'_{K,C}) = \lambda_{\min}(S'_1\bar{\Phi}'_{K,C}\bar{\Phi}_{K,C}S_1) \geq \lambda_{\min}(S'_1S_1)\lambda_{\min}(\bar{\Phi}'_{K,C}\bar{\Phi}_{K,C}) \geq \lambda_{\min}(S'S)\bar{c}_\Phi^{-2} = \bar{c}_\Phi^{-2}$, where the last inequality follows from the fact that $\lambda_{\min}(\bar{\Phi}'_{K,C}\bar{\Phi}_{K,C}) \geq \lambda_{\min}([\Phi_K(A_\tau)]^{-1}[\Phi_K(A_\tau)]^{-1}) \geq \bar{c}_\Phi^{-2}$ and that $\lambda_{\min}(S'_1S_1) \geq \lambda_{\min}(S'S) = 1$ (see, e.g., Theorem 8.4.5. in Bernstein (2005)). In addition, $\lambda_{\min}(S_{1K}J'_{K,A}\bar{\Phi}'_{K,C}M_KM_K\bar{\Phi}_{K,C}J_{K,A}S_{1K}) \geq \underline{c}_M\lambda_{\min}(S_{1K}J'_{K,A}\bar{\Phi}'_{K,C}M_K\bar{\Phi}_{K,C}J_{K,A}S_{1K}) = \underline{c}_M\lambda_{\min}(S_{1K}) \geq \underline{c}_M[\bar{c}_M\underline{c}_\Phi^{-2}\bar{c}_J]^{-1}$. It follows that $\lambda_{\min}(A - BC^{-1}B') \geq \bar{c}_\Phi^{-2}\underline{c}_M[\bar{c}_M\underline{c}_\Phi^{-2}\bar{c}_J]^{-1} > 0$ uniformly in K . Analogously, using the fact that S_{2K} is idempotent with rank $(k_2 + k_3 - k_1)K$, we can show that $\lambda_{\min}(C)$ is bounded from below by a positive constant, c , say, under Assumptions A5(i)-(iv). As a result $\lambda_{\min}(\Omega_\tau\Omega'_\tau) \geq \underline{c}_\Omega > 0$ by taking $\underline{c}_\Omega = \min(\lambda_{\min}(S'_1S_1)\bar{c}_\Phi^{-2}\underline{c}_M[\bar{c}_M\underline{c}_\Phi^{-2}\bar{c}_J]^{-1}, c)$. This completes the proof of the lemma. ■

D Extension to the panel data models with individual fixed effects

Recall that $\chi_{it} = (U'_{it}, D'_{it}, W'_{it})'$ and $v_{it}(A, \Theta) = P_K(D_{it}, U_{it})'A + \phi_{W_{it}}^K(U_{it})'\Theta - \alpha_\tau(U_{it})'D_{it} - \beta_{1\tau}(U_{it})'\mathbf{X}_{1,it} - \beta'_{2\tau}\mathbf{X}_{2,it}$, $H = \text{diag}(I_{k_2K}, NK^{-1/2}I_N, I_{k_3K})$, and $H_1 = \text{diag}(I_{k_2K}, N^{1/2}I_N, I_{k_3K})$. Let $g_{it}(A, \Theta) =$

$\phi_{W_{it}}^K(U_{it}) \psi_\tau(\varepsilon_{it} - v_{it}(A, \Theta))$ and $\bar{H}_1 = H_1 H^{-1} H_1 = \text{diag}(I_{k_2 K}, K^{1/2} I_N, I_{k_3 K})$. Define

$$\begin{aligned}\Phi_{NT,K}(A) &= (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T f_{\varepsilon_{it}}(v_{it}(A, \Theta_\tau(A)) | \chi_{it}) H_1 \phi_{W_{it}}^K(U_{it}) \phi_{W_{it}}^K(U_{it})' H_1, \\ \hat{\Psi}_K &\equiv (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T H_1 \phi_{W_{it}}^K(U_{it}) \phi_{W_{it}}^K(U_{it})' H_1,\end{aligned}$$

Let $\Phi_K(A) = E[\Phi_{NT,K}(A)]$ and $\Psi_K = E[\hat{\Psi}_K]$. Further, define

$$\begin{aligned}\Omega_{\Theta_\tau} &\equiv [H_1 H^{-1} \Phi_K(A) H_1^{-1} H]^{-1} [I_{(k_2+k_3)K+N} + H_1 H^{-1} J_{K,A} \Omega_{A_\tau}], \\ \Omega_{B_\tau} &\equiv \bar{\Phi}_{K,B} [I_{(k_2+k_3)K+N} + H_1 H^{-1} J_{K,A} \Omega_{A_\tau}].\end{aligned}$$

To take into account the non-identical distributions of $(Y_{it}, U_{it}, D_{it}, \mathbf{X}_{1,it}, Z_{it})$ over either i or t , we re-define $\Pi_\tau(A, B)$ and $\Pi_{\tau|A}(B, C)$ as follows

$$\begin{aligned}\Pi_\tau(A, B) &\equiv (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T E[\psi_\tau(Y_{it} - P_K(D_{it}, U_{it})' A - P_K^\dagger(\mathbf{X}_{it}, U_{it})' B) \phi_{W_{it}}^K(U_{it})], \text{ and} \\ \Pi_{\tau|A}(B, C) &\equiv (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T E[\psi_\tau(Y_{it} - P_K(D_{it}, U_{it})' A - P_K^\dagger(\mathbf{X}_{it}, U_{it})' B - P_K(Z_{it}, U_{it})' C) \phi_{W_{it}}^K(U_{it})],\end{aligned}$$

where we suppress the potential dependence of $\Pi_\tau(A, B)$ and $\Pi_{\tau|A}(B, C)$ on (N, T) in the case of non-identical distributions. Let $\alpha_{\tau,l}(\cdot)$ and $\gamma_{\tau,l}(\cdot)$ be as defined in Section 2.2. Let $\beta_{1\tau,l}(\cdot)$ denote the l^{th} element of $\beta_{1\tau}(\cdot)$. Let $A_\tau \equiv (A'_{\tau,1}, \dots, A'_{\tau,k_1})'$, $B_{1\tau} \equiv (B'_{1\tau,1}, \dots, B'_{1\tau,k_2})'$, $B_\tau = (B'_{1\tau}, \beta'_{2\tau})'$, and $C_\tau \equiv (C'_{\tau,1}, \dots, C'_{\tau,k_3})'$. Let $\mathcal{A}_K, \mathcal{B}_K$, and \mathcal{C}_K denote the supports of A_τ, B_τ , and C_τ , respectively. We use $(N, T) \rightarrow \infty$ to denote that N and T pass to infinity jointly.

We make the following set of assumptions.

Assumption D1. (i) $(\varepsilon_{it}, U_{it}, D_{it}, \mathbf{X}_{1,it}, Z_{it})$ are independently distributed over i . For each $i = 1, \dots, N$, $\{(\varepsilon_{it}, U_{it}, D_{it}, \mathbf{X}_{1,it}, Z_{it}) : t = 1, 2, \dots\}$ is strong mixing with mixing coefficients $\{\alpha_i(\cdot)\}$ such that $\alpha(\cdot) \equiv \alpha_{NT}(\cdot) \equiv \max_{1 \leq i \leq N} \alpha_i(\cdot)$ satisfies $\alpha(s) = O(\rho^s)$ for some $\rho \in (0, 1)$. The supports of the exogenous variables $U_{it}, \mathbf{X}_{1,it}$, and Z_{it} are compact. $k_3 \geq k_1$.

(ii) $P(Y_{it} \leq \alpha_\tau(U_{it})' D_{it} + \beta_{1\tau}(U_{it})' \mathbf{X}_{1,it} + \beta_{2\tau,i} | U_{it}, \mathbf{X}_{1,it}, \beta_{2\tau,i}, Z_{it}) = \tau$ a.s.

(iii) The CDF of ε_{it} conditional on χ_{it} , $F_{\varepsilon_{it}}(\cdot | \chi_{it})$, exhibits a PDF $f_{\varepsilon_{it}}(\cdot | \chi_{it})$ that is bounded from above by \bar{c}_{f_ε} a.s.; $f_{\varepsilon_{it}}(\cdot | \chi_{it})$ is continuously differentiable in the neighborhood of 0 a.s. with first derivative bounded from above by \bar{c}_{f_ε} ; $E\{\sup_{v \in \mathbb{R}} \mathbf{1}\{|\varepsilon_{it} - v| \leq \vartheta(\chi_{it})\} | \chi_{it}\} \leq 2\bar{c}_{f_\varepsilon} \vartheta(\chi_{it})$ for any measurable function $\vartheta(\cdot)$.

(iv) The distribution of U_{it} is absolutely continuous with respect to the Lebesgue measure.

Assumption D2. (i) For $l = 1, \dots, k_1, k_2$, or k_3 , $\alpha_{\tau,l}(\cdot)$, $\beta_{1\tau,l}(\cdot)$, and $\gamma_{\tau,l}(\cdot)$ belong to the class of λ -smooth functions with $\lambda > 0$.

(ii) Assumption A2(ii) holds.

(iii) (A_τ, B_τ, C_τ) lies in the interior of $\mathcal{A}_K \times \mathcal{B}_K \times \mathcal{C}_K$, where $\mathcal{A}_K \subset \mathbb{R}^{k_1 K}$, $\mathcal{B}_K \subset \mathbb{R}^{k_2 K + N}$ and $\mathcal{C}_K \subset \mathbb{R}^{k_3 K}$ are compact and convex for all K and N , and \mathcal{C}_K contains 0 for all K .

Assumption D3. Assumption A3 holds.

Assumption D4. Assumption A4 holds.

Assumption D5. Assumption A5 holds.

Assumption D6. (i) Let $\zeta_K \equiv \sup_{u \in \mathcal{U}} \|p^K(u)\|$. As $(N, T) \rightarrow \infty$, $\zeta_K^2 K^3 [\ln(NT)]^2 / (NT) \rightarrow 0$ and $NTK^{-(1+2\lambda/d)} / \ln(NT) \rightarrow c_0 \in [0, \infty)$.

(ii) As $(N, T) \rightarrow \infty$, $NT\zeta_{K,u}^{-2} K^{-2\lambda/d} \rightarrow 0$ where $\zeta_{K,u} \equiv \|\Pi(u)\| > 0$

(iii) As $(N, T) \rightarrow \infty$, $\zeta_K^2 N \ln(NT) / T \rightarrow 0$.

Assumptions D1-D6(i)-(ii) parallel Assumptions A1-A6(i)-(ii). Note that we do not require identical distributions of $(\varepsilon_{it}, U_{it}, D_{it}, \mathbf{X}_{1,it}, Z_{it})$ over either i or t in D1(i) and D1(iii). D1(ii) is the quantile identification condition. Assumption D6(iii) is new. It is used in the proof of Lemmas D.2 and D.3 below and signifies the incidental parameter problem caused by the $N \times 1$ fixed effects parameter $\beta_{2\tau}$. For B-splines, $\zeta_K = O(K^{1/2})$ and D6(iii) becomes

$$KN \ln(NT) / T \rightarrow 0,$$

which is much weaker than the requirement $N^2 (\ln N)^3 / T \rightarrow 0$ as $(N, T) \rightarrow \infty$ used in Kato et al. (2012) because of the difference in the proof strategies. [Please note that $K = 1$ in Kato et al. (2012) and they implicitly require that T diverge to infinity at a rate that is a polynomial function of N .]

To prove Theorem 3.6, we first prove some technical lemmas that parallel Lemmas A.1-A.6 under the conditions stated in Theorem 3.6.

Lemma D.1 (i) $\sup_{A \in \mathcal{A}_K} \|\Phi_{NT,K}(A) - \Phi_K(A)\| = O_P(\zeta_K [K \ln(NT) / (NT)]^{1/2} + [(K+N) \ln(NT) / T]^{1/2}) = o_P(1)$.

(ii) $\underline{c}_\Phi / 2 \leq \inf_{A \in \mathcal{A}_K} \lambda_{\min}(\Phi_{NT,K}(A)) \leq \sup_{A \in \mathcal{A}_K} \lambda_{\max}(\Phi_{NT,K}(A)) \leq 2\bar{c}_\Phi$ w.p.a.1.

(iii) $\|\hat{\Psi}_K - \Psi_K\| = O_P(\zeta_K [K / (NT)]^{1/2} + [(K+N) / T]^{1/2}) = o_P(1)$.

(iv) $\underline{c}_\Psi / 2 \leq \lambda_{\min}(\hat{\Psi}_K) \leq \lambda_{\max}(\hat{\Psi}_K) \leq 2\bar{c}_\Psi$ w.p.a.1.

Proof. Let $f_{it,A} = f_{\varepsilon_{it}}(v_{it}(A, \Theta_\tau(A)) | \chi_{it})$. We partition the symmetric matrix $\Phi_{NT,K}(A)$ as follows

$$\Phi_{NT,K}(A) = \begin{pmatrix} \Phi_{NT,K}(A; 1, 1) & \Phi_{NT,K}(A; 1, 2) & \Phi_{NT,K}(A; 1, 3) \\ \Phi_{NT,K}(A; 2, 1) & \Phi_{NT,K}(A; 2, 2) & \Phi_{NT,K}(A; 2, 3) \\ \Phi_{NT,K}(A; 3, 1) & \Phi_{NT,K}(A; 3, 2) & \Phi_{NT,K}(A; 3, 3) \end{pmatrix}$$

where

$$\Phi_{NT,K}(A; 1, 1) = (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T f_{it,A} P_K(\mathbf{X}_{1,it}, U_{it}) P_K(\mathbf{X}_{1,it}, U_{it})',$$

$$\Phi_{NT,K}(A; 1, 2) = (NT)^{-1} N^{1/2} \sum_{i=1}^N \sum_{t=1}^T f_{it,A} P_K(\mathbf{X}_{1,it}, U_{it}) \mathbf{X}_{2,it}',$$

$$\Phi_{NT,K}(A; 1, 3) = (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T f_{it,A} P_K(\mathbf{X}_{1,it}, U_{it}) P_K(Z_{it}, U_{it})',$$

$$\Phi_{NT,K}(A; 2, 2) = (NT)^{-1} N \sum_{i=1}^N \sum_{t=1}^T f_{it,A} \mathbf{X}_{2,it} \mathbf{X}_{2,it}',$$

$$\begin{aligned}\Phi_{NT,K}(A; 2, 3) &= (NT)^{-1} N^{1/2} \sum_{i=1}^N \sum_{t=1}^T f_{it,A} \mathbf{X}_{2,it} P_K(Z_{it}, U_{it})', \\ \Phi_{NT,K}(A; 3, 3) &= (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T f_{it,A} P_K(Z_{it}, U_{it}) P_K(Z_{it}, U_{it})' .\end{aligned}$$

Partition $\Phi_K(A)$ analogously. Following the proof of Lemma A.1 (see also Newey (1997)), we can readily show that

$$\begin{aligned}\sup_{A \in \mathcal{A}_K} \|\Phi_{NT,K}(A; s, s) - \Phi_K(A; s, s)\| &= O_P(\zeta_K [K \ln(NT)/(NT)]^{1/2}) \text{ for } s = 1, 3, \text{ and} \\ \sup_{A \in \mathcal{A}_K} \|\Phi_{NT,K}(A; 1, 3) - \Phi_K(A; 1, 3)\| &= O_P(\zeta_K [K \ln(NT)/(NT)]^{1/2}).\end{aligned}$$

Let $\bar{f}_{i,A} = T^{-1} \sum_{t=1}^T f_{it,A}$. Recall that $\mathbf{X}_{2,it}$ denotes the i th column of I_N for each t . This implies that $\mathbf{X}_{2,i1} \mathbf{X}'_{2,i1}$ is an $N \times N$ matrix with 1 as its (i, i) th element and zeros everywhere else and $\bar{\Phi}_{NT,K}(A; 2, 2) = \sum_{i=1}^N \bar{f}_{i,A} \mathbf{X}_{2,i1} \mathbf{X}'_{2,i1}$ is a diagonal matrix with its (i, i) th diagonal element given by $\bar{f}_{i,A}$. With this observation, we can readily show that

$$\|\Phi_{NT,K}(A; 2, 2) - \Phi_K(A; 2, 2)\|^2 = \sum_{i=1}^N [\bar{f}_{i,A} - E(\bar{f}_{i,A})]^2 = \frac{1}{T^2} \sum_{i=1}^N \left[\sum_{t=1}^T [f_{it,A} - E(f_{it,A})] \right]^2 = O_P(NT^{-1}),$$

and $\sup_{A \in \mathcal{A}_K} \|\Phi_{NT,K}(A; 2, 2) - \Phi_K(A; 2, 2)\| = O_P([N \ln(NT)/T]^{1/2})$. Let $\zeta_{it,A} = f_{it,A} P_K(\mathbf{X}_{1,it}, U_{it})$ and $\bar{\zeta}_{i,A} = T^{-1} \sum_{t=1}^T \zeta_{it,A}$. Then $\Phi_{NT,K}(A; 1, 2) = N^{-1/2} \sum_{i=1}^N \bar{\zeta}_{i,A} \mathbf{X}'_{2,i1}$ denotes a $k_2 K \times N$ matrix whose i th column is given by $N^{-1/2} \bar{\zeta}_{i,A}$. Then we can readily show that

$$\begin{aligned}\|\Phi_{NT,K}(A; 1, 2) - \Phi_K(A; 1, 2)\|^2 &= \frac{1}{N} \sum_{i=1}^N [\bar{\zeta}_{i,A} - E(\bar{\zeta}_{i,A})]' [\bar{\zeta}_{i,A} - E(\bar{\zeta}_{i,A})] \\ &= \frac{1}{NT^2} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T [\zeta_{it,A} - E(\zeta_{it,A})]' [\zeta_{it,s} - E(\zeta_{it,s})] = O_P(K/T),\end{aligned}$$

and $\sup_{A \in \mathcal{A}_K} \|\Phi_{NT,K}(A; 1, 2) - \Phi_K(A; 1, 2)\| = O_P([K \ln(NT)/T]^{1/2})$. By the same token, $\sup_{A \in \mathcal{A}_K} \|\Phi_{NT,K}(A; 2, 3) - \Phi_K(A; 2, 3)\| = O_P([K \ln(NT)/T]^{1/2})$. In sum, we have

$$\sup_{A \in \mathcal{A}_K} \|\Phi_{NT,K}(A) - \Phi_K(A)\| = O_P(\zeta_K [K \ln(NT)/(NT)]^{1/2} + [(K+N) \ln(NT)/T]^{1/2}) = o_P(1).$$

Similarly, we can show that $\|\hat{\Psi}_K - \Psi_K\| = O_P(\zeta_K [K/(NT)]^{1/2} + [(K+N)/T]^{1/2}) = o_P(1)$. This proves (i) and (iii). The proofs of (ii) and (iv) are analogous to that of Lemma A.1(ii) and thus omitted. ■

Lemma D.2 $\sup_{A \in \mathcal{A}_K} \left\| (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T \bar{H}_1 g_{it}(A, \hat{\Theta}_\tau(A)) \right\| = O_P(\zeta_K [K/(NT) + 1/N]) = o_P((NT)^{-1/2})$.

Proof. The proof follows from that of Lemma A.2. Alternatively, we can apply Lemma A.2 in Ruppert and Carroll (1980) (see also Lemma A.5 in Koenker and Zhao (1995)) to obtain

$$\begin{aligned}\sup_{A \in \mathcal{A}_K} \left\| (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T \bar{H}_1 g_{it}(A, \hat{\Theta}_\tau(A)) \right\| &\leq (NT)^{-1} [(k_2 + k_3) K + N] \max_{i,t} \left\| \bar{H}_1 \phi_{W_{it}}^K(U_{it}) \right\| \\ &= (NT)^{-1} [(k_2 + k_3) K + N] O_P(\zeta_K + K^{1/2}) \\ &= O_P(\zeta_K [K/(NT) + 1/N]) = o_P((NT)^{-1/2}).\end{aligned}$$

■

Lemma D.3 $\sup_{A \in \mathcal{A}_K} \left\| (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T \bar{H}_1 g_{it}(A, \Theta_\tau(A)) \right\| = O_P \left([K \ln(NT)/(NT)]^{1/2} \right).$

Proof. As in the proof of Lemma D.1(i), we can readily show that

$$\begin{aligned} \left\| (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T \bar{H}_1 g_{it}(A, \Theta_\tau(A)) \right\|^2 &= \left\| (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T P_K(\mathbf{X}_{1,it}, U_{it}) \psi_\tau(\varepsilon_{it} - v_{it}(A, \Theta_\tau(A))) \right\|^2 \\ &\quad + \left\| (NT)^{-1} K^{1/2} \sum_{i=1}^N \sum_{t=1}^T \mathbf{X}_{2,it} \psi_\tau(\varepsilon_{it} - v_{it}(A, \Theta_\tau(A))) \right\|^2 \\ &\quad + \left\| (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T P_K(Z_{it}, U_{it}) \psi_\tau(\varepsilon_{it} - v_{it}(A, \Theta_\tau(A))) \right\|^2 \\ &= O_P(K/(NT)) + O_P(K/(NT)) + O_P(K/(NT)) = O_P(K/(NT)) \end{aligned}$$

and $\sup_{A \in \mathcal{A}_K} \left\| (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T g_{it}(A, \Theta_\tau(A)) \right\| = O_P \left([K \ln(NT)/(NT)]^{1/2} \right).$ ■

Lemma D.4 Let $\ell_{NT} \equiv L[K \ln(NT)/(NT)]^{1/2}$ and $\mathcal{S}_L(A) \equiv \{\Theta : \|H^{-1}(\Theta - \Theta_\tau(A))\| \leq \ell_{NT}\}$. Then

$$\sup_{\|\mathbf{c}\|=1} \sup_{A \in \mathcal{A}_K} \sup_{\Theta \in \mathcal{S}_L(A)} \left| (NT)^{-1} \mathbf{c}' \bar{H}_1 \sum_{i=1}^N \sum_{t=1}^T E[g_{it}(A, \Theta) - g_{it}(A, \Theta_\tau(A))] + \mathbf{c}' \bar{\Theta} \right| = o_P((NT)^{-1/2})$$

for any constant $L > 0$, where $\bar{\Theta} = \bar{H}_1 H_1^{-1} \Phi_K(A) H_1^{-1} [\Theta - \Theta_\tau(A)] = H_1 H^{-1} \Phi_K(A) H_1^{-1} [\Theta - \Theta_\tau(A)]$.

Proof. Following the proof of Lemma A.4 and using the weight matrices H , H_1 and \bar{H}_1 , we have

$$\begin{aligned} &\sup_{\|\mathbf{c}\|=1} \sup_{A \in \mathcal{A}_K} \sup_{\Theta \in \mathcal{S}_L(A)} \left| (NT)^{-1} \mathbf{c}' \bar{H}_1 \sum_{i=1}^N \sum_{t=1}^T E[g_{it}(A, \Theta) - g_{it}(A, \Theta_\tau(A))] + \mathbf{c}' \bar{\Theta} \right| \\ &\leq \sup_{A \in \mathcal{A}_K} \sup_{\Theta \in \mathcal{S}_L(A)} \left| (NT)^{-1} \mathbf{c}' \bar{H}_1 \sum_{i=1}^N \sum_{t=1}^T E \left\{ [F_{\varepsilon_{it}}(v_{it}(A, \Theta) | \chi_{it}) - F_{\varepsilon_{it}}(v_{it}(A, \Theta_\tau(A)) | \chi_{it})] \phi_{W_{it}}^K(U_{it}) \right\} - \mathbf{c}' \bar{\Theta} \right| \\ &\leq \frac{1}{2} \bar{c}_{f_\varepsilon} \left(\zeta_K + K^{1/2} \right) \sup_{A \in \mathcal{A}_K} \sup_{\Theta \in \mathcal{S}_L(A)} [H_1^{-1} H H^{-1} (\Theta - \Theta_\tau(A))]' \\ &\quad \times \left\{ (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T H_1 E \left[\phi_{W_{it}}^K(U_{it}) \phi_{W_{it}}^K(U_{it})' \right] H_1 \right\} H_1^{-1} H H^{-1} [\Theta - \Theta_\tau(A)] \\ &= \bar{c}_\Psi \bar{c}_{f_\varepsilon} O \left(\zeta_K + K^{1/2} \right) \sup_{A \in \mathcal{A}_K} \sup_{\Theta \in \mathcal{S}_L(A)} [H^{-1} (\Theta - \Theta_\tau(A))]' H H_1^{-1} H_1^{-1} H [H^{-1} (\Theta - \Theta_\tau(A))] \\ &= c \bar{c}_\Psi \bar{c}_{f_\varepsilon} O \left(\zeta_K + K^{1/2} \right) \ell_{NT}^2 N K^{-1} = O_P \left(\zeta_K \ln(NT) / T \right) = o_P \left((NT)^{-1/2} \right), \end{aligned}$$

where the second inequality follows from a second order Taylor expansion and the fact that $\max_{i,t} \sup_{\|\mathbf{c}\|=1} \left| \mathbf{c}' \bar{H}_1 \phi_{W_{it}}^K(U_{it}) \right|^2 = O(\zeta_K^2 + K)$ and the fourth inequality follows from the fact that $\lambda_{\max}(H H_1^{-1} H_1^{-1} H) = N/K$. ■

Lemma D.5 Let $\eta_{it}(A; \Theta_1, \Theta_2) \equiv g_{it}(A, \Theta_1) - g_{it}(A, \Theta_2) - E[g_{it}(A, \Theta_1)] + E[g_{it}(A, \Theta_2)]$. Then

(i) $\max_{i,t} E[\sup_{A \in \mathcal{A}_K} \sup_{\|H^{-1}(\Theta_1 - \Theta_2)\| \leq \Delta} \|\bar{H}_1 \eta_{it}(A; \Theta_1, \Theta_2)\|^2] \leq (NT)^{1/2} \Delta$ for sufficiently large N and T ;

(ii) for any constant $L > 0$ and $\mathbf{c} \in \mathbb{R}^{(k_2+k_3)K+N}$ with $\|\mathbf{c}\| = 1$, $E[\sup_{A \in \mathcal{A}_K} \sup_{\mathcal{S}_L(A)} \sum_{i=1}^N \sum_{t=1}^T \left| \mathbf{c}' \bar{H}_1 \eta_{it}(A; \Theta, \Theta_\tau(A)) \right|^2] = O(V(N, T, K))$ where $V(N, T, K) = NT \zeta_K K [K \ln(NT)/(NT)]^{1/2}$.

Proof. As in the proof of Lemma A.5, we have for sufficiently large N and T

$$\begin{aligned} \max_{i,t} E \left[\sup_{A \in \mathcal{A}_K} \sup_{\|H^{-1}(\Theta_1 - \Theta_2)\| \leq \Delta} \left\| \bar{H}_1 \eta_{it}(A; \Theta_1, \Theta_2) \right\|^2 \right] &\leq 2c(\zeta_K + 1) \max_{i,t} E \left\| \bar{H}_1 \phi_{W_{it}}^K(U_{it}) \right\|^2 \Delta \\ &\leq 2c^2(\zeta_K + 1) K \Delta \leq (NT)^{1/2} \Delta, \end{aligned}$$

where c is a generic large constant and we use the fact that

$$E \left\| \bar{H}_1 \phi_{W_{it}}^K(U_{it}) \right\|^2 = \text{tr} \left\{ E \left[\bar{H}_1 \phi_{W_{it}}^K(U_{it}) \phi_{W_{it}}^K(U_{it})' \bar{H}_1 \right] \right\} = O(K).$$

So (i) follows. Analogously, we can show (ii). ■

Lemma D.6 For any $\mathbf{c} \in \mathbb{R}^{(k_2+k_3)K+N}$ with $\|\mathbf{c}\| \leq 1$ and $r > 0$, we have

$$\sup_{\|(A, \Theta_1)\| \leq (NT)^r, \|(A, \Theta_2)\| \leq (NT)^r} \frac{\sum_{i=1}^N \sum_{t=1}^T \|\mathbf{c}' \bar{H}_1 \eta_{it}(A; \Theta_1, \Theta_2)\|}{t_{1NT}(A; \Theta_1, \Theta_2) + t_{2NT}(A; \Theta_1, \Theta_2) + (NT)^{-2}} = O_P \left([K \ln(NT)]^{1/2} \right),$$

where $t_{1NT}(A; \Theta_1, \Theta_2) = \left\{ \sum_{i=1}^N \sum_{t=1}^T E \left\| \mathbf{c}' \bar{H}_1 \eta_{it}(A; \Theta_1, \Theta_2) \right\|^2 \right\}^{1/2}$ and $t_{2NT}(A; \Theta_1, \Theta_2) = \left\{ \sum_{i=1}^N \sum_{t=1}^T \left\| \mathbf{c}' \bar{H}_1 \eta_{it}(A; \Theta_1, \Theta_2) \right\|^2 \right\}^{1/2}$.

Proof. If $(Y_{it}, U_{it}, D_{it}, \mathbf{X}_{1,it}, Z_{it})$ is independent over both i and t , the above result follows from Lemma 3.2 in He and Shao (2000) directly. Since we only assume that $(Y_{it}, U_{it}, D_{it}, \mathbf{X}_{1,it}, Z_{it})$ is independent over i and strong mixing over t , we need to modify the proof of Lemma 3.2 in He and Shao (2000) by using Bernstein inequality for strong mixing processes over the time dimension. The details are omitted. ■

Proof of Theorem 3.6. (i) Let $\Theta_{\tau, \mathbf{c}}(A) = \Theta_\tau(A) + L[K \ln(NT)/(NT)]^{1/2} H \mathbf{c}$, $J_{1NT}(A) = \left\{ \sum_{i=1}^N \sum_{t=1}^T E |\mathbf{c}' \eta_{it}(A; \Theta_{\tau, \mathbf{c}}(A), \Theta_\tau(A))|^2 \right\}^{1/2}$ and $J_{2NT}(A) = \left\{ \sum_{i=1}^N \sum_{t=1}^T |\mathbf{c}' \eta_{it}(A; \Theta_{\tau, \mathbf{c}}(A), \Theta_\tau(A))|^2 \right\}^{1/2}$. By Lemmas D.4-D.6, we have that uniformly in $A \in \mathcal{A}_K$

$$\begin{aligned} &\sum_{i=1}^N \sum_{t=1}^T \mathbf{c}' \bar{H}_1 g_{it}(A, \Theta_{\tau, \mathbf{c}}(A)) \\ &= \sum_{i=1}^N \sum_{t=1}^T \mathbf{c}' \bar{H}_1 g_{it}(A, \Theta_\tau(A)) + \sum_{i=1}^N \sum_{t=1}^T \mathbf{c}' \bar{H}_1 E [g_{it}(A, \Theta_{\tau, \mathbf{c}}(A)) - g_{it}(A, \Theta_\tau(A))] \\ &\quad + \left[J_{1NT}(A) + J_{2NT}(A) + (NT)^{-2} \right] O_P([K \ln(NT)]^{1/2}) \\ &= \sum_{i=1}^N \sum_{t=1}^T \mathbf{c}' \bar{H}_1 g_{it}(A, \Theta_\tau(A)) - L[NTK \ln(NT)]^{1/2} \mathbf{c}' \bar{H}_1 H_1^{-1} \Phi_K(A) H_1^{-1} H \mathbf{c} + o_P([NTK \ln(NT)]^{1/2}) \\ &= \sum_{i=1}^N \sum_{t=1}^T \mathbf{c}' \bar{H}_1 g_{it}(A, \Theta_\tau(A)) - L[NTK \ln(NT)]^{1/2} \mathbf{c}' H_1 H^{-1} \Phi_K(A) H_1^{-1} H \mathbf{c} + o_P([NTK \ln(NT)]^{1/2}). \end{aligned}$$

Consequently,

$$\begin{aligned}
& P \left(\inf_{\|\mathbf{c}\|=1} - \sum_{i=1}^N \sum_{t=1}^T \mathbf{c}' \bar{H}_1 g_{it}(A, \Theta_{\tau, \mathbf{c}}(A)) > 0 \text{ for all } A \in \mathcal{A}_K \right) \\
& \geq P \left([NTK \ln(NT)]^{-1/2} \inf_{\|\mathbf{c}\|=1} - \sum_{i=1}^N \sum_{t=1}^T \mathbf{c}' \bar{H}_1 g_{it}(A, \Theta_{\tau}(A)) > -\frac{L}{2} \mathbf{c}' H_1 H^{-1} \Phi_K(A) H_1^{-1} H \mathbf{c} \text{ for all } A \in \mathcal{A}_K \right) \\
& \geq P \left([NTK \ln(NT)]^{-1/2} \inf_{\|\mathbf{c}\|=1} - \sum_{i=1}^N \sum_{t=1}^T \mathbf{c}' \bar{H}_1 g_{it}(A, \Theta_{\tau}(A)) > -\frac{L}{2} \underline{c}_{\Phi} \text{ for all } A \in \mathcal{A}_K \right) \\
& = P \left([NTK \ln(NT)]^{-1/2} \sup_{A \in \mathcal{A}_K} \sup_{\|\mathbf{c}\|=1} - \sum_{i=1}^N \sum_{t=1}^T \mathbf{c}' \bar{H}_1 g_{it}(A, \Theta_{\tau}(A)) < \frac{L}{2} \underline{c}_{\Phi} \right) \rightarrow 1 \text{ as } L, N, T \rightarrow \infty,
\end{aligned}$$

where we use the fact that $\lambda_{\min}(H_1 H^{-1} \Phi_K(A) H_1^{-1} H) \geq \underline{c}_{\Phi}$ and the last line follows by Lemma D.3. It follows that $\|H^{-1}[\hat{\Theta}_{\tau}(A) - \Theta_{\tau}(A)]\| = O_P([K \ln(NT)]/(NT)^{1/2})$.

For part (ii), we apply Lemma D.6 with $(\Theta_1, \Theta_2) = (\hat{\Theta}_{\tau}(A), \Theta_{\tau}(A))$ to obtain $\sum_{i=1}^N \sum_{t=1}^T \mathbf{c}' \bar{H}_1 \eta_{it}(A; \hat{\Theta}_{\tau}(A), \Theta_{\tau}(A)) = [\hat{J}_{NT}(A) + (NT)^{-2}] O_P([K \ln(NT)]^{1/2})$, where $\hat{J}_{NT}(A) = O_P(V(N, T, K)^{1/2})$ by (i) and Lemma D.5(ii). Then by Lemmas D.2 and D.4, we have

$$\begin{aligned}
& \mathbf{c}' H_1 H^{-1} \Phi_K(A) H_1^{-1} H H^{-1} [\hat{\Theta}_{\tau}(A) - \Theta_{\tau}(A)] \\
& = (NT)^{-1} \mathbf{c}' \bar{H}_1 \sum_{i=1}^n g_{it}(A, \Theta_{\tau}(A)) + O_P\left((NT)^{-1} [V(N, T, K) K \ln(NT) NT]^{1/2}\right) + o_P((NT)^{-1/2})
\end{aligned}$$

uniformly in $A \in \mathcal{A}_K$ and \mathbf{c} with $\|\mathbf{c}\| = 1$. It follows that

$$H^{-1} [\hat{\Theta}_{\tau}(A) - \Theta_{\tau}(A)] = (NT)^{-1} [H_1 H^{-1} \Phi_K(A) H_1^{-1} H]^{-1} \bar{H}_1 \sum_{i=1}^N \sum_{t=1}^T g_{it}(A, \Theta_{\tau}(A)) + o_P((NT)^{-1/2}) + r_{NT},$$

where $\|r_{NT}\| = O_P\left((NT)^{-1} [V(N, T, K) K \ln(NT) NT]^{1/2}\right) = O_P(\zeta_K^{1/2} K^{5/4} (NT)^{-3/4} \ln(NT))$. ■

Proof of Theorem 3.7. Following the proof of Theorem 3.3 closely, we can show that

$$\begin{aligned}
\hat{A}_{\tau} - A_{\tau} &= \Omega_{A_{\tau}} (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T \bar{H}_1 g_{it}(A_{\tau}, \Theta_{\tau}) + o_P\left((NT)^{-1/2}\right), \\
H^{-1} (\hat{\Theta}_{\tau} - \Theta_{\tau}) &= \Omega_{\Theta_{\tau}} (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T \bar{H}_1 g_{it}(A_{\tau}, \Theta_{\tau}) + o_P\left((NT)^{-1/2}\right), \\
H^{-1} (\hat{B}_{1\tau} - B_{1\tau}) &= \Omega_{B_{1\tau}} (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T \bar{H}_1 g_{it}(A_{\tau}, \Theta_{\tau}) + o_P\left((NT)^{-1/2}\right).
\end{aligned}$$

Using arguments as used in the proof of Theorem 3.4, we can show that $\sup_{u \in \mathcal{U}} \|\hat{\alpha}_{\tau}(u) - \alpha_{\tau}(u)\| = O_P[\zeta_K((NT/K)^{-1/2} + K^{-\lambda/d})]$ and $\sup_{u \in \mathcal{U}} \|\hat{\beta}_{1\tau}(u) - \beta_{1\tau}(u)\| = O_P[\zeta_K((NT/K)^{-1/2} + K^{-\lambda/d})]$.

As in the proof of Theorem 3.4, noting that $\hat{\delta}_{\tau}(u) = \begin{pmatrix} \hat{\alpha}_{\tau}(u) \\ \hat{\beta}_{1\tau}(u) \end{pmatrix} = \begin{pmatrix} \Pi_{\alpha}(u) \hat{A}_{\tau} \\ \Pi_{\beta_1}(u) \hat{B}_{1\tau} \end{pmatrix} = \Pi(u) \begin{pmatrix} \hat{A}_{\tau} \\ \hat{B}_{1\tau} \end{pmatrix}$,

we have

$$\hat{\delta}_{\tau}(u) - \delta_{\tau}(u) = \Pi(u) \begin{pmatrix} \hat{A}_{\tau} - A_{\tau} \\ \hat{B}_{1\tau} - B_{1\tau} \end{pmatrix} - \begin{pmatrix} \alpha_{\tau}(u) - \Pi_{\alpha}(u) A_{\tau} \\ \beta_{1\tau}(u) - \Pi_{\beta_1}(u) B_{1\tau} \end{pmatrix} = \Pi(u) \mathbf{V}_{\tau} - \mathbf{B}_{\tau}(u),$$

where $\mathbf{V}_\tau = \begin{pmatrix} \hat{A}_\tau - A_\tau \\ \hat{B}_\tau - B_\tau \end{pmatrix}$ and $\mathbf{B}_\tau(u) = \begin{pmatrix} \alpha_\tau(u) - \Pi_\alpha(u) A_\tau \\ \beta_{1\tau}(u) - \Pi_{\beta_1}(u) B_{1\tau} \end{pmatrix}$. Let $\Sigma_\tau(u) \equiv \tau(1-\tau)\Pi(u)\Omega_\tau\Psi_K \times \Omega'_\tau\Pi(u)'$ and $\Lambda_\tau(u) = \Sigma_\tau(u)^{-1/2}\Pi(u)\Omega_\tau$. Then

$$\begin{aligned} \Sigma_\tau(u)^{-1/2}\sqrt{NT}\left[\hat{\delta}_\tau(u) - \delta_\tau(u)\right] &= \Lambda_\tau(u)(NT)^{-1/2}\sum_{i=1}^N\sum_{t=1}^T\bar{H}_1\phi_{W_{it}}^K(U_{it})\psi_\tau(\varepsilon_{it}) \\ &\quad + \Lambda_\tau(u)(NT)^{-1/2}\sum_{i=1}^n\phi_{W_{it}}^K(U_{it})[\psi_\tau(\varepsilon_{it} - v_{it}(A_\tau, \Theta_\tau)) - \psi_\tau(\varepsilon_{it})] \\ &\quad + \sqrt{NT}\Sigma_\tau(u)^{-1/2}\Pi(u)r_{NT} - \sqrt{NT}\Sigma_\tau(u)^{-1/2}\mathbf{B}_\tau(u) \\ &\equiv D_{1NT} + D_{2NT} + D_{3NT} + D_{4NT}, \text{ say,} \end{aligned}$$

where $\|r_{NT}\| = o_P((NT)^{-1/2})$. Using analogous arguments to those used in the proof of Theorem 3.5, we can show that (i) $D_{1NT} \xrightarrow{d} N(0, I_{k_1+k_2})$ and (ii) $D_{sNT} = o_P(1)$ for $s = 2, 3, 4$. It follows that $\Sigma_\tau(u)^{-1/2}\sqrt{NT}\left[\hat{\delta}_\tau(u) - \delta_\tau(u)\right] \xrightarrow{d} N(0, I_{k_1+k_2})$. ■

E Justification for the asymptotic validity of \hat{T}_n^{**} defined in Remark 12

In this appendix, we prove the asymptotic validity of \hat{T}_n^{**} defined in Remark 12.

Let $\sum_{j<k} = \sum_{j=1}^{n-1}\sum_{k=j+1}^n$ and $\sum_{j,k} = \sum_{j=1}^n\sum_{k=1}^n$. Define $\sum_{i<j<k}$ and $\sum_{i<j<k<l}$ similarly. Let $\hat{\Omega}_\tau = \hat{\Omega}'_\tau\hat{\Upsilon}\hat{\Omega}_\tau$. We assume that the probability limits of $\lambda_{\min}(\hat{\Psi}_K^{**})$ and $\lambda_{\min}(\hat{\Omega}_\tau)$ are bounded away from zero and the those of $\lambda_{\max}(\hat{\Psi}_K)$ and $\lambda_{\max}(\hat{\Omega}_\tau)$ are bounded away from infinity, both of which can be verified under either the null hypothesis or the local alternative. To justify the asymptotic validity of the above bootstrap method, it suffices to show that conditional on the original sample \mathcal{D}_n , \hat{T}_n^{**} converges to $N(0, 1)$ in distribution no matter whether \mathbb{H}_0 holds in the original data or not. Let $\varphi_n^{**}(w_j, w_k) = \hat{\varphi}_n(\hat{\zeta}_j, \hat{\zeta}_k)w_jw_k$. Let

$$T_n^{**} = \frac{2}{n}\sum_{j<k}\varphi_n^{**}(w_j, w_k).$$

Note that $\hat{T}_n^{**} = \sigma_n^{**,-1}T_n^{**}$ and T_n^{**} plays the role of the score function (or influence function) in Kline and Santos (2012). Let E^* denote the conditional expectation under the probability law introduced by the bootstrap conditional on the data. Apparently, $\varphi_n^{**}(w_j, w_k) = \varphi_n^{**}(w_k, w_j)$ and $E^*[\varphi_n^{**}(w_j, w_k)|w_k] = 0$ for any $j \neq k$. So T_n^{**} is a degenerate second order U-statistic with asymptotic variance (conditional on \mathcal{D}_n) given by

$$\begin{aligned} \sigma_n^{**2} &= \frac{2}{n^2}\sum_{j,k}E_j^*E_k^*\left[\varphi_n^{**}(w_j, w_k)^2\right] = \frac{2}{n^2}\sum_{j,k}\left[\psi_\tau(\hat{\varepsilon}_j)\phi_{W_j}^K(U_j)'\hat{\Omega}'_\tau\hat{\Upsilon}\hat{\Omega}_\tau\phi_{W_k}^K(U_k)\psi_\tau(\hat{\varepsilon}_k)\right]^2 \\ &= 2\tau^2(1-\tau)^2\text{tr}\left\{\hat{\Omega}_\tau\hat{\Psi}_K^{**}\hat{\Omega}_\tau\hat{\Psi}_K^{**}\right\} \geq 2\tau^2(1-\tau)^2\lambda_{\min}^2(\hat{\Omega}_\tau)\lambda_{\min}^2(\hat{\Psi}_K^{**})\text{tr}(I_{(k_2+k_3)K}) \\ &= 2K\tau^2(1-\tau)^2\lambda_{\min}^2(\hat{\Omega}_\tau)\lambda_{\min}^2(\hat{\Psi}_K^{**}), \end{aligned}$$

where E_j^* denotes that expectation with respect to w_j conditional on \mathcal{D}_n . We prove $\sigma_n^{**,-1}T_n^{**} \xrightarrow{d} N(0, 1)$ conditional on \mathcal{D}_n by verifying all the conditions of Proposition 3.2 in de Jong (1987) are satisfied. Let

$\varphi_{ij}^* = n^{-1} \varphi_n^{**}(w_j, w_k)$. Define

$$\begin{aligned} G_I &= \sum_{j < k} E^*(\varphi_{ij}^{*4}), \\ G_{II} &= \sum_{i < j < k} [E^*(\varphi_{ij}^{*2} \varphi_{ik}^{*2}) + E^*(\varphi_{ji}^{*2} \varphi_{jk}^{*2}) + E^*(\varphi_{ki}^{*2} \varphi_{kj}^{*2})], \\ G_{IV} &= \sum_{i < j < k < l} [E^*(\varphi_{ij}^* \varphi_{ik}^* \varphi_{lj}^* \varphi_{lk}^*) + E^*(\varphi_{ij}^* \varphi_{il}^* \varphi_{kj}^* \varphi_{kl}^*) + E^*(\varphi_{ik}^* \varphi_{il}^* \varphi_{jk}^* \varphi_{jl}^*)]. \end{aligned}$$

By Proposition 3.2 in de Jong (1987), it suffices to show that G_I , G_{II} , and G_{IV} are of smaller probability order than σ_n^{**4} .

(i) Observe that

$$\begin{aligned} \sigma_n^{**4} G_I &= \frac{\mu_4^2}{n^4 \sigma_n^{**4}} \sum_{j < k} \hat{\varphi}_n(\hat{\zeta}_j, \hat{\zeta}_k)^4 \\ &= \frac{\mu_4^2}{n^4 \sigma_n^{**4}} \sum_{j < k} [\psi_\tau(\hat{\varepsilon}_j) \phi_{W_j}^K(U_j)' \hat{\Omega}_\tau \phi_{W_k}^K(U_k) \psi_\tau(\hat{\varepsilon}_k)]^4 \\ &\leq \frac{\mu_4^2}{n^4 \sigma_n^{**4}} \sum_{j < k} \phi_{W_j}^K(U_j)' \hat{\Omega}_\tau \phi_{W_k}^K(U_k) \phi_{W_k}^K(U_k)' \hat{\Omega}_\tau \phi_{W_j}^K(U_j) \phi_{W_j}^K(U_j)' \hat{\Omega}_\tau \phi_{W_k}^K(U_k) \phi_{W_k}^K(U_k)' \hat{\Omega}_\tau \phi_{W_j}^K(U_j) \\ &\leq \frac{\mu_4^2 c_\phi^2 \lambda_{\max}^2(\hat{\Omega}_\tau)}{n^4 \sigma_n^{**4}} \sum_{j < k} \text{tr} \left[\phi_{W_j}^K(U_j)' \hat{\Omega}_\tau \phi_{W_k}^K(U_k) \phi_{W_k}^K(U_k)' \phi_{W_k}^K(U_k) \phi_{W_k}^K(U_k)' \hat{\Omega}_\tau \phi_{W_j}^K(U_j) \right] \\ &\leq \frac{\mu_4^2 c_\phi^2 \lambda_{\max}^2(\hat{\Omega}_\tau)}{n^3 \sigma_n^{**4}} \text{tr} \left[\hat{\Omega}_\tau \sum_{k=1}^n \phi_{W_k}^K(U_k) \phi_{W_k}^K(U_k)' \phi_{W_k}^K(U_k) \phi_{W_k}^K(U_k)' \hat{\Omega}_\tau \hat{\Psi}_K \right] \\ &\leq \frac{\mu_4^2 c_\phi^2 \lambda_{\max}^2(\hat{\Omega}_\tau) \lambda_{\max}(\hat{\Psi}_K)}{n^3 \sigma_n^{**4}} \text{tr} \left[\hat{\Omega}_\tau \sum_{k=1}^n \phi_{W_k}^K(U_k) \phi_{W_k}^K(U_k)' \phi_{W_k}^K(U_k) \phi_{W_k}^K(U_k)' \hat{\Omega}_\tau \right] \\ &\leq \frac{\mu_4^2 c_\phi^4 \lambda_{\max}^2(\hat{\Omega}_\tau) \lambda_{\max}(\hat{\Psi}_K)}{n^2 \sigma_n^{**4}} \text{tr} \left(\hat{\Omega}_\tau \hat{\Psi}_K \hat{\Omega}_\tau \right) \leq \frac{\mu_4^2 c_\phi^4 \lambda_{\max}^2(\hat{\Omega}_\tau) \lambda_{\max}(\hat{\Psi}_K) \text{tr}(I_{(k_2+k_3)K})}{n^2 \sigma_n^{**4}} \\ &= O_P \left(\frac{\zeta_K^4 K}{n^2 \sigma_n^{**4}} \right) = O_P \left(\frac{\zeta_K^4}{n^2 K} \right) = o_P(1), \end{aligned}$$

where $c_\phi = \sup_{(w,u)} \left\| \phi_w^K(w) \right\| = O(\zeta_K)$ and recall that $\hat{\Psi}_K = \frac{1}{n} \sum_{i=1}^n \phi_{W_i}^K(U_i) \phi_{W_i}^K(U_i)'$.

(ii) Write $G_{II} = \sum_{i < j < k} [E^*(\varphi_{ij}^{*2} \varphi_{ik}^{*2}) + E^*(\varphi_{ji}^{*2} \varphi_{jk}^{*2}) + E^*(\varphi_{ki}^{*2} \varphi_{kj}^{*2})] = G_{II}^{(1)} + G_{II}^{(2)} + G_{II}^{(3)}$, say. By

moment calculations,

$$\begin{aligned}
\sigma_n^{**4} G_{II}^{(1)} &= \frac{1}{\sigma_n^{**4}} \sum_{i < j < k} E^* (\varphi_{ij}^{*2} \varphi_{ik}^{*2}) = \frac{\mu_4}{n^4 \sigma_n^{**4}} \sum_{1 \leq i < j < k \leq n} \hat{\varphi}_n(\hat{\zeta}_i, \hat{\zeta}_j)^2 \hat{\varphi}_n(\hat{\zeta}_i, \hat{\zeta}_k)^2 \\
&= \frac{\mu_4}{n^4 \sigma_n^{**4}} \sum_{i < j < k} \left[\psi_\tau(\hat{\varepsilon}_i) \phi_{W_i}^K(U_i)' \hat{\Omega}_\tau \phi_{W_j}^K(U_j) \psi_\tau(\hat{\varepsilon}_j) \right]^2 \left[\psi_\tau(\hat{\varepsilon}_i) \phi_{W_i}^K(U_i)' \hat{\Omega}_\tau \phi_{W_k}^K(U_k) \psi_\tau(\hat{\varepsilon}_k) \right]^2 \\
&\leq \frac{\mu_4}{n^4 \sigma_n^{**4}} \sum_{i < j < k} \phi_{W_i}^K(U_i)' \hat{\Omega}_\tau \phi_{W_j}^K(U_j) \phi_{W_j}^K(U_j)' \hat{\Omega}_\tau \phi_{W_i}^K(U_i) \phi_{W_i}^K(U_i)' \hat{\Omega}_\tau \phi_{W_k}^K(U_k) \phi_{W_k}^K(U_k)' \hat{\Omega}_\tau \phi_{W_i}^K(U_i) \\
&\leq \frac{\mu_4}{n^2 \sigma_n^{**4}} \sum_{i=1}^n \phi_{W_i}^K(U_i)' \hat{\Omega}_\tau \hat{\Psi}_K \hat{\Omega}_\tau \phi_{W_i}^K(U_i) \phi_{W_i}^K(U_i)' \hat{\Omega}_\tau \hat{\Psi}_K \hat{\Omega}_\tau \phi_{W_i}^K(U_i) \\
&\leq \frac{c_\phi^2 \mu_4}{n \sigma_n^{**4}} \text{tr} \left(\hat{\Omega}_\tau \hat{\Psi}_K \hat{\Omega}_\tau \hat{\Psi}_K \hat{\Omega}_\tau \hat{\Psi}_K \hat{\Omega}_\tau \right) \leq \frac{c_\phi^2 \mu_4 \lambda_{\max}^4(\hat{\Omega}_\tau) \lambda_{\max}^3(\hat{\Psi}_K) \text{tr}(I_{(k_2+k_3)K})}{n \sigma_n^{**4}} \\
&= O_P \left(\frac{\zeta_K^2}{nK} \right) = o_P(1).
\end{aligned}$$

By the same token, $\sigma_n^{**4} G_{II}^{(s)} = o_P(1)$ for $s = 2, 3$.

(iii) Write $G_{IV} = \sum_{i < j < k < l} \left[E^*(\varphi_{ij}^* \varphi_{ik}^* \varphi_{lj}^* \varphi_{lk}^*) + E^*(\varphi_{ij}^* \varphi_{il}^* \varphi_{kj}^* \varphi_{kl}^*) + E^*(\varphi_{ik}^* \varphi_{il}^* \varphi_{jk}^* \varphi_{jl}^*) \right] = G_{IV}^{(1)} + G_{IV}^{(2)} + G_{IV}^{(3)}$, say. Then

$$\begin{aligned}
\sigma_n^{**4} G_{IV}^{(1)} &= \frac{1}{\sigma_n^{**4}} \sum_{i < j < k < l} E^* (\varphi_{ij}^* \varphi_{ik}^* \varphi_{lj}^* \varphi_{lk}^*) \\
&= \frac{1}{n^4 \sigma_n^{**4}} \sum_{i < j < k < l} \hat{\varphi}_n(\hat{\zeta}_i, \hat{\zeta}_j) \hat{\varphi}_n(\hat{\zeta}_i, \hat{\zeta}_k) \hat{\varphi}_n(\hat{\zeta}_l, \hat{\zeta}_j) \hat{\varphi}_n(\hat{\zeta}_l, \hat{\zeta}_k) \\
&\leq \frac{1}{n^4 \sigma_n^{**4}} \sum_{i < j < k < l} \phi_{W_j}^K(U_j)' \hat{\Omega}_\tau \phi_{W_i}^K(U_i) \phi_{W_i}^K(U_i)' \hat{\Omega}_\tau \phi_{W_k}^K(U_k) \phi_{W_k}^K(U_k)' \hat{\Omega}_\tau \phi_{W_l}^K(U_l) \phi_{W_l}^K(U_l)' \hat{\Omega}_\tau \phi_{W_j}^K(U_j) \\
&\leq \frac{1}{\sigma_n^{**4}} \text{tr} \left(\hat{\Omega}_\tau \hat{\Psi}_K \hat{\Omega}_\tau \hat{\Psi}_K \hat{\Omega}_\tau \hat{\Psi}_K \hat{\Omega}_\tau \hat{\Psi}_K \right) \leq \frac{\lambda_{\max}^4(\hat{\Omega}_\tau) \lambda_{\max}^4(\hat{\Psi}_K) \text{tr}(I_{(k_2+k_3)K})}{n \sigma_n^{**4}} \\
&= O_P \left(\frac{1}{K} \right) = o_P(1).
\end{aligned}$$

By the same token, $\sigma_n^{**4} G_{IV}^{(s)} = o_P(1)$ for $s = 2, 3$.

That is, the conditions in Proposition 3.2 in de Jong (1987) are all satisfied and we can conclude $\hat{T}_n^{**} \xrightarrow{d} N(0, 1)$ conditional on the data.

F Some additional simulation results

In this appendix, we report some additional simulation results for the bootstrap test based on \hat{T}_n^{**} defined in Remark 12 in the text. Tables A.1 and A.2 report the finite sample rejection frequency for our \hat{T}_n^{**} -based bootstrap test when the weights $\{w_i\}$ are generated as independent Rademacher and standard normal random variables, respectively. From both tables, we can see the tests are severely undersized for both DGPs under consideration. Despite this, the power performance is comparable with our \hat{T}_n^* -based bootstrap test.

Table A.1: Finite sample rejection frequency of \hat{T}_n^{**} -based bootstrap test: Rademacher weights

DGP	Δ_0	ρ	n	$\mathbb{H}_{0,\alpha}$			$\mathbb{H}_{0,\beta}$			$\mathbb{H}_{0,\alpha\beta}$		
				1%	5%	10%	1%	5%	10%	1%	5%	10%
1'	0	0.2	200	0.004	0.022	0.034	0.004	0.022	0.040	0.002	0.018	0.032
			400	0.004	0.038	0.064	0.004	0.020	0.044	0.008	0.026	0.036
			800	0.002	0.028	0.042	0.008	0.022	0.048	0.010	0.020	0.040
		0.5	200	0.004	0.020	0.056	0.004	0.034	0.062	0.002	0.022	0.054
			400	0.002	0.012	0.038	0.002	0.014	0.028	0.004	0.012	0.022
			800	0.006	0.026	0.050	0.008	0.040	0.062	0.004	0.032	0.050
		0.8	200	0.000	0.028	0.048	0.008	0.032	0.046	0.006	0.030	0.044
			400	0.006	0.026	0.044	0.000	0.016	0.036	0.000	0.010	0.032
			800	0.002	0.022	0.046	0.004	0.026	0.044	0.002	0.016	0.034
	1	0.2	200	0.998	1.000	1.000	0.618	0.844	0.894	0.970	0.988	0.990
			400	1.000	1.000	1.000	0.932	0.984	0.986	1.000	1.000	1.000
			800	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.5	200	1.000	1.000	1.000	0.698	0.866	0.932	0.984	0.994	1.000
			400	1.000	1.000	1.000	0.940	0.990	0.994	0.998	1.000	1.000
			800	1.000	1.000	1.000	0.998	1.000	1.000	1.000	1.000	1.000
		0.8	200	0.994	0.998	0.998	0.646	0.826	0.892	0.960	0.984	0.990
			400	1.000	1.000	1.000	0.952	0.988	0.992	0.998	1.000	1.000
			800	1.000	1.000	1.000	0.996	0.998	1.000	1.000	1.000	1.000
3'	0	0.2	200	0.014	0.048	0.086	0.008	0.028	0.066	0.014	0.030	0.070
			400	0.006	0.014	0.048	0.002	0.016	0.034	0.002	0.018	0.032
			800	0.008	0.028	0.046	0.004	0.014	0.030	0.004	0.014	0.038
		0.5	200	0.012	0.046	0.072	0.008	0.036	0.060	0.004	0.024	0.050
			400	0.010	0.046	0.076	0.008	0.044	0.076	0.012	0.038	0.068
			800	0.008	0.018	0.060	0.000	0.016	0.038	0.000	0.018	0.042
		0.8	200	0.016	0.044	0.074	0.012	0.048	0.080	0.010	0.042	0.074
			400	0.004	0.028	0.068	0.012	0.054	0.084	0.008	0.048	0.076
			800	0.012	0.056	0.082	0.020	0.046	0.068	0.018	0.050	0.070
	1	0.2	200	0.890	0.950	0.968	0.212	0.388	0.506	0.632	0.770	0.848
			400	0.942	0.964	0.980	0.432	0.614	0.724	0.812	0.902	0.934
			800	0.992	0.998	0.998	0.872	0.952	0.980	0.984	0.992	0.996
		0.5	200	0.870	0.934	0.954	0.224	0.368	0.492	0.594	0.740	0.814
			400	0.942	0.968	0.982	0.484	0.714	0.792	0.858	0.926	0.962
			800	0.996	0.996	0.998	0.890	0.952	0.968	0.982	0.990	0.994
		0.8	200	0.848	0.932	0.954	0.256	0.416	0.548	0.662	0.802	0.870
			400	0.936	0.960	0.968	0.520	0.706	0.798	0.870	0.926	0.952
			800	0.988	0.994	1.000	0.890	0.956	0.976	0.976	0.992	0.998

Table A.2: Finite sample rejection frequency of \hat{T}_n^{**} -based bootstrap test: standard normal weights

DGP	Δ_0	ρ	n	$\mathbb{H}_{0,\alpha}$			$\mathbb{H}_{0,\beta}$			$\mathbb{H}_{0,\alpha\beta}$		
				1%	5%	10%	1%	5%	10%	1%	5%	10%
1'	0	0.2	200	0.006	0.016	0.038	0.008	0.022	0.044	0.004	0.018	0.040
			400	0.008	0.032	0.060	0.012	0.024	0.050	0.008	0.024	0.034
			800	0.004	0.018	0.052	0.006	0.026	0.046	0.008	0.020	0.048
		0.5	200	0.004	0.032	0.056	0.008	0.022	0.044	0.006	0.022	0.034
			400	0.004	0.016	0.040	0.008	0.032	0.056	0.004	0.024	0.044
			800	0.008	0.022	0.042	0.006	0.018	0.038	0.004	0.016	0.038
		0.8	200	0.008	0.026	0.042	0.014	0.024	0.056	0.012	0.026	0.040
			400	0.000	0.018	0.042	0.004	0.022	0.054	0.002	0.010	0.038
			800	0.004	0.014	0.044	0.006	0.016	0.040	0.006	0.010	0.032
	1	0.2	200	0.998	1.000	1.000	0.676	0.856	0.896	0.980	0.988	0.990
			400	1.000	1.000	1.000	0.954	0.984	0.988	1.000	1.000	1.000
			800	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.5		200	0.994	1.000	1.000	0.672	0.854	0.896	0.966	0.982	0.990	
		400	0.998	1.000	1.000	0.934	0.986	0.994	0.996	0.998	0.998	
		800	1.000	1.000	1.000	0.998	0.998	0.998	1.000	1.000	1.000	
0.8		200	0.998	1.000	1.000	0.624	0.808	0.882	0.972	0.992	0.996	
		400	0.998	1.000	1.000	0.936	0.982	0.986	0.996	0.998	0.998	
		800	1.000	1.000	1.000	0.994	1.000	1.000	1.000	1.000	1.000	
3'	0	0.2	200	0.018	0.056	0.082	0.012	0.038	0.070	0.014	0.044	0.068
			400	0.004	0.016	0.048	0.004	0.018	0.038	0.002	0.014	0.040
			800	0.008	0.034	0.050	0.006	0.018	0.026	0.008	0.020	0.040
		0.5	200	0.008	0.044	0.098	0.018	0.036	0.058	0.022	0.040	0.068
			400	0.006	0.020	0.052	0.004	0.026	0.046	0.004	0.026	0.038
			800	0.006	0.040	0.078	0.010	0.024	0.066	0.006	0.030	0.060
		0.8	200	0.008	0.040	0.062	0.006	0.030	0.062	0.002	0.032	0.056
			400	0.010	0.026	0.044	0.010	0.042	0.064	0.010	0.030	0.056
			800	0.004	0.044	0.072	0.010	0.028	0.048	0.010	0.032	0.050
	1	0.2	200	0.904	0.960	0.968	0.234	0.398	0.510	0.666	0.790	0.850
			400	0.934	0.966	0.982	0.444	0.622	0.732	0.828	0.898	0.936
			800	0.992	0.998	0.998	0.896	0.962	0.978	0.978	0.992	0.996
0.5		200	0.868	0.914	0.934	0.240	0.370	0.488	0.668	0.756	0.814	
		400	0.954	0.976	0.986	0.504	0.706	0.796	0.858	0.942	0.956	
		800	0.988	0.994	0.998	0.888	0.950	0.976	0.978	0.988	0.998	
0.8		200	0.866	0.916	0.944	0.278	0.472	0.586	0.706	0.826	0.870	
		400	0.932	0.952	0.964	0.542	0.724	0.826	0.878	0.914	0.942	
		800	0.992	0.998	1.000	0.916	0.958	0.986	0.990	0.994	0.996	

References

- Bernstein, D. S., 2005. *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory*. Princeton University Press, Princeton.
- de Jong, P., 1987. A central limit theorem for generalized quadratic forms. *Probability Theory and Related Fields* 75, 261-277.
- He, X., Shao, Q., 2000. On parameters of increasing dimensions. *Journal of Multivariate Analysis* 73, 120-135.
- Horowitz, J., Lee, S., 2005. Nonparametric estimation of an additive quantile regression model. *Journal of the American Statistical Association*, 100, 1238-1249.
- Kato, K., Galvao, A. F., Montes-Rojas, G. V., 2012. Asymptotics for panel quantile regression models with individual effects. *Journal of Econometrics* 10, 76-91.
- Kline, P. Santos, A., 2012. A score based approach to wild bootstrap inference. *Journal of Econometric Methods* 1, 23-41.
- Koenker, R., Basset, G., 1978. Regression quantiles. *Econometrica* 46, 33-50.
- Koenker, R., Zhao, Q., 1996. Conditional quantile estimation and inference for ARCH models. *Econometric Theory* 12, 793-813.
- Newey, W. K., 1997. Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics* 79, 147-168.
- Ruppert, D., Carroll, R.J., 1980. Trimmed least squares estimation in the linear model. *Journal of the American Statistical Association* 75, 828-838.
- Su, L., Jin, S., 2012. Sieve estimation of panel data models with cross section dependence. *Journal of Econometrics* 169, 34-47.