Lancaster University Management School

Economics Working Paper Series

2018/010

Inference in Nonparametric Series Estimation with Specification Searches for the Number of Series Terms

Byunghoon Kang

LUMS home page: http://www.lancaster.ac.uk/lums/

Inference in Nonparametric Series Estimation with Specification Searches for the Number of Series Terms

Byunghoon Kang*

Department of Economics, Lancaster University

First version December 9, 2014; Revised March 31, 2018

Abstract

Nonparametric series estimation often involves specification search over the different number of series terms due to the unknown smoothness of underlying function. This paper considers pointwise inference in the nonparametric series regression for the conditional mean and introduces test based on the supremum of t-statistics over different series terms. I show that proposed test has correct asymptotic size and it can be used to construct confidence intervals that have correct asymptotic coverage probability uniform in the number of series terms. With possibly large bias in this setup, I also consider infimum of the t-statistics which is shown to reduce size distortions in such case. Asymptotic distribution of the test statistics, asymptotic size, and local power results are derived. I investigate the performance of the proposed tests and CIs in various simulation setups as well as an illustrative example, nonparametric estimation of wage elasticity of the expected labor supply from Blomquist and Newey (2002). I also extend our inference methods to the partially linear model setup.

Keywords: Nonparametric series regression, Pointwise confidence interval, Smoothing parameter choice, Specification search, Undersmoothing.

JEL classification: C12, C14.

1 Introduction

I consider the following nonparametric regression model;

$$y_i = g_0(x_i) + \varepsilon_i,$$
$$E(\varepsilon_i|x_i) = 0$$

(1.1)

where $\{y_i, x_i\}_{i=1}^n$ is i.i.d. with scalar response variable y_i, vector of covariates $x_i \in \mathbb{R}^{d_x}$, and $g_0(x) = E(y_i|x_i = x)$ is the conditional mean function. Theory of estimation and inference is well developed for nonparametric series (sieves) methods in large econometrics and statistics literature,[1] and series estimation has also received attention in applied economics as it can easily impose shape restrictions such as additive separability or monotonicity. However, implementation requires a choice of smoothing parameter, the number of series terms $K = K_n$, and this often involves specification searches, i.e., search over different series terms $K \in \mathcal{K}_n = [\underline{K}, \bar{K}]$. Existing theory for the asymptotic normality and valid inference imposes so-called undersmoothing (i.e., overfitting) condition that is a faster rate of K than the mean-squared error (MSE) optimal convergence rates, however, to achieve undersmoothing one has to know the smoothness of $g_0(x)$ which is usually unknown in practice. Specification searches seem necessary in such case, but it may lead to misleading inference without taking into account the first step specification search.

This paper provides inference methods in nonparametric series regression given the range of the different number of series terms. I consider the testing problem for a regression function at a point, and I show that (1) test based on the absolute value of supremum of the studentized t-statistics over different series terms and its asymptotic critical value control the asymptotic size; (2) this test can be used to construct asymptotically valid confidence intervals (CI) which are uniform in $K \in \mathcal{K}_n$; (3) in consequence, CI with any $\widehat{K} \in \mathcal{K}_n$ has a correct asymptotic coverage for $g_0(x)$ by adjusting the conventional normal critical value to the critical value from supremum of the t-statistics.

The main contribution of this paper is to derive a uniform asymptotic distribution theory for the entire sequences of t-statistics over a range of the number of series terms K. Existing asymptotic normality of the t-statistic in the literature holds under a deterministic sequence of $K \to \infty$ as $n \to \infty$. First, I develop an asymptotic distribution of t-statistics over the set $\mathcal{K}_n$ that includes a finite number of sequences allowing a broad range of K from oversmoothing rates to undersmoothing rates as well as optimal MSE rates. In addition, I provide an empirical process theory for the t-statistics indexed by continuous parameter π, a fraction of the largest series terms $\bar{K}$, by considering different $\mathcal{K}_n$; although this set assumption only allow same rates of K (put it differently, $\mathcal{K}_n$ can be small), it is sufficient to show the weak convergence of the empirical process. I also show that, in the special case, limiting Gaussian process coincides with the scaled Brownian motion process so that the asymptotic critical value of the test statistics can be tabulated easily as it is only a function of $\underline{\pi} = \underline{K}/\bar{K}$ with the smallest $\underline{K}$ and the largest $\bar{K}$, and this can be viewed

---

[1]For examples, Andrews (1991a), Newey (1997), Huang (2003a), Chen (2007), Belloni, Chernozhukov, Chetverikov, and Kato (2015), and Chen and Christensen (2015b), among many others.

as an analogous result developed in the kernel estimation literature (for example, Armstrong and Kolesár (2015)).

With an asymptotic distribution theory established in the paper, I also consider tests based on infimum of the t-statistics, and searching for "small" t-statistics in this setup has a similar motivation with the undersmoothing assumption; using "large" K (or faster rates of K than the optimal MSE rate) which has a small bias and large variance, thus lead to "small" t-statistics. For valid inference, many papers in nonparametric series estimation literature typically suggested to increase the number of series terms and include additional terms than that cross-validation chooses (for example, see Newey, Powell, and Vella (1999), Newey (2013)), mainly due to the lack of an explicit asymptotic bias formula and bias-corrections for the series estimator. I formally justify this conventional wisdom by introducing the infimum test statistic over $\mathcal{K}_n$ and provide an inference method based on its asymptotic distribution. We can, in principle, consider other "small" t-statistics (e.g., 2nd smallest or median), but a valid pointwise CI can be constructed easily by inverting infimum t-statistics; it is obtained as the union of all CIs by replacing the standard normal critical value with the critical value from the infimum t-statistic.

This paper also establishes asymptotic local power results for a particular sequence of slower than $n^{-1/2}$-local alternatives and show the size-power trade-off between supremum and infimum test statistics; similar to using undersmoothing rates, the test based on the infimum t-statistics can reduce size distortion, but may also reduce the power of the tests. Formally, I show that the test based on the supremum test statistics is consistent against some sequences of local alternatives while infimum test statistic is consistent only against a limited range of local alternatives. However, the test based on infimum t-statistics controls the asymptotic size even allowing large asymptotic bias, and it has nontrivial power against some local alternatives, and has better power than the test based on single t-statistic using normal critical value with "large" K that is faster than or equal to the optimal MSE rates.

I also provide a construction of valid CIs by test statistic inversion and show coverage results of the proposed CIs. The critical values can be easily implemented using the asymptotic Gaussian distribution. I investigate finite sample coverage and length properties of the proposed CIs in various simulation setups. As an illustrative example, I revisit nonparametric estimation of labor supply function using entire individual piecewise-linear budget set as in Blomquist and Newey (2002). Imposing additive separability, derived by economic theory, Blomquist and Newey (2002) estimate conditional mean of labor supply function using series estimation and report wage elasticity of the expected labor supply as well as other welfare measures with various specifications of the different number of series terms.

Finally, I provide inference methods in partially linear model setup focusing on the common parametric part. Unlike the nonparametric object of interest that has a slower convergence rate than $n^{1/2}$ (e.g., regression function or regression derivative), t-statistics for the parametric object of interest are asymptotically equivalent for all sequences of K under standard rate conditions $K/n \to 0$ as $n \to \infty$. To account the dependency of the t-statistics with the different sequences of

Ks in this setup, I develop an asymptotic distribution of the studentized t-statistics over $K \in \mathcal{K}_n$ using the results of Cattaneo, Jansson, and Newey (2015a) under the faster rate of K that grows as fast as the sample size n. I also discuss methods to construct CIs that are similar to the nonparametric regression setup and provide coverage properties.

The supremum t-statistics has been used as a correction for the multiple testing problems or to construct simultaneous confidence bands, and the importance of multiple testing problems (data mining or data snooping) has long been alerted in various other contexts (see Leamer (1983), White (2000), Romano and Wolf (2005), Hansen (2005)). While considering infimum t-statistics is closely related to the undersmoothing assumption that conceptually requires increasing K until t-statistic is "small enough" in our nonparametric setup, it is also related to the "stepup" procedures in the multiple testing literature (see, for example, Romano and Shaikh (2006)) which start by considering smallest test statistics (the least significant hypotheses) and then move up to the larger test statistics. Furthermore, inference methods using the union of confidence intervals also has been used in various contexts, for examples, instrumental variable (IV) setup without imposing instrument exclusion restriction as in Conley, Hansen, and Rossi (2012), sensitivity analysis in parametric setup as in Levine and Renelt (1992) following Leamer (1983). With the bias-variance trade-off of nonparametric estimators in our different setup, we provide less conservative inference methods from our asymptotic distribution results of the infimum test statistic with the critical values smaller than the standard normal critical values.

Several important papers have investigated the asymptotic properties of series (and sieves) estimators, including papers by Andrews (1991a), Eastwood and Gallant (1991), Newey (1997), Chen and Shen (1998), Huang (2003a), Chen (2007), Chen and Liao (2014), Chen, Liao, and Sun (2014), Belloni et al. (2015), and Chen and Christensen (2015b), among many others. This paper extends the asymptotic normality of the t-statistic under a single sequence of K to the uniform central limit theorem of the t-statistic for the sequences of K over a set $\mathcal{K}_n$, and focuses on a pointwise inference on $g_0(x)$ which is an irregular (i.e., slower than $n^{1/2}$ rate) and linear functional under i.i.d. setup.

There is also a growing literature on data-dependent series term selection and its impact on estimation and inference in econometrics and statistics. Asymptotic optimality results of cross-validation have been developed, including papers by Li (1987), Andrews (1991b), Hansen (2015). Recent papers by Horowitz (2014), Chen and Christensen (2015a) develop data-driven methods for choosing sieve dimension in the nonparametric instrumental variables (NPIV) estimation so that resulting NPIV estimators attain the optimal sup-norm or L^2 norm rates adaptive to the unknown smoothness of $g_0(x)$. Although we do not pursue adaptive inference in this paper, there is also a large statistics literature on adaptive inference (see Giné and Nickl (2015, Section 8) for comprehensive lists of references). For example, Giné and Nickl (2010), Chernozhukov, Chetverikov, and Kato (2014a) construct adaptive confidence bands in density estimation problem. For regression setup, similar techniques developed in Chernozhukov et al. (2014a) can be used to approximate distributions of suprema of empirical processes with Gaussian multiplier bootstrap, but obtaining

asymptotic distributions in this paper can be useful to consider other test statistics, such as infimum t-statistic, in pointwise inference. Furthermore, once data-driven choice is obtained for adaptive estimation (e.g., Lepski (1990)-type procedures), one still require undersmoothing condition for inference to eliminate asymptotic bias terms (see Theorem 1 of Giné and Nickl (2010), and Corollary 3.1 of Chernozhukov et al. (2014a)), and this may come up with similar specification search issues to choose sufficiently "large" K in practice.

We can, in principle, consider kernel-based estimation where several data-dependent bandwidth selections or explicit bias-corrections have been proposed,[2] however, there exist many examples estimating $g_0(x)$ using series estimation with imposing shape constraints easily (such as additive separability) and also interested in pointwise inference (see Section 7 for the example as in Blomquist and Newey (2002)). Unlike the kernel-based methods, little is known about the statistical properties of data-dependent selection rules or an asymptotic theory for the bias-correction due to the lack of an explicit asymptotic bias formula for the series estimator.[3] With the issues of specification search, this paper is closely related to a recent paper by Armstrong and Kolesár (2015) which considers similar inference methods for $g_0(x)$ using kernel estimation with bandwidth snooping. We provide coverage results that are uniform in series terms, as well as analogous limiting distributions for the supremum of Gaussian processes as in Armstrong and Kolesár (2015).

The rest of the paper is organized as follows. I introduce basic nonparametric series regression setup and the asymptotic distribution in Section 2 and the asymptotic size and power properties of the test statistics in Section 3. In Section 4, I introduce CIs, implementation of the critical values and provide coverage results. Section 5 extends our inference methods to the partially linear model setup. Section 6 summarizes Monte Carlo experiments in various setups, and Section 7 illustrates empirical example as in Blomquist and Newey (2002), then Section 8 concludes. Appendix A includes all proofs, and Appendix B includes figures and tables.

1.1 Notation

I introduce some notation will be used in the following sections. I use $||A|| = \sqrt{tr(A'A)}$ for the Euclidean norm. Let $\lambda_{min}(A), \lambda_{max}(A)$ denote the minimum and maximum eigenvalues of a symmetric matrix A, respectively. $o_p(\cdot)$ and $O_p(\cdot)$ denote the usual stochastic order symbols, convergence in probability and bounded in probability. $\xrightarrow{d}$ denotes convergence in distribution and $\Rightarrow$ denotes weak convergence. I use the notation $a \wedge b = \min\{a, b\}, a \vee b = \max\{a, b\}$, and denote $\lfloor a \rfloor$ as the largest integer less than the real number a. For two sequences of positive real numbers a_n and b_n, $a_n \lesssim b_n$ denotes $a_n \leq cb_n$ for all n sufficiently large with some constant $c > 0$ that is independent of n. $a_n \asymp b_n$ denotes $a_n \lesssim b_n$ and $b_n \lesssim a_n$. For a given random variable $\{X_i\}$ and $1 \leq p < \infty$, $L^p(X)$ is the space of all L^p norm bounded functions with $||f||_{L^p} = [E||f(X_i)||^p]^{1/p}$ and $\ell^\infty(X)$ denotes the space of all bounded functions under sup-norm, $||f||_\infty = \sup_{x \in \mathcal{X}} |f(x)|$ for

---

[2]See Härdle and Linton (1994), Li and Racine (2007) for references. See also Hall and Horowitz (2013), Calonico, Cattaneo and Farrell (2015), Schennach (2015) and references therein for various recent works on related bias issues and inference for the kernel estimator.

[3]Zhou, Shen, and Wolfe (1998) and Huang (2003b) for exceptions with some specific sieves.

the bounded real-valued functions f on the support $\mathcal{X}$. Let also $\mathbb{R}_{[\pm\infty]} = \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$.

2   Model framework and asymptotic distribution

I first introduce the nonparametric series regression setup in the model (1.1). Given a random sample $\{y_i, x_i\}_{i=1}^n$, we are interested in the conditional mean $g_0(x) = E(y_i|x_i = x)$ at a point $x \in \mathcal{X} \subset \mathbb{R}^{d_x}$. All the results derived in this paper are the pointwise inference in x, and I will omit the dependence on x if there is no confusion.

Let $\widehat{g}_n(K, x)$ be an estimator of $g_0(x)$ using the first $K = K_n \geq 1$ series terms $P_K(x) = (p_1(x), \cdots, p_K(x))'$ from basis functions $p(x) = (p_1(x), p_2(x), \cdots)'$. Standard examples for the basis functions are power series, Fourier series, orthogonal polynomials (e.g., Hermite polynomials), or splines with evenly sequentially spaced knots. Series estimator is then obtained by standard least square (LS) estimation of y_i on regressors $P_K(x_i)$

$$\widehat{g}_n(K, x) = P_K(x)'\widehat{\beta}_K, \qquad \widehat{\beta}_K = (P^{K'}P^K)^{-1}P^{K'}Y \tag{2.1}$$

where $P^K = [P_{K1}, \cdots, P_{Kn}]', P_{Ki} \equiv P_K(x_i) = (p_1(x_i), p_2(x_i), \cdots, p_K(x_i))', Y = (y_1, \cdots y_n)'$. For simplicity of notation, I define the true regression function at a point as $\theta_0 \equiv g_0(x)$ and let $\widehat{\theta}_n(K) \equiv \widehat{g}_n(K, x)$. Define the series variance

$$\begin{aligned} V_n(K) &\equiv V_n(K, x) = P_K(x)'Q_K^{-1}\Omega_K Q_K^{-1}P_K(x), \\ Q_K &= E(P_{Ki}P_{Ki}'), \quad \Omega_K = E(P_{Ki}P_{Ki}'\varepsilon_i^2) \end{aligned} \tag{2.2}$$

where $Q_K^{-1}\Omega_K Q_K^{-1}$ is the conventional asymptotic variance formula for the LS estimator $\widehat{\beta}_K$.

We use the notion of testing setup and consider two-sided testing for θ

$$H_0 : \theta = \theta_0, \quad H_1 : \theta \neq \theta_0. \tag{2.3}$$

The studentized t-statistic for H_0 is

$$T_n(K, \theta_0) \equiv \frac{\sqrt{n}(\widehat{g}_n(K, x) - g_0(x))}{V_n(K, x)^{1/2}} = \frac{\sqrt{n}(\widehat{\theta}_n(K) - \theta_0)}{V_n(K)^{1/2}}. \tag{2.4}$$

Under standard regularity conditions (will be discussed below in Assumption 2.2), t-statistic under H_0 can be decomposed as follows

$$T_n(K, \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{P_K(x)'Q_K^{-1}P_{Ki}\varepsilon_i}{V_n(K)^{1/2}} - \frac{r_n(K)}{\sqrt{V_n(K)/n}} + o_p(1) \tag{2.5}$$

where the first term converges to a standard normal distribution for deterministic sequence $K \to \infty$ as $n \to \infty$, and the second term does not necessarily converge to 0 due to approximation errors, $r_n(K) \equiv r_n(K, x) = g_0(x) - P_K(x)'\beta_K$ with the best linear L^2 projection coefficients $\beta_K \equiv$

6

$(E[P_{Ki}P'_{Ki}])^{-1}E[P_{Ki}y_i]$. With undersmoothing assumption, the asymptotic distribution of the t-statistic, $T_n(K, \theta_0) \xrightarrow{d} N(0, 1)$, is well known in the literature (See, for example, Andrews (1991a), Newey (1997), Belloni et al. (2015), Chen and Christensen (2015b) among many others), and the confidence interval for the nonparametric regression function can be constructed.

Here, I develop an asymptotic distribution theory of $T_n(K, \theta_0)$ over a set $\mathcal{K}_n$. The following set assumption is constructed to allow a broad range of Ks so that $\mathcal{K}_n$ can allow (unknown) optimal mean squared error rate of K as well as oversmoothing rate which increases slower than the optimal MSE rate.

Assumption 2.1. *(Set of number of series terms) Let* $\mathcal{K}_n$ *as*

$\mathcal{K}_n = \{\underline{K} = K_1, \cdots, K_m, \cdots, \bar{K} = K_M\}$ *where* $K_m \equiv \lfloor \tau n^{\phi_m} \rfloor$ *for constant* $\tau > 0$, $0 < \phi_1 < \phi_2 < \cdots < \phi_M$, *and fixed* M. *Define asymptotic bias for the m-th model as* $\nu(m) \equiv -\lim_{n \to \infty} \sqrt{n} V_n(K_m)^{-1/2} r_n(K_m)$. *Assume that* $\inf_m |\nu(m)| = O(1)$.

Remark 2.1. Suppose $g_0(x)$ belongs to the Hölder space of smoothness $s > 0$, $\Sigma(s, \mathcal{X})$, then we obtain optimal L^2 convergence rates $O_p(n^{-s/(2s+d_x)})$ with $K \asymp n^{d_x/(d_x+2s)}$. Due to the unknown smoothness of $g_0(x)$, $\mathcal{K}_n$ may include small Ks corresponding to oversmoothing rates. If the unknown optimal MSE rates satisfy $\phi_p < d_x/(d_x + 2s) \leq \phi_{p+1}$ for some $p \in \{1, \cdots, M-1\}$, then Assumption 2.1 contains oversmoothing rates of K ($|\nu(m)| = +\infty$ for all $m = 1, \cdots, p$), as well as optimal MSE rate and undersmoothing rates ($|\nu(p+1)| < +\infty$, $|\nu(m)| = 0$ for all $m = p+2, \cdots, M$). Also note that $\inf_m |\nu(m)| = O(1)$ excludes the possibility of $|\nu(m)| = +\infty$ for all $m = 1, \cdots, M$.[4] We may consider growing set of the possible number of series terms with $\mathcal{K}_n = [\underline{K}, \bar{K}] \cap \mathbb{N}$ (similar assumption is used in the literature, for example, Newey (1994a, 1994b)), but this complicates the derivation of the limiting distribution. We can use the general results of Chernozhukov, Chetverikov, and Kato (2013) in this case, however, obtaining asymptotic distributions over $K \in \mathcal{K}_n$ is useful in our pointwise inference setup to consider other test statistics. We also consider different $\mathcal{K}_n$ including infinite sequences of K in Section 2.1 to establish weak convergence results.

Next, I impose mild regularity conditions that are standard in nonparametric series regression literature and are satisfied by well-known basis functions. For each $K \in \mathcal{K}_n$, define $\zeta_K \equiv \sup_{x \in \mathcal{X}} ||P_K(x)||$ as the largest normalized length of the regressor vector and $\lambda_K \equiv (\lambda_{min}(Q_K))^{-1/2}$ for $K \times K$ design matrix $Q_K = E(P_{Ki}P'_{Ki})$.

Assumption 2.2. *(Regularity conditions)*

(i) $\{y_i, x_i\}_{i=1}^n$ *are i.i.d random variables satisfying the model* (1.1).

---

[4]Assumption 2.1 allows to have optimal L^2 rates of K in a large set of classes of functions; by setting $\phi_m = d_x/(d_x + 2s_m)$ for all $s_m \in S = \{\underline{s} = s_1, \cdots, \bar{s} = s_M\}$, Assumption 2.1 contains the number of series terms which obtain optimal L^2 rate of convergence for every $g_0(x) \in \bigcup_{s \in S} \Sigma(s, \mathcal{X})$. A similar assumption is used in the literature on adaptive inference, although we do not pursue this direction and we only consider finite sequences of smoothing parameters.

(ii) $\sup_{x \in \mathcal{X}} E(\varepsilon_i^2 | x_i = x) < \infty$, $\inf_{x \in \mathcal{X}} E(\varepsilon_i^2 | x_i = x) > 0$, and $\sup_{x \in \mathcal{X}} E(\varepsilon_i^2 \{|\varepsilon_i| > c(n)\} | x_i = x) \to$
    0 *for any sequence* $c(n) \to \infty$ *as* $n \to \infty$.

(iii) For each $K \in \mathcal{K}_n$, *as* $K \to \infty$, *there exists* η *and* c_K, ℓ_K *such that*

$$\sup_{x \in \mathcal{X}} |g_0(x) - P_K(x)'\eta| \leq \ell_K c_K, \quad E[(g_0(x_i) - P_K(x_i)'\eta)^2]^{1/2} \leq c_K.$$

(iv) $\sup_{K \in \mathcal{K}_n} \lambda_K \lesssim 1$.

(v) $\sup_{K \in \mathcal{K}_n} \zeta_K \sqrt{(\log K)/n}(1 + \sqrt{K}\ell_K c_K) + \ell_K c_K \to 0$ *as* $n \to \infty$.

Assumption 2.2(ii) imposes moment conditions and standard uniform integrability conditions. Assumption 2.2(iii)-(v) are similar to those imposed in Belloni et al. (2015), Chen and Christensen (2015b) except that we impose rate conditions of K uniformly over $\mathcal{K}_n$. Other standard regularity conditions in the literature (e.g., Newey (1997), Chen (2007)) can also be used here, so ζ_K, c_K, ℓ_K in Assumption 2.2(iii)-(v) are satisfied with various sieve bases and all the discussions made there also apply here.[5]

Next theorem provides an asymptotic distribution under Assumption 2.1 with a broad range of K. Allowing $\sup_m |\nu(m)| = +\infty$ in the result will be useful for considering the effect of bias on the asymptotic size of the test, as well as for obtaining asymptotic distributions of the test statistics under the wide range of local alternatives. For this, we define the continuous function on the extended real space as follows; $S : A \to B$ is continuous at $t \in A$ if $t' \to t$ for $t \in A$ implies $S(t') \to S(t)$ for any set A.

Theorem 2.1. *Under Assumptions 2.1 and 2.2, following holds for any continuous function* $S(t)$ *at all* $t \in \mathcal{T} \subset \mathbb{R}^M_{[\pm\infty]}$,

$$S(T_n(\theta_0)) \xrightarrow{d} S(Z_\Sigma + \nu)$$

where $T_n(\theta) = (T_n(K_1, \theta), \cdots, T_n(K_M, \theta))'$, $Z_\Sigma = (Z_1, \cdots, Z_M)' \sim N(0, \Sigma)$, $\nu = (\nu(1), \cdots, \nu(M))'$ *are* $M \times 1$ *vectors provided that* Σ *exists and is a finite positive definite matrix with* $\Sigma(j, l) = \lim_{n \to \infty} \Sigma_n(j, l)$, *and*

$$\Sigma_n(j, l) = \frac{P_{K_j}(x)' E(P_{K_j i} P'_{K_j i})^{-1} E(P_{K_j i} P'_{K_l i} \varepsilon_i^2) E(P_{K_l i} P'_{K_l i})^{-1} P_{K_l}(x)}{V_n(K_j)^{1/2} V_n(K_l)^{1/2}}.$$

---

[5]For example, if the support $\mathcal{X}$ is a cartesian product of compact connected intervals (e.g. $\mathcal{X} = [0, 1]^{d_x}$) and the probability density of x_i is bounded below zero, then $\zeta_K \lesssim K$ for power series and other orthogonal polynomial series, and $\zeta_K \lesssim \sqrt{K}$ for regression splines, Fourier series, and wavelet series. Moreover, c_K and ℓ_K can be replaced by series specific bounds. If $g_0(x) \in \Sigma(s, \mathcal{X}), s > 0$, then $c_K \lesssim K^{-s/d_x}, \ell_K \lesssim K$ for power series, $c_K \lesssim K^{-(s \wedge s_0)/d_x}, \ell_K \lesssim 1$ for spline and wavelet series of order s_0. Furthermore, when the probability density function of x_i is uniformly bounded above and bounded away from zero over the compact support $\mathcal{X}$ and orthonormal basis is used, then $\lambda_K \lesssim 1$ (see, for example, Proposition 2.1 in Belloni et al. (2015) and Remark 2.2 in Chen and Christensen (2015b)). Then, the rate conditions in Assumption 2.2(v) can be replaced by the specific bounds of ζ_K, c_K, ℓ_K. For example, for the power series, Assumption 2.2(v) reduced to $\sup_{K \in \mathcal{K}_n} \sqrt{K^2(\log K)/n}(1 + K^{3/2-s/d_x}) + K^{1-s/d_x} \to 0$.

If $\nu(m) = \pm\infty$, then the corresponding element of $Z_\Sigma + \nu$ equals $\pm\infty$.

This is an asymptotic theory for the entire sequence of t-statistics $T_n(K, \theta_0), K \in \mathcal{K}_n$. Note that standard inference methods in nonparametric regression setup typically consider a singleton set $\mathcal{K}_n = \{K\}$ with $K \to \infty$ as $n \to \infty$.

Remark 2.2. We note that Theorem 2.1 does not require either undersmoothing assumption for all $K \in \mathcal{K}_n$ or $\sup_m |\nu(m)| < \infty$. The t-statistics may not converge in distribution to a bounded random vector if $|\nu(m)| = +\infty$ for some m; corresponding t-statistic $T_n(K_m, \theta_0)$ diverges in probability to $\pm\infty$. This matters when we obtain the asymptotic distribution of the test statistic because continuous mapping theorem cannot be directly applied. To obtain the asymptotic distribution under $\sup_m |\nu(m)| = +\infty$, we provide formal proofs which combine arguments in inference on CIs for the parameters in moment inequality literature as in Andrews and Guggenberger (2009).[6] Theorem 2.1 coincides with standard continuous mapping theorem when we control maximum asymptotic bias or impose uniform undersmoothing assumption (i.e., $\sup_m |\nu(m)| = 0$), and thus useful to analyze the effect of large bias on the size of the tests, and to compare the power of asymptotic level α tests under wide range of local alternatives (see Section 3.1 and 3.2 for details).

Remark 2.3 (Variance-covariance matrix)**.** Under conditional homoskedasticity, $E(\varepsilon_i^2 | x_i = x) = \sigma^2$, variance-covariance matrix Σ reduces to the simple form as $\Sigma(j, l) = \lim_{n \to \infty} \frac{V_n(K_j \wedge K_l)^{1/2}}{V_n(K_j \vee K_l)^{1/2}}$, i.e., the limit of the ratio of standard deviations. If we assume $V_n(K)^{1/2} \asymp K^\eta$ at some point x with $\eta > 0$, then for any $j < l$, $\Sigma_n(j, l) \leq C \frac{V_n(K_j)^{1/2}}{V_n(K_l)^{1/2}}$ for some constant $C > 0$ by Assumption 2.2(ii) and the upper bound converges to 0 as $n \to 0$ by Assumption 2.1, thus $\Sigma_n(j, l) \to 0$ and variance-covariance matrix also reduces to $\Sigma = I_M$.

Remark 2.4 (Other functionals)**.** Here, we focus on the leading example, where $\theta_0 = g_0(x)$ for some fixed point $x \in \mathcal{X}$, but we can consider other linear functionals $\theta_0 = a(g_0(\cdot))$, such as the regression derivates $a(g_0(x)) = \frac{d}{dx} g_0(x)$. All the results in this paper can be applied to irregular (slower than $n^{1/2}$ rate) linear functionals using estimators $\widehat{\theta}_n(K) = a(\widehat{g}_n(K, x)) = a_K(x)' \widehat{\beta}_K$ and appropriate transformation of basis $a_K(x) = (a(p_1(x)), \cdots, a(p_K(x)))'$ with proper smoothness condition on the functional and continuity conditions on the derivative as in Newey (1997). While verification of previous results for regular ($n^{1/2}$ rate) functionals, such as integrals and weighted average derivative, is beyond the scope of this paper, I examine similar results for the partially linear model setup in Section 5.

Remark 2.5 (Largest K)**.** Although there exist rate restrictions for $\bar{K}$ to be used for the asymptotic normal approximation, formal guidance or data-dependent results for the range of $\mathcal{K}_n = [\underline{K}, \bar{K}]$ are beyond the scope of this paper. Together with Assumption 2.2, there exist explicit rate restrictions on ϕ_m in the Assumption 2.1 uniformly over m to guarantee asymptotic normality of single t-statistic. For example, under $\sup_{K \in \mathcal{K}_n} K^2 \log K / n \to 0$ for the power series, rates are allowed only

---

[6]This technical difficulty does not arise when we consider t-statistics centered at the pseudo-true value $\theta(K) = P_K(x)' \beta_K$ (or $\theta_0 - r_n(K)$). But we do not pursue this approach because (1) we are mainly interested in conditional mean function (θ_0); (2) it is difficult to implement bias-correction in this setup.

if $\phi_M < 1/2$. Some formal data-dependent methods to obtain optimal L^2 norm or sup-norm rates, such as cross-validation can be useful guidelines for $\mathcal{K}_n$, and it would be of interest to extend all inference results developed in this paper with data-dependent $\mathcal{K}_n$.

2.1 Weak convergence of t-statistic process

In this section, I consider different set assumption $\mathcal{K}_n$ to construct empirical process theory of the t-statistics over $K \in \mathcal{K}_n$ which I shall call 't-statistic process'. The t-statistic process is indexed by the continuous and fixed parameter π, which is a fraction of the largest series term $\bar{K}$.

Assumption 2.3. *(Same Rates of K) Let $\mathcal{K}_n$ as*

$$\mathcal{K}_n = \{K : K \in [\underline{\pi}\bar{K}, \bar{K}]\}$$

where $K = \pi\bar{K}$ with a fixed constant $\pi \in \Pi = [\underline{\pi}, \bar{\pi}]$, $0 < \underline{\pi} < \bar{\pi} = 1$, and $\bar{K} \equiv \bar{K}(n) \to \infty$ as $n \to \infty$.

Assumption 2.3 considers a range of the number of series terms and considers an infinite sequence of approximations indexed by $\pi \in \Pi$. By imposing the rate conditions for the largest series term $\bar{K}$ with the Assumption 2.2, set $\mathcal{K}_n$ considers the sequence of models that has the same rate of K, i.e., $K \asymp K'$ for any $K, K' \in \mathcal{K}_n$.

Next, define the following empirical process, $T_n^*(\pi, \theta)$, as

$$T_n^*(\pi, \theta) \equiv T_n(\lfloor \pi\bar{K} \rfloor, \theta), \quad \pi \in \Pi, \tag{2.6}$$

where $T_n(K, \theta)$ is defined in (2.4), i.e., $T_n^*(\pi, \theta)$ is a t-statistic evaluated at the parameter θ using $\lfloor \pi\bar{K} \rfloor$ number of series terms. Note that $T_n^*(\pi, \theta)$ is indexed by $\pi \in \Pi$ and is a step function of π. This empirical process has a covariance kernel

$$\Sigma_n(\pi_1, \pi_2) \equiv \frac{P_{\pi_1}(x)' Q_{\pi_1}^{-1} E(P_{\pi_1 i} P_{\pi_2 i}' \varepsilon_i^2) Q_{\pi_2}^{-1} P_{\pi_2}(x)}{V_n(\pi_1)^{1/2} V_n(\pi_2)^{1/2}}, \quad \pi_1, \pi_2 \in \Pi, \tag{2.7}$$

where $P_\pi(x) \equiv P_{\lfloor \bar{K}\pi \rfloor}(x)$, $P_{\pi i} \equiv P_\pi(x_i) = P_{\lfloor \bar{K}\pi \rfloor i}$, and the series variance $V_n(\pi) \equiv V_n(\lfloor \pi\bar{K} \rfloor, x) = ||\Omega_\pi^{1/2} Q_\pi^{-1} P_\pi(x)||^2$, $\Omega_\pi = E(P_{\pi i} P_{\pi i}' \varepsilon_i^2)$, $Q_\pi = E(P_{\pi i} P_{\pi i}')$. We expect that the limiting Gaussian process has a covariance function as a limit of the sequence of covariance functions $\Sigma_n(\pi_1, \pi_2)$, which is assumed to exist by the following assumption. To establish weak convergence of the empirical process $\{T_n^*(\pi, \theta_0) : \pi \in \Pi\}$, I also impose rate restrictions on series variances.

Assumption 2.4.

(i) $\Sigma(\pi_1, \pi_2) = \lim_{n \to \infty} \Sigma_n(\pi_1, \pi_2)$ *exists and $\Sigma(\pi_1, \pi_2) < 1$ for any $\pi_1, \pi_2 \in \Pi$.*

(ii) $\lim_{n \to \infty} V_n(\pi)^{1/2} (\bar{K}\pi)^{-\eta} = c$ *uniformly in $\pi \in \Pi$ for some constants $c, \eta > 0$.*

Assumption 2.4(i) guarantees well-defined covariance function for the tight Gaussian process in $\ell^\infty(\Pi)$ and a positive definite variance-covariance matrix for its finite-dimensional limit distributions. Assumption 2.4(ii) is a high-level assumption but required to prove weak convergence of the t-statistic process. Assumption 2.4(ii) requires that series variance is increasing in K at some rates uniformly in $K \in \mathcal{K}_n$ under Assumption 2.3, i.e., $\lim_{n\to\infty} V_n(K)^{1/2} K^{-\eta} = c$. This assumption holds with $\eta = 1/2$ when we consider a point $x \in \mathcal{X}$ where $V_n(K)^{1/2} \propto K^{1/2}$. Moreover, Assumption 2.4(ii) is a sufficient condition for Assumption 2.4(i) under homoskedasticity. Under conditional homoskedasticity, $E(\varepsilon_i^2 | x_i = x) = \sigma^2$, the limit of the covariance kernel $\Sigma(\pi_1, \pi_2)$ reduces to the

$$\Sigma(\pi_1, \pi_2) = \lim_{n\to\infty} \frac{V_n(\pi_1 \wedge \pi_2)^{1/2}}{V_n(\pi_1 \vee \pi_2)^{1/2}} \tag{2.8}$$

for any $\pi_1, \pi_2 \in \Pi$. If we further assume Assumption 2.4(ii), then $\Sigma(\pi_1, \pi_2) = (\frac{\pi_1 \wedge \pi_2}{\pi_1 \vee \pi_2})^\eta$. With $\eta = 1/2$, this coincides with the covariance kernel of a scaled Brownian motion process $\mathbb{Z}(\pi)/\sqrt{\pi}$, $\pi \in \Pi$.

Next, we define the asymptotic bias as $\nu(\pi) \equiv \lim_{n\to\infty} -\sqrt{n} V_n(\pi)^{-1/2} r_n(\pi)$ for $\pi \in \Pi$, where $r_n(\pi) \equiv r_n(\lfloor \pi \bar{K} \rfloor, x)$ and consider undersmoothing assumption. Under Assumption 2.5, $\nu(\pi) = 0$ for all $\pi \in \Pi$.[7]

Assumption 2.5. *(Undersmoothing)* $\sup_{K \in \mathcal{K}_n} |\sqrt{n} V_n(K)^{-1/2} r_n(K)| \to 0$ *as* $n \to \infty$.

Next theorem provides uniform central limit theorem of the t-statistic process for nonparametric LS series estimation.

Theorem 2.2. *Under Assumptions 2.2, 2.3, 2.4 and* $\sup_\pi |\nu(\pi)| < \infty$,

$$T_n^*(\pi, \theta_0) \Rightarrow \mathbb{T}(\pi) + \nu(\pi), \quad \pi \in \Pi, \tag{2.9}$$

where $\mathbb{T}(\pi)$ *is a mean zero Gaussian process on* $\ell^\infty(\Pi)$ *with covariance kernel,* $E(\mathbb{T}(\pi_1)\mathbb{T}(\pi_2)) = \Sigma(\pi_1, \pi_2)$ *for any* $\pi_1, \pi_2 \in \Pi$, *defined in Assumption 2.4. In addition, if Assumption 2.5 is satisfied, then*

$$T_n^*(\pi, \theta_0) \Rightarrow \mathbb{T}(\pi), \quad \pi \in \Pi. \tag{2.10}$$

Theorem 2.2 provides weak convergence of the t-statistic process $T_n^*(\pi, \theta_0), \pi \in \Pi$. The t-statistic process converges weakly to a mean zero Gaussian process $\mathbb{T}(\pi)$ plus the asymptotic bias $\nu(\pi)$. Proof of Theorem 2.2 needs to verify a uniform-entropy condition and apply empirical process theory in van der Vaart and Wellner (1996, Theorem 2.11.22).

---

[7] When we use explicit bounds $\ell_K c_K \lesssim K^{-s/d_x}$ for spline or wavelet series, sufficient condition for Assumption 2.5 is $\sup_{K \in \mathcal{K}_n} \sqrt{n} V_n(K)^{-1/2} K^{-s/d_x} = o(1)$. When we further consider $V_n(K)^{1/2} \propto K^{1/2}$, Assumption 2.3 and 2.5 together imply that Assumption 2.5 is provided by $\sqrt{n} \bar{K}^{1/2 - s/d_x} \to 0$ for power series.

Remark 2.6 (Rate conditions). Theorem 2.2 requires $\sup_\pi |\nu(\pi)| < \infty$ so that t-statistics has either zero asymptotic bias for all $K \in \mathcal{K}_n$ or non-zero bounded asymptotic bias for all $K \in \mathcal{K}_n$. A set $\mathcal{K}_n$ in Assumption 2.3 allows narrower ranges of K than the Assumption 2.1 because it only considers sequences of K with the same rates which only differ in constant π. However, these are the class of sequences to be able to provide uniform central limit theorem of the t-statistic process; as discussed in Remark 2.3, studentized t-statistic can be asymptotically independent with different rates of K. Moreover, weak convergence results can be used for tabulating critical value (see Remark 4.1 for more details).

3 Test statistic

In this section, I introduce test statistics and analyze its asymptotic null distribution based on Theorem 2.1. Then, I provide an asymptotic size result of the tests and consider local power analysis (I restrict attention to Assumption 2.1 which considers a more general range of $\mathcal{K}_n$ than Assumption 2.3, and thus power results under Theorem 2.2 are omitted for brevity).

I consider following supremum and infimum test statistics,

$$
\begin{aligned}
\text{Sup}\, T_n(\theta) &= \sup_{K \in \mathcal{K}_n} |T_n(K, \theta)|, \\
\text{Inf}\, T_n(\theta) &\equiv \inf_{K \in \mathcal{K}_n} |T_n(K, \theta)|.
\end{aligned}
\tag{3.1}
$$

As I denoted in the introduction, considering $\text{Sup}\, T_n(\theta)$ is appropriate in the context of multiple testing and is known to control the size of the simultaneous tests over $\mathcal{K}_n$. There are also several reasons to consider $\text{Inf}\, T_n(\theta)$ in the series regression context; small t-statistic centered at the true θ_0 corresponds to the approximation that has a small bias and large variance, similar to under-smoothing assumption. Using similar ideas of some rule-of-thumb methods suggested to choose undersmoothed series terms, I provide formal inference methods based on asymptotic distribution results of the infimum test statistic with appropriate critical values smaller than the standard normal critical values.[8] Using infimum t-statistics is also related to certain sensitivity analysis in the parametric setup, such as Levine and Renelt (1992) following Leamer (1983)'s idea, which is so-called Leamer's "Extreme Bounds Analysis."[9]

Asymptotic null limiting distribution of the test statistic follows immediately from Theorem 2.1.

---

[8]For example, among many others, Newey (2013) suggested increasing K until standard errors are large relative to small changes in objects of interest, Newey, Powell, and Vella (1999) suggested using more terms than that cross-validation chooses, and Horowitz and Lee (2012) suggested increasing K until the integrated variance suddenly increases and then adding an additional term.

[9]In the parametric setup, a variable is defined as "robust" when 0 is not included in the lower and upper extremes of $\widehat{\theta}_n(m) \pm 1.96 s(\widehat{\theta}_n(m))$ over different specifications $m = 1, \cdots, M$, otherwise "fragile". This is easily shown to be equivalent to testing $\text{Inf}\, T_n(0) = \inf_m |\frac{\widehat{\theta}_n(m)}{s(\widehat{\theta}_n(m))}| > 1.96$.

Corollary 3.1. *Under Assumptions 2.1, 2.2 and,* $\sup_m |\nu(m)| < \infty$ *, then*

$$\mathrm{Sup}\, T_n(\theta_0) \xrightarrow{d} \xi_{\mathrm{sup}}(\Sigma, \nu) \equiv \sup_{m=1,\cdots,M} |Z_m + \nu(m)|, \tag{3.2}$$

where Z_m *is an element of* $M \times 1$ *normal vector* $Z_\Sigma \sim N(0, \Sigma)$ *and* $\nu = (\nu(1), \cdots, \nu(M))'$ *is defined in Theorem 2.1. Under Assumptions 2.1, 2.2,*

$$\mathrm{Inf}\, T_n(\theta_0) \xrightarrow{d} \xi_{\mathrm{inf}}(\Sigma, \nu) \equiv \inf_{m=1,\cdots,M} |Z_m + \nu(m)|. \tag{3.3}$$

Corollary 3.1 gives the asymptotic distribution of our test statistics, $\mathrm{Sup}\, T_n(\theta)$ and $\mathrm{Inf}\, T_n(\theta)$. In addition, if Assumption 2.5 holds, then the limiting distributions $\xi_{\mathrm{sup}}(\Sigma, 0_M)$ and $\xi_{\mathrm{inf}}(\Sigma, 0_M)$ do not contain bias components. Allowing $\sup_m |\nu(m)| = +\infty$ for some m, $\mathrm{Inf}\, T_n(\theta_0)$ converges in distribution to the bounded random variable, while $\mathrm{Sup}\, T_n(\theta_0)$ converges in probability to infinity. Using $\mathrm{Inf}\, T_n(\theta)$ can reduce size distortions, however, it may have low power; similar to the size and power trade-off using (deterministic) undersmoothing rates compare with slower than or equal to the optimal MSE rates. We investigate this trade-off between size and power of the test in the following sections.

3.1 Asymptotic size

I start by defining critical value $c_{1-\alpha}^{\mathrm{sup}}(\nu), c_{1-\alpha}^{\mathrm{inf}}(\nu)$ for $0 < \alpha < 1/2$,

$$c_{1-\alpha}^{\mathrm{sup}}(\nu) = F_{\xi_{\mathrm{sup}}(\Sigma, \nu)}^{-1}(1 - \alpha), \quad c_{1-\alpha}^{\mathrm{inf}}(\nu) = F_{\xi_{\mathrm{inf}}(\Sigma, \nu)}^{-1}(1 - \alpha) \tag{3.4}$$

where $F_X^{-1}(1 - \alpha)$ denotes the $(1 - \alpha)$ quantile of the random variable X. I also define $z_{1-\alpha/2}$ as $(1 - \alpha/2)$ quantile of the standard normal random variable, which solves $P(|Z| > z_{1-\alpha/2}) = \alpha$, $Z \sim N(0, 1)$.

Next Corollary provides the asymptotic size of the test based on $\mathrm{Sup}\, T_n(\theta)$ with the critical value $c_{1-\alpha}^{\mathrm{sup}}(\nu)$.

Corollary 3.2. *Under Assumptions 2.1, 2.2, and* $\sup_m |\nu(m)| < \infty$, *following holds*

$$\lim_{n\to\infty} P(\mathrm{Sup}\, T_n(\theta_0) > c_{1-\alpha}^{\mathrm{sup}}(\nu)) = \alpha. \tag{3.5}$$

Moreover,

$$\limsup_{n\to\infty} P(T_n(\widehat{K}, \theta_0) > c_{1-\alpha}^{\mathrm{sup}}(\nu)) \leq \alpha \tag{3.6}$$

for any $\widehat{K} \in \mathcal{K}_n$.

Corollary 3.2 shows that the tests based on $\mathrm{Sup}\, T_n(\theta)$ asymptotically control size, and the tests based on $|T_n(\widehat{K}, \theta_0)|$ for any $\widehat{K} \in \mathcal{K}_n$ with adjusting standard normal critical value to the critical

13

value from supremum test statistics will also be valid, but conservative.

Next Corollary provides the asymptotic size of the tests based on $\operatorname{Inf} T_n(\theta)$.

Corollary 3.3. *Under Assumptions 2.1 and 2.2, following holds,*

$$\lim_{n\to\infty} P(\operatorname{Inf} T_n(\theta_0) > c_{1-\alpha}^{\inf}(\nu)) = \alpha. \tag{3.7}$$

Corollary 3.3 shows that the tests based on $\operatorname{Inf} T_n(\theta)$ asymptotically control size without assuming $\sup_m |\nu(m)| < \infty$. Even when we allow "large" asymptotic bias ($|\nu(m)| = \infty$) for several Ks in $\mathcal{K}_n$, size of the test is asymptotically level α.

Remark 3.1. Note that $\operatorname{Inf} T_n(\theta_0) \leq |T_n(K_m, \theta_0)|$ and $|T_n(K_m, \theta_0)| \overset{d}{\to} |N(\nu(m), 1)|$ for any single $K_m \in \mathcal{K}_n$, thus we can also show that the test based on $\operatorname{Inf} T_n(\theta)$ using critical value $\inf_m F_{|N(\nu(m),1)|}^{-1}(1-\alpha)$ controls the asymptotic size, but conservative. Note that $\inf_m F_{|N(\nu(m),1)|}^{-1}(1-\alpha)| = F_{|N(\nu(\hat{m}),1)|}^{-1}(1-\alpha)$ with $\hat{m}$ such that $|\nu(\hat{m})| = \inf_m |\nu(m)|$ and this coincides with normal critical value $z_{1-\alpha/2}$ when $\inf_m |\nu(m)| = 0$, which implies that the test based on $\operatorname{Inf} T_n(\theta_0)$ with normal critical value controls size asymptotically if the smallest asymptotic bias is 0, which is the case when $\mathcal{K}_n$ includes at least one undersmoothing sequence.

3.2   Asymptotic local power properties

In this section, we investigate the local asymptotic power of the tests based on $\operatorname{Sup} T_n(\theta)$ and $\operatorname{Inf} T_n(\theta)$.[10]

In order to establish the power properties of our test, we impose following assumptions.

Assumption 3.1.

   (i) *Let $\theta_n = \theta_0 + \bar{\mu} n^{-\gamma}$ for some $\bar{\mu} \in \mathbb{R}$ and $\gamma > 0$.*

   (ii) $\sqrt{n/V_n(K_m)} \asymp n^{\delta_m}$ *for all $K_m \in \mathcal{K}_n$ defined in Assumption 2.1 such that $1/2 > \delta_1 > \delta_2 > \cdots > \delta_M > 0$.*

The local alternatives are defined in 3.1(i) and high-level assumption 3.1(ii) is used to characterize the asymptotic distribution under local alternatives. Since nonparametric estimator $\hat{\theta}_n(K)$ is converging at a rate $\sqrt{n/V_n(K)}$, we need rate restrictions on series variance $V_n(K)$ similar to Assumption 2.4(ii) to compare rates of $\sqrt{n/V_n(K)}$ and $n^{-\gamma}$, as well as asymptotic bias term.

Under Assumptions 2.1, 2.2 and 3.1, limiting distribution of the test statistics under local alternatives are as follows;

$$S(T_n(\theta_n)) \overset{d}{\longrightarrow} S(Z_\Sigma + \nu + \mu) \tag{3.8}$$

---

[10]Both tests are consistent against all fixed alternatives, so we focus on the local power analysis.

for any continuous function $S(t)$ at all $t \in \mathcal{T} \subset \mathbb{R}^M_{[\pm\infty]}$, where $T_n(\theta) = (T_n(K_1, \theta), \cdots, T_n(K_M, \theta))'$, $Z_\Sigma = (Z_1, \cdots, Z_M)' \sim N(0, \Sigma)$, $\nu = (\nu(1), \cdots, \nu(M))'$ are defined in Theorem 2.1, and $\mu = (\mu(1), \cdots, \mu(M))'$, $\mu(m) \equiv -\lim_{n\to\infty} \sqrt{n} V_n(K_m)^{-1/2} \bar{\mu} n^{-\gamma}$.

The following Corollary shows the consistency of the test based on $\operatorname{Sup} T_n(\theta)$ against all $n^{-\gamma}$-local alternatives such that $\gamma < \delta_1$. The test also has nontrivial power against $n^{-\gamma}$-local alternative with $\gamma = \delta_1$, but not against alternatives that converges faster.

Corollary 3.4. *Suppose Assumptions 2.1, 2.2, 3.1 and* $\sup_m |\nu(m)| < \infty$ *holds. Then, following holds for all* $\gamma < \delta_1$,

$$\lim_{n\to\infty} P(\operatorname{Sup} T_n(\theta_n) > c^{\sup}_{1-\alpha}(\nu)) = 1. \tag{3.9}$$

For $\gamma = \delta_1$,

$$\operatorname{Sup} T_n(\theta_n) \xrightarrow{d} \sup_{m=1,\cdots,M} |Z_m + \nu(m) + \mu(m)|, \tag{3.10}$$

where $Z_\Sigma = (Z_1, \cdots, Z_M)' \sim N(0, \Sigma)$ *and* $\nu = (\nu(1), \cdots, \nu(M))'$ *are defined in Theorem 2.1, and* $\mu = (\mu(1), \cdots, \mu(M))' = (\mu(1), 0'_{M-1})'$.

Local asymptotic power results for the test based on $\operatorname{Inf} T_n(\theta)$ and the critical value $c^{\inf}_{1-\alpha}(\nu)$ are as follows.

Corollary 3.5. *Suppose Assumptions 2.1, 2.2, and 3.1 holds. For some fixed* $p \in \{0, 1, \cdots, M-1\}$, *assume that* $|\nu(m)| = +\infty$ *for all* $m < p+1$ *and* $|\nu(m)| < +\infty$ *otherwise (* $p = 0$ *denotes the case* $\sup |\nu(m)| < +\infty$*). Then, following holds for all* $\gamma < \delta_M$,

$$\lim_{n\to\infty} P(\operatorname{Inf} T_n(\theta_n) > c^{\inf}_{1-\alpha}(\nu)) = 1. \tag{3.11}$$

Moreover, for all $\gamma \in [\delta_M, \delta_{p+1}]$,

$$\lim_{n\to\infty} P(\operatorname{Inf} T_n(\theta_n) > c^{\inf}_{1-\alpha}(\nu)) \geq \alpha \tag{3.12}$$

with strict inequality holds for $\gamma = \{\delta_{p+1}, \cdots, \delta_M\}$.

Corollary 3.5 show the consistency of the test based on infimum statistics against all $n^{-\gamma}$-local alternatives such that $\gamma < \delta_M$ and asymptotic unbiasedness of the test for all $\gamma \in [\delta_M, \delta_{p+1}]$ when the first p elements of $|\nu| = (|\nu(1)|, \cdots, |\nu(M)|)'$ is $+\infty$. The test has nontrivial power against some $n^{-\gamma}$-local alternatives with $\gamma \in \{\delta_{p+1}, \cdots, \delta_M\}$, but has no power against $\gamma > \delta_{p+1}$.

Remark 3.2 (Power comparisons of $\operatorname{Sup} T_n(\theta)$ and $\operatorname{Inf} T_n(\theta)$). When $\sup |\nu(m)| < +\infty$, results from Corollary 3.2 and 3.3 reveal that the test based on $\operatorname{Sup} T_n(\theta)$ and $\operatorname{Inf} T_n(\theta)$ both control size asymptotically. Thus, we can compare the power of two asymptotic level α tests from Corollary 3.4 and Corollary 3.5 with $p = 0$. We conclude that the test based on $\operatorname{Sup} T_n(\theta)$ has better asymptotic

15

power than the test based on $\operatorname{Inf} T_n(\theta)$ as the former test is consistent for all $n^{-\gamma}$-local alternatives $\gamma < \delta_1$, but the latter is consistent only against $\gamma < \delta_M (< \delta_1)$ although its power is no less than the size α for $\gamma \in [\delta_M, \delta_1]$.[11]

Remark 3.3 (Power of $\operatorname{Inf} T_n(\theta)$). Although $\operatorname{Inf} T_n(\theta)$ leads to the test that has lower power property of the test compare with $\operatorname{Sup} T_n(\theta)$, the test controls the asymptotic size even when $\sup |\nu(m)| = +\infty$, in which $\operatorname{Sup} T_n(\theta)$ can suffer from size distortions. Furthermore, the test based on $\operatorname{Inf} T_n(\theta)$ with its asymptotic critical value $c_{1-\alpha}^{\text{inf}}(\nu)$ still has nontrivial power against $\gamma \in \{\delta_{p+1}, \cdots, \delta_M\}$, and power is no less than the size for all $\gamma \in [\delta_M, \delta_{p+1}]$. Moreover, it has better power than the test based on single t-statistic $T_n(K, \theta)$ using normal critical value with K that is faster than or equal to the optimal MSE rates. It can be shown (in the proof of Corollary 3.4 with $\mathcal{K}_n = \{K_m\}$) that the single t-statistic $T_n(K_m, \theta)$ only has nontrivial power against $n^{-\gamma}$-local alternatives with $\gamma = \delta_m$, and has no power against strictly slower. Since $\operatorname{Inf} T_n(\theta)$ has nontrivial power for all $\gamma \in \{\delta_{p+1}, \cdots, \delta_M\}$, it has better power properties over single t-statistic with K_m for all $m = p+1, \cdots M$. Intuitively, suppose that $\operatorname{Inf} T_n(\theta_n) = |T_n(\tilde{K}, \theta_n)|$ for some $\tilde{K}$ under the alternative, then the power of the test based on $\operatorname{Inf} T_n(\theta)$ is no less than the power of the test using $|T_n(\tilde{K}, \theta_n)|$ because $c_{1-\alpha}^{\text{inf}}(\nu) \leq z_{1-\alpha/2}$.

4 Confidence interval

In this section, I introduce confidence intervals for $\theta_0 = g_0(x)$ and provide their coverage properties as well as methods to obtain critical values.

We consider a confidence interval based on inverting a test statistic for $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$. Let $T_{n,\widehat{V}}(K, \theta) = \sqrt{n/\widehat{V}_n(K)}(\widehat{\theta}_n(K) - \theta)$ as a t-statistic replacing $V_n(K)$ with $\widehat{V}_n(K)$, and $s(\widehat{\theta}_n(K)) \equiv \sqrt{\widehat{V}_n(K)/n}$ as a standard error of series estimator $\widehat{\theta}_n(K)$. First, we consider following standard CI using the normal critical value $z_{1-\alpha/2}$,

$$CI^{\texttt{Naive}} \equiv \{\theta : |T_{n,\widehat{V}}(\widehat{K}, \theta)| \leq z_{1-\alpha/2}\} = [\widehat{\theta}_n(\widehat{K}) - z_{1-\alpha/2}s(\widehat{\theta}_n(\widehat{K})), \widehat{\theta}_n(\widehat{K}) + z_{1-\alpha/2}s(\widehat{\theta}_n(\widehat{K}))] \quad (4.1)$$

where $\widehat{K}$ is a possibly data-dependent rule chosen from $\mathcal{K}_n$. Conventional CI using normal critical values in (4.1) comes from the asymptotic normality of the t-statistic under deterministic sequence, i.e., when $\mathcal{K}_n = \{K\}$. However, it is not clear whether $CI^{\texttt{Naive}}$ has correct coverage probability; (1) $T_n(\widehat{K}, \theta_0) \xrightarrow{d} N(0, 1)$ may not hold with a random sequence of $\widehat{K}$, even if we assume the asymptotic bias is negligible; (2) $\widehat{K}$ with some data-dependent rules (e.g., cross-validation) may not satisfy the undersmoothing rate conditions which ensure the asymptotic normality without bias terms.

Next, I consider the following robust CI using the critical value $\widehat{c}_{1-\alpha}^{\text{sup}}$ rather than the normal

---

[11]We may also consider other test statistics, for example, average or median of the t-statistics, and we can provide similar size and power results from (3.8). Investigating some optimal property of the test (e.g., optimal weighted average power of the test as in Andrews and Ploberger (1994)) in this nonparametric regression setup will be of interest, but this is beyond the scope of this paper.

critical value $z_{1-\alpha/2}$ compare to $CI^{\texttt{Naive}}$,

$$CI_{\texttt{sup}}^{\texttt{Robust}} \equiv [\widehat{\theta}_n(\widehat{K}) - \widehat{c}_{1-\alpha}^{\sup} s(\widehat{\theta}_n(\widehat{K})), \widehat{\theta}_n(\widehat{K}) + \widehat{c}_{1-\alpha}^{\sup} s(\widehat{\theta}_n(\widehat{K}))], \quad \widehat{K} \in \mathcal{K}_n. \tag{4.2}$$

where $\widehat{c}_{1-\alpha}^{\sup}$ is the critical value from $1 - \alpha$ quantile of $\xi_{\sup}(\widehat{\Sigma}, 0_M)$ defined in Corollary 3.1 with Assumption 2.5 (see Remark 4.2 for general cases without Assumption 2.5). We can obtain critical values by using Monte Carlo simulation based method which will be discussed later in this section.

Finally, I define $CI_{\texttt{inf}}^{\texttt{Robust}}$ as the nominal level $1 - \alpha$ CI for θ based on infimum test statistics,

$$\begin{aligned} CI_{\texttt{inf}}^{\texttt{Robust}} &\equiv \{\theta : \inf_{K \in \mathcal{K}_n} |T_{n,\widehat{V}}(K, \theta)| \leq \widehat{c}_{1-\alpha}^{\inf}\} = \bigcup_{K \in \mathcal{K}_n} \{\theta : |T_{n,\widehat{V}}(K, \theta)| \leq \widehat{c}_{1-\alpha}^{\inf}\} \\ &= [\inf_K (\widehat{\theta}_n(K) - \widehat{c}_{1-\alpha}^{\inf} s(\widehat{\theta}_n(K))), \sup_K (\widehat{\theta}_n(K) + \widehat{c}_{1-\alpha}^{\inf} s(\widehat{\theta}_n(K)))] \end{aligned} \tag{4.3}$$

where $\widehat{c}_{1-\alpha}^{\inf}$ is the critical value from $1 - \alpha$ quantile of $\xi_{\inf}(\widehat{\Sigma}, 0_M)$. Note that $CI_{\texttt{inf}}^{\texttt{Robust}}$ can be easily obtained by using estimates $\widehat{\theta}_n(K)$, standard errors $s(\widehat{\theta}_n(K))$, and a critical value $\widehat{c}_{1-\alpha}^{\inf}$; it is the lower and the upper-end point of confidence intervals for all $K \in \mathcal{K}_n$ using $\widehat{c}_{1-\alpha}^{\inf}$.

Next, I discuss detail implementations to obtain critical values using simple simulation methods. This requires estimators of the variance $V_n(K)$ that are consistent uniformly over $K \in \mathcal{K}_n$. Define least square residuals as $\widehat{\varepsilon}_{Ki} = y_i - P'_{Ki}\widehat{\beta}_K$, and let $\widehat{V}_n(K)$ as the simple plug-in estimator for $V_n(K)$

$$\begin{aligned} \widehat{V}_n(K) &= P_K(x)' \widehat{Q}_K^{-1} \widehat{\Omega}_K \widehat{Q}_K^{-1} P_K(x), \\ \widehat{Q}_K &= \frac{1}{n} \sum_{i=1}^{n} P_{Ki} P'_{Ki}, \quad \widehat{\Omega}_K = \frac{1}{n} \sum_{i=1}^{n} P_{Ki} P'_{Ki} \widehat{\varepsilon}_{Ki}^2. \end{aligned} \tag{4.4}$$

We also define $\widehat{V}_n(K_j, K_l)$ as a sample analog estimator of $V_n(K_j, K_l) \equiv P_{K_j}(x)' Q_{K_j}^{-1} \Omega_{K_j, K_l} Q_{K_l}^{-1} P_{K_l}(x)$, $\Omega_{K_j, K_l} = E(P_{K_j i} P'_{K_l i} \varepsilon_i^2)$,

$$\begin{aligned} \widehat{V}_n(K_j, K_l) &= P_{K_j}(x)' \widehat{Q}_{K_j}^{-1} \widehat{\Omega}_{K_j, K_l} \widehat{Q}_{K_l}^{-1} P_{K_l}(x), \\ \widehat{\Omega}_{K_j, K_l} &= \frac{1}{n} \sum_{i=1}^{n} P_{K_j i} P'_{K_l i} \widehat{\varepsilon}_{K_j i} \widehat{\varepsilon}_{K_l i}. \end{aligned}$$

Then, I define $\widehat{c}_{1-\alpha}^{\sup}$ based on the asymptotic null distribution of $\sup T_n(\theta_0)$ as follows,

$$\begin{aligned} \widehat{c}_{1-\alpha}^{\sup} &\equiv (1 - \alpha) \text{ quantile of } \xi_{\sup}(\widehat{\Sigma}, 0_M) = \sup_{m=1,\cdots,M} |Z_{m,\widehat{\Sigma}}|, \\ &\text{where } Z_{\widehat{\Sigma}} = (Z_{1,\widehat{\Sigma}}, \cdots, Z_{M,\widehat{\Sigma}})' \sim N(0, \widehat{\Sigma}), \\ \widehat{\Sigma}(j,j) &= 1, \quad \widehat{\Sigma}(j,l) = \frac{\widehat{V}_n(K_j, K_l)}{\widehat{V}_n(K_j)^{1/2} \widehat{V}_n(K_l)^{1/2}} \end{aligned} \tag{4.5}$$

where $\widehat{\Sigma}$ is a consistent estimator of variance-covariance matrix Σ defined in Theorem 2.1. To make implementation procedures even simpler, we can construct variance-covariance matrix using

17

conditional homoskedasticity assumption by replacing $\widehat{\Sigma}(j,l) = \widehat{V}_n(K_j)^{1/2}/\widehat{V}_n(K_l)^{1/2}$; this only requires standard error $\widehat{V}_n(K)^{1/2}$ for each $K \in \mathcal{K}_n$. One can compute $\widehat{c}_{1-\alpha}^{\sup}$ by simulating B (typically $B = 1000$ or 5000) i.i.d. random vectors $Z_{\widehat{\Sigma}}^b \sim N(0, \widehat{\Sigma})$ and by taking $(1-\alpha)$ sample quantile of $\{\sup_m |Z_{m,\widehat{\Sigma}}^b| : b = 1, \cdots, B\}$.

Similarly, I define $\widehat{c}_{1-\alpha}^{\inf}$ as follows

$$\widehat{c}_{1-\alpha}^{\inf} \equiv (1-\alpha) \text{ quantile of } \inf_{m=1,\cdots,M} |Z_{m,\widehat{\Sigma}}|, \tag{4.6}$$

where $Z_{\widehat{\Sigma}} = (Z_{1,\widehat{\Sigma}}, \cdots, Z_{M,\widehat{\Sigma}})' \sim N(0, \widehat{\Sigma})$ and $\widehat{\Sigma}$ are defined in (4.5).

Remark 4.1 (Tabulating critical values). Alternatively, we can use the weak convergence of empirical process in Theorem 2.2 to tabulate critical values. Under the same assumptions as in Theorem 2.2,

$$\sup_{K \in \mathcal{K}_n} |T_n(\lfloor K \rfloor, \theta)| = \sup_{\pi \in \Pi} |T_n^*(\pi, \theta)| \xrightarrow{d} \sup_{\pi \in \Pi} |\mathbb{T}(\pi)| \tag{4.7}$$

where $\mathbb{T}(\pi)$ is the mean zero Gaussian process defined in Theorem 2.2. The asymptotic null distribution can be completely defined by covariance kernel of the limiting Gaussian process $\mathbb{T}(\pi)$. In general, the limiting Gaussian process cannot be written as some transformation of Brownian motion, so that the asymptotic critical value cannot be tabulated. However, in the special case where Assumption 2.4(ii) holds with $\eta = 1/2$ under homoskedasticity (as discussed below the equation (2.8)), $\sup_{\pi \in [\underline{\pi}, 1]} |\mathbb{T}(\pi)|$ can be approximated by $\sup_{\pi \in [\underline{\pi}, 1]} |\mathbb{Z}(\pi)/\sqrt{\pi}|$ with a Brownian motion process $\mathbb{Z}(\pi)$. Then the critical value can be tabulated easily as it is only a function of $\underline{\pi} = \underline{K}/\bar{K}$ with the smallest $\underline{K}$ and the largest $\bar{K}$, which can be viewed as an analogous result in kernel estimation literature (See Section 2 of Armstrong and Kolesár (2015) with the uniform Kernel and other references therein).[12]

I impose following assumption to provide consistency of the variance-covariance matrix $\widehat{\Sigma}$ and the validity of critical values $\widehat{c}_{1-\alpha}^{\sup}$, $\widehat{c}_{1-\alpha}^{\inf}$.

Assumption 4.1. $\displaystyle\sup_{(K_j, K_l) \in \mathcal{K}_n \times \mathcal{K}_n} |\frac{\widehat{V}_n(K_j, K_l)}{V_n(K_j, K_l)} - 1| = o_p(1)$ as $n, K \to \infty$.

Assumption 4.1 not only impose the consistency of variance estimator $\widehat{V}_n(K)$ uniformly in $K \in \mathcal{K}_n$, but also impose consistent covariance term $\widehat{V}_n(K_j, K_l)$ uniformly in $\mathcal{K}_n \times \mathcal{K}_n$. Assumption 4.1 is satisfied under the regularity conditions (Assumption 2.2) with an additional assumption. For example, if we further assume $\displaystyle\sup_{(K_j, K_l) \in \mathcal{K}_n \times \mathcal{K}_n} ||\sum_{i=1}^n \tilde{P}_{K_j i} \tilde{P}_{K_l i}' \varepsilon_i^2 - E[\tilde{P}_{K_j i} \tilde{P}_{K_l i}' \varepsilon_i^2]|| = o_p(1)$ with

---

[12]Using Table 1 from Armstrong and Kolesár (2015) with a uniform kernel, we can use $\widehat{c}_{1-\alpha}^{\sup} = 1.96, 2.24, 2.48, 2.60,$ 2.80 when $\bar{K}/\underline{K} = 1, 1.2, 2, 3, 10$, respectively. Results in Armstrong and Kolesár (2015) not only depend on the ratio of the maximum and minimum bandwidths but also depends on the kernel to be used. Unlike Kernel estimation, basis functions play no specific roles in our theory as we only impose assumption on rates of $V_n(K)$.

an orthonormalized vector of basis functions $\tilde{P}_K(x) \equiv Q_K^{-1/2} P_K(x)$, then Assumption 4.1 holds. See Lemma 5.1 of Belloni et al. (2015), and also Lemma 3.1 and 3.2 of Chen and Christensen (2015b) for different sufficient conditions under mild rate restrictions and unconditional moment of the error terms.

Following Corollary provides valid coverage property of the CIs considered, and it immediately follows from Corollary 3.2 and 3.3.

Corollary 4.1. *Under the Assumptions 2.1, 2.2, 2.5, and 4.1, $\widehat{c}_{1-\alpha}^{\sup} \overset{p}{\longrightarrow} c_{1-\alpha}^{\sup}(0_M)$ and $\widehat{c}_{1-\alpha}^{\inf} \overset{p}{\longrightarrow} c_{1-\alpha}^{\inf}(0_M)$ holds where $\widehat{c}_{1-\alpha}^{\sup}, \widehat{c}_{1-\alpha}^{\inf}$ are defined in (4.5), (4.6) and $c_{1-\alpha}^{\sup}(\nu), c_{1-\alpha}^{\inf}(\nu)$ are defined in (3.4). Moreover, following coverage properties holds*

$$\lim_{n \to \infty} P(\theta_0 \in [\widehat{\theta}_n(K) \pm \widehat{c}_{1-\alpha}^{\sup} s(\widehat{\theta}_n(K))] \quad \forall K \in \mathcal{K}_n) = 1 - \alpha, \tag{4.8}$$

$$\liminf_{n \to \infty} P(\theta_0 \in CI_{sup}^{Robust}) = \liminf_{n \to \infty} P(\theta_0 \in [\widehat{\theta}_n(\widehat{K}) \pm \widehat{c}_{1-\alpha}^{\sup} s(\widehat{\theta}_n(\widehat{K}))]) \geq 1 - \alpha \tag{4.9}$$

for $\widehat{K} \in \mathcal{K}_n$. Furthermore,

$$\lim_{n \to \infty} P(\theta_0 \in CI_{inf}^{Robust}) = 1 - \alpha. \tag{4.10}$$

By using an appropriate critical value from the distribution of $\mathrm{Sup}\, T_n(\theta)$, (4.8) gives asymptotic coverage of the uniform confidence intervals over $K \in \mathcal{K}_n$ for the true regression function θ_0. Corollary also shows that the $CI_{\sup}^{\mathrm{Robust}}$ guarantees the asymptotic coverage as $1 - \alpha$ for $\widehat{K} \in \mathcal{K}_n$, as well as the validity of $CI_{\inf}^{\mathrm{Robust}}$.

Remark 4.2 (Undersmoothing assumption). Note that the Corollary 4.1 requires undersmoothing assumption. Without Assumption 2.5, coverage in (4.8) can be understood as the uniform confidence intervals for the pseudo-true value $\theta(K) = P_K(x)' \beta_K$, i.e.,

$$\lim_{n \to \infty} P(\theta(K) \in [\widehat{\theta}_n(K) \pm \widehat{c}_{1-\alpha}^{\sup} s(\widehat{\theta}_n(K))] \quad \forall K \in \mathcal{K}_n) = 1 - \alpha. \tag{4.11}$$

Furthermore, imposing undersmoothing assumption does not undermine coverage results established here because an asymptotic distribution of supremum test statistic is only valid under $\sup_m |\nu(m)| < +\infty$ in Corollary 3.1. Only case where Assumption 2.5 fails but still $\sup_m |\nu(m)| < \infty$ is that when $\underline{K}$ satisfies MSE rates, and other $K \in \mathcal{K}_n$ satisfy undersmoothing, i.e., $\nu = (\nu(1), 0'_{M-1})'$. Although asymptotic bias terms $\nu(1)$ cannot be consistently estimated, we can use following conservative critical values,

$$\sup_{|\nu(1)| \in [0, \bar{\nu}]} F_{\xi_{\sup}(\hat{\Sigma}, \nu)}^{-1}(1 - \alpha)$$

where $\nu = (\nu(1), 0'_{M-1})'$ with some upper bound of the asymptotic bias $\bar{\nu}$. Similarly, under the same assumptions as in Corollary 3.5 and allowing $\sup_m |\nu(m)| = +\infty$ we have $\mathrm{Inf}\, T_n(\theta_0) \overset{d}{\longrightarrow} \xi_{\inf}(\Sigma, \nu) = \inf_{m=p+1, \cdots, M} |Z_m + \nu(m)|$, thus we can use $\sup_{|\nu(p+1)| \in [0, \bar{\nu}]} F_{\xi_{\inf}(\hat{\Sigma}, \nu)}^{-1}(1 - \alpha)$ where $\nu = (0'_p, \nu(p+1),$

$0'_{M-p-1})'$ (or $F^{-1}_{\xi_{\text{inf}}(\hat{\Sigma},0_M)}(1-\alpha)$ if $|\nu(p+1)| = 0$) for $CI_{\text{inf}}^{\text{Robust}}$ without undersmoothing assumption. The validity of these critical values without Assumption 2.5 (so that optimal MSE rates are allowed) can be shown, but additional efforts do not add neither new insights nor practical takeaway, so it is omitted for brevity.[13]

Remark 4.3 (Length of the interval)**.** Note that the last equality from the definition of $CI_{\text{inf}}^{\text{Robust}}$ in (4.3) holds only when there is no dislocated CI, i.e., the intersection is nonempty at least for some two CIs using $\hat{c}_{1-\alpha}^{\text{inf}}$.[14] Otherwise, using the superset widens the length of CI. Potential large length of the $CI_{\text{inf}}^{\text{Robust}}$ is related to the possible low power property of the test in Section 3.2, and this can be avoidable if $\underline{K}$ is reasonably large. Also note that increasing $\bar{K}$ does not necessarily increase the length of $CI_{\text{inf}}^{\text{Robust}}$ as it can decrease critical values $\hat{c}_{1-\alpha}^{\text{inf}}$, although theory in this paper does not consider the data-dependent $\mathcal{K}_n$.

5 Extension: partially linear model setup

In this section, I provide inference methods for the partially linear model (PLM) setup. For notational simplicity, I use the similar notation defined in nonparametric regression setup. Suppose we observe random samples $\{y_i, w_i, x_i\}_{i=1}^n$, where y_i is scalar response variable, $w_i \in \mathcal{W} \subset \mathbb{R}$ is treatment/policy variable of interest, and $x_i \in \mathcal{X} \subset \mathbb{R}^{d_x}$ is a set of explanatory variables. For simplicity, we shall assume w_i is a scalar. I consider following partially linear model

$$y_i = \theta_0 w_i + g_0(x_i) + \varepsilon_i, \qquad E(\varepsilon_i | w_i, x_i) = 0. \tag{5.1}$$

We are interested in inference on θ_0 after approximating unknown function $g_0(x)$ by series terms/regressors $p(x_i)$ among a set of potential control variables. Specification searches can be done for the number of different approximating terms or the number of covariates in estimating the nonparametric part.

Series estimator $\hat{\theta}_n(K)$ for θ_0 using the first $K = K_n$ terms is obtained by standard LS estimation of y_i on w_i and P_{Ki}, and has the usual "partialling out" formula

$$\hat{\theta}_n(K) = \left(W'M_K W\right)^{-1} W'M_K Y \tag{5.2}$$

where $W = (w_1, \cdots, w_n)'$, $M_K = I_K - P^K(P^{K'}P^K)^{-1}P^{K'}$, $P^K = [P_{K1}, \cdots, P_{Kn}]'$, $Y = (y_1, \cdots, y_n)'$. The asymptotic normality and valid inference for $\hat{\theta}_n(K)$ have been developed in the literature.[15]

---

[13]Similar to the discussion in Remark 3.1, using standard normal critical value for $CI_{\text{inf}}^{\text{Robust}}$ achieve nominal coverage probability $1 - \alpha$. In an earlier version of the paper, allowing $\sup_m |\nu(m)| = +\infty$, we also show that the coverage probability of $CI_{\text{inf}}^{\text{Robust}}$ using critical value $\hat{c}_{1-\alpha}^{\text{inf}}$ is bounded below by $P(|Z_m| < c_{1-\alpha}^{\text{inf}}(0))$, which is the coverage of single CI with the smallest asymptotic bias. For example, this lower bound is 0.87 when $c_{1-\alpha}^{\text{inf}}(0) = 1.5$ for $\alpha = 0.05$.

[14]As the variance of series estimator increases with K, we expect that the union of all confidence intervals using $\hat{c}_{1-\alpha}^{\text{inf}}$ may only be determined by some large Ks, so that there is no dislocated CI. Although dislocated confidence interval may show evidence of significant bias for some specific models, there is no guarantee to exclude those Ks a *priori*.

[15]See also Robinson (1988), Linton (1995) and references therein for the results of the kernel estimators.

Donald and Newey (1994) derived the asymptotic normality of $\widehat{\theta}_n(K)$ under standard rate conditions where $K/n \to 0$. Belloni, Chernozukhov, and Hansen (2014) analyzed asymptotic normality and uniformly valid inference for the post-double-selection estimator even when K is much larger than n. A recent paper by Cattaneo, Jansson, and Newey (2015a) provided a valid approximation theory for $\widehat{\theta}_n(K)$ even when K grows at the same rate of n.

Different approximation theory using faster rate of K $(K/n \to c > 0)$ than the standard rate conditions $(K/n \to 0)$ is particularly useful for our purpose to establish asymptotic distribution of t-statistics over $K \in \mathcal{K}_n$. From the results in Cattaneo, Jansson, and Newey (2015a), we have following decomposition for any K,

$$
\begin{aligned}
\sqrt{n}(\widehat{\theta}_n(K) - \theta_0) &= (\frac{1}{n}W'M_K W)^{-1}\frac{1}{\sqrt{n}}W'M_K Y \\
&= \widehat{\Gamma}_K^{-1}(\frac{1}{\sqrt{n}}\sum_i v_i M_{ii}^K \varepsilon_i + \frac{1}{\sqrt{n}}\sum_{i=1}^n \sum_{j=1,j\neq i}^n v_i M_{ij}^K \varepsilon_j) + o_p(1)
\end{aligned}
\tag{5.3}
$$

where $\widehat{\Gamma}_K = W'M_K W/n$. Under $E[\varepsilon_i^2|w_i, x_i] = \sigma_\varepsilon^2$, $\sqrt{n}(\widehat{\theta}_n(K) - \theta_0)$ is asymptotically normal with variance $V = \sigma_\varepsilon^2 E[v_i^2]^{-1}$ with any deterministic sequence $K \to \infty$ satisfying the standard rate conditions $K/n \to 0$. Unlike the nonparametric object of interest in the fully nonparametric model where variance term increases with K, $\widehat{\theta}_n(K)$ in (5.3) has parametric $(n^{1/2})$ convergence rate and asymptotic variance V coincides with the semiparametric efficiency bound for all sequences under $K/n \to 0$, i.e., all estimators $\widehat{\theta}_n(K)$ with different sequences of K are asymptotically equivalent.[16] However, under the faster rate conditions $K/n \to c$ for $c > 0$, the second term in (5.3) is not negligible and converges to bounded random variables. Cattaneo, Jansson, and Newey (2015a) apply the central limit theorem of degenerate U-statistics for the second term, similar to the many instrument asymptotics analyzed in Chao, Swanson, Hausman, Newey and Woutersen (2012). The limiting normal distribution under $K/n \to c > 0$ has a larger variance than the standard first-order asymptotic variance, and the adjusted variances depend on the number of terms K so that I can provide an asymptotic distribution of the t-statistics with the different sequence of K over $\mathcal{K}_n$.

Following assumption on $\mathcal{K}_n$ is considered similar to Assumption 2.3. I also impose regularity conditions that are used in Cattaneo, Jansson, and Newey (2015a, Assumption PLM) uniformly over $K \in \mathcal{K}_n$. Let $v_i \equiv w_i - g_{w0}(x_i)$ where $g_{w0}(x_i) \equiv E[w_i|x_i]$.

Assumption 5.1. *(Set of finite number of series terms)*

$\mathcal{K}_n = \{\underline{K} \equiv K_1, \cdots, K_m, \cdots, \bar{K} \equiv K_M\}$ *where* $K_m = \pi_m \bar{K}$ *for constant* π_m, $0 < \underline{\pi} = \pi_1 < \pi_2 < \cdots < \pi_M = 1$, *fixed* M, *and* $\bar{K} = \bar{K}(n) \to \infty$ *as* $n \to \infty$.

Assumption 5.2. *(Regularity conditions for Partially Linear Model)*

(i) $\{y_i, w_i, x_i\}$ *are i.i.d random variables satisfying the model* (5.1).

---

[16]This is also related to the well-known results of the two-step semiparametric estimation; asymptotic variance of two-step semiparametric estimators does not depend on the type of the first-step estimator and smoothing parameter sequences under certain conditions (see Newey (1994b)).

(ii) *There exists constant $0 < c \le C < \infty$ such that $E[\varepsilon_i^2|w_i, x_i] \ge c$ and $E[v_i^2|x_i] \ge c$, $E[\varepsilon_i^4|w_i, x_i] \le C$ and $E[v_i^4|x_i] \le C$.*

(iii) *$rank(P_K) = K$ (a.s.) and $M_{ii,K} \ge C$ for $C > 0$ for all $K \in \mathcal{K}_n$.*

(iv) *For all $K \in \mathcal{K}_n$, there exists γ_g, γ_{g_w},*

$$\min_{\eta_g} E[(g_0(x_i) - \eta_g' P_{Ki})^2] = O(K^{-2\gamma_g}), \quad \min_{\eta_{g_w}} E[(g_{w0}(x_i) - \eta_{g_w}' P_{Ki})^2] = O(K^{-2\gamma_{g_w}}).$$

Assumption 5.2 does not require $K/n \to 0$ which is required to get asymptotic normality in the literature (e.g., Donald and Newey (1994)). Similar to the Assumption 2.2(iii) in nonparametric setup, Assumption 5.2(iv) holds for the polynomials and splines basis. For example, 5.2(iv) holds with $\gamma_g = s_g/d_x, \gamma_{g_w} = s_w/d_x$ when $\mathcal{X}$ is compact and unknown functions $g_0(x), g_{w0}(x)$ has s_g, s_w continuous derivates, respectively.

Under Assumptions 5.1, 5.2 and undersmoothing condition ($nK^{-2(\gamma_g + \gamma_{g_w})} \to 0$), we have following asymptotic distributions of t-statistics $T_n(K, \theta)$ over $K \in \mathcal{K}_n$ assuming conditional homoskedasticity;

$$T_n(K, \theta_0) = \sqrt{n}V_n(K)^{-1/2}(\hat{\theta}_n(K) - \theta_0) \xrightarrow{d} N(0, 1),$$
$$V_n(K) = (1 - K/n)^{-1}V, \quad V = \sigma_\varepsilon^2 E[v_i^2]^{-1},$$

where $V_n(K)$ coincides with standard first-order asymptotic variance under $K/n \to 0$. Allowing K/n need not converge to zero requires "correction" term, $(1 - K/n)^{-1}$ taking into account for the remainder terms in (5.3) that are assumed "small" with the classical condition $K/n \to 0$. Note that the adjusted variance $V_n(K)$ is always greater than V when $K/n \nrightarrow 0$ and is an increasing function of K.

Next theorem is the main result of the partially linear model setup, analogous to nonparametric setup. Theorem 5.1 provides joint asymptotic distribution of the t-statistics $T_n(K, \theta_0)$ over $K \in \mathcal{K}_n$. It also provides the asymptotic coverage results of the CIs that are similarly defined as in Section 4.

Theorem 5.1. *Suppose Assumptions 5.1 and 5.2 hold. Also, $n\bar{K}^{-2(\gamma_g + \gamma_{g_w})} \to 0$ as $\bar{K} \to \infty$. Assume $\bar{K}/n \to c$ ($0 < c < 1$) and $E[\varepsilon_i^2|w_i, x_i] = \sigma_\varepsilon^2, E[v_i^2|x_i] = E[v_i^2]$. Then the joint null limiting distribution is given by*

$$(T_n(K_1, \theta_0), \cdots, T_n(K_M, \theta_0))' \xrightarrow{d} Z_\Sigma = (Z_1, \cdots Z_M)' \sim N(0, \Sigma)$$

with variance-covariance matrix Σ where $\Sigma(j, l) \equiv \lim_{n \to \infty} V_n(K_{j \wedge l})^{1/2}/V_n(K_{j \vee l})^{1/2}$ for $j \ne l$, and

$\Sigma(j, l) = 1$ *for* $j = l$. *In addition, if* $\sup\limits_{K \in \mathcal{K}_n} |\frac{\widehat{V}_n(K)}{V_n(K)} - 1| = o_p(1)$ *as* $n, K \to \infty$, *then*

$$\lim_{n \to \infty} P(\theta_0 \in [\widehat{\theta}_n(K) \pm \widehat{c}_{1-\alpha}^{\text{sup}} s(\widehat{\theta}_n(K))] \quad \forall K \in \mathcal{K}_n) = 1 - \alpha, \tag{5.4}$$

$$\liminf_{n \to \infty} P(\theta_0 \in CI_{sup}^{Robust}) \geq 1 - \alpha \quad \text{for } \widehat{K} \in \mathcal{K}_n, \tag{5.5}$$

$$\lim_{n \to \infty} P(\theta_0 \in CI_{inf}^{Robust}) = 1 - \alpha \tag{5.6}$$

where CI_{inf}^{Robust} *and* CI_{sup}^{Robust} *are similarly defined as in Section 4 with PLM estimator* $\widehat{\theta}_n(K)$ *and variance estimator* $\widehat{V}_n(K)$, *and the critical values* $\widehat{c}_{1-\alpha}^{\text{inf}}, \widehat{c}_{1-\alpha}^{\text{sup}}$.

Theorem 5.1 derives the joint asymptotic distribution of the $T_n(K, \theta_0)$ over $K \in \mathcal{K}_n$ for the parametric part in the partially linear model. Note that the variance-covariance matrix Σ is same as in nonparametric model setup under homoskedasticity, and it can be reduced under the condition $\bar{K}/n \to c$,

$$\Sigma(j, l) = \lim_{n \to \infty} \frac{V_n(K_{j \wedge l})^{1/2}}{V_n(K_{j \vee l})^{1/2}} = \lim_{n \to \infty} \frac{(1 - \pi_{j \wedge l} \bar{K}/n)^{-1/2}}{(1 - \pi_{j \vee l} \bar{K}/n)^{-1/2}} = (\frac{1 - c\pi_{j \vee l}}{1 - c\pi_{j \wedge l}})^{1/2}$$

for any $j \neq l$.

Remark 5.1. Note that construction of CIs also requires consistent variance estimators $\widehat{V}_n(K)$,

$$\widehat{V}_n(K) = s^2 \widehat{\Gamma}_K^{-1}, \quad s^2 = \frac{1}{n - 1 - K}(Y - W\widehat{\theta}_K)' M_K (Y - W\widehat{\theta}_K).$$

For consistent variance estimation results under $K/n \to c > 0$ and more discussions, see Theorem 2 of Cattaneo, Jansson, and Newey (2015a) and also Cattaneo, Jansson, and Newey (2015b) even under conditional heteroskedastic error terms.

6 Simulations

This section investigates the small sample performance of the proposed inference methods. I report empirical coverage, and the length of CIs considered in Section 4 with various simulation setups. I also report the power comparison of the tests considered in Section 3.

I consider the following data generating process

$$y_i = g(x_i) + \varepsilon_i,$$

$$x_i = \Phi(x_i^*), \quad \begin{pmatrix} x_i^* \\ \varepsilon_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & \sigma^2 \end{pmatrix}\right) \tag{6.1}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function need to ensure compact support. I investigate following four functions for $g(x)$: $g_1(x) = 4x - 1$, $g_2(x) = \ln(|6x - 3| + 1)sgn(x - 1/2)$, $g_3(x) = \frac{\sin(7\pi x/2)}{1 + 2x^2(sgn(x) + 1)}$, $g_4(x) = x - 1/2 + 5\phi(10(x - 1/2))$, where $\phi(\cdot)$ is the standard normal

probability density function, and $sgn(\cdot)$ is the sign function. $g_1(x)$ and $g_2(x)$ are used in Newey and Powell (2003), Chen and Christensen (2015a) and we label these functions as "linear" and "nonlinear" designs. $g_3(x)$ and $g_4(x)$ are rescaled versions used in Hall and Horowitz (2013), and we denote these as "highly nonlinear" designs. Also, I set $\sigma^2 = 1$ for all simulations results below. Results for $\sigma^2 = 0.5, 0.1$ show similar patterns, thus omitted.

I generate 5000 simulation replications for each different design with sample size $n = 200$. Then, I implement nonparametric series estimators using power series bases and quadratic splines with evenly placed knots. In either case, K denotes the total number of estimated coefficients. I set $\mathcal{K}_n = \{3, 4, 5, 6\}$ for the polynomials and $\mathcal{K}_n = \{6, 7, \cdots, 11, 12\}$ for the splines by setting $\underline{K} = n^{1/5}$ and $\bar{K} = n^{1/3}$ (or $\underline{K} = 2n^{1/5}$ and $\bar{K} = 2n^{1/3}$ for splines) rounded up to the nearest integer. Then, I calculate a pointwise coverage of various CIs for all 40 grid points of x on the support $\mathcal{X} = [0, 1]$. To calculate critical values, 1000 additional Monte Carlo replications are also performed on each simulation iteration. Results for different sample sizes $n = 100, 500, 1000$ and results for the cubic spline regressions show similar patterns, thus omitted for brevity.

As a benchmark, I consider $CI^{\texttt{Naive}}$ in (4.1), standard CI with $\widehat{K}_{\texttt{cv}} \in \mathcal{K}_n$ selected to minimize leave-one-out cross-validation. Then, I consider coverage of proposed CIs in this paper; (1) $CI^{\texttt{Robust}}_{\texttt{sup,cv}}$ defined in (4.2) using $\widehat{K}_{\texttt{cv}}$; (2) $CI^{\texttt{Robust}}_{\texttt{sup,cv+}}$ using $\widehat{K}_{\texttt{cv}} + 2$; (3) $CI^{\texttt{Robust}}_{\texttt{inf}}$ based on $\text{Inf}\, T_n(\theta)$ defined in (4.3). Finally, I examine $CI_{\texttt{maxK}} = [\widehat{\theta}_n(\bar{K}) - z_{1-\alpha/2}s(\widehat{\theta}_n(\bar{K})), \widehat{\theta}_n(\bar{K}) + z_{1-\alpha/2}s(\widehat{\theta}_n(\bar{K}))]$ using the largest number of series terms $\bar{K}$.[17] The critical values, $\widehat{c}^{\texttt{inf}}_{1-\alpha}$ and $\widehat{c}^{\texttt{sup}}_{1-\alpha}$ are constructed using the Monte Carlo methods.

Figure 1 reports nominal 95% coverage probability and the length of CIs for "linear" and "nonlinear" designs ($g_1(x)$ and $g_2(x)$) using polynomials. Figure 2 displays results for "highly nonlinear" designs ($g_3(x)$ and $g_4(x)$) using quadratic splines with a different number of knots. Overall, we find (1) coverage of CIs based on the $\widehat{K}$ (e.g., selected by cross-validation or using more terms than the cross-validation) using larger critical value $\widehat{c}^{\texttt{sup}}_{1-\alpha}$ ($CI^{\texttt{Robust}}_{\texttt{sup,cv}}$ and $CI^{\texttt{Robust}}_{\texttt{sup,cv+}}$) is close to or above 95%; (2) coverage of CI based on the infimum t-statistic ($CI^{\texttt{Robust}}_{\texttt{inf}}$) is close to the nominal 95% level and performs well across the different simulation designs. In terms of length, $CI^{\texttt{Robust}}_{\texttt{inf}}$ and $CI^{\texttt{Robust}}_{\texttt{sup,cv}}$ are quite similar and are narrower than $CI_{\texttt{maxK}}$ (standard CI with undersmoothed K) over the support. $CI^{\texttt{Naive}}$ (using optimal MSE rate) has shortest length, but its coverage is far less than 95% in many cases. Also note that $CI^{\texttt{Robust}}_{\texttt{sup,cv+}}$ is robust to specification search as well as bias because it uses undersmoothed K, and thus length of $CI^{\texttt{Robust}}_{\texttt{sup,cv+}}$ seems wide. However, coverage of $CI^{\texttt{Robust}}_{\texttt{sup,cv+}}$ is no less than 95% at almost all points in different setups.

For the linear function $g_1(x)$, finite sample bias is expected to be small over $K \in \mathcal{K}_n$ as polynomials approximate unknown function well for all K. In this case, as theory predicted in Corollary 4.1, Figure 1 shows that coverage of $CI^{\texttt{Robust}}_{\texttt{inf}}, CI_{\texttt{maxK}}$ are close to the nominal 95% and coverage of $CI^{\texttt{Robust}}_{\texttt{sup,cv}}, CI^{\texttt{Robust}}_{\texttt{sup,cv+}}$ are slightly above 95% for many points of the support. Given the small sample size, coverage of $CI^{\texttt{Robust}}_{\texttt{inf}}$ performs well even at the boundary and length is narrower than $CI_{\texttt{maxK}}$.

---

[17]I also consider coverage of CI using the smallest number of series terms $\underline{K}$, but it is omitted as its coverage is far below 95% at most points in nonlinear designs. In terms of coverage, $CI^{\texttt{Robust}}_{\texttt{inf}}$ using standard normal critical value performs well, but it can be quite conservative, thus also omitted here for brevity.

When there exists bias for some K for the nonlinear function $g_2(x)$, Figure 1 also shows that coverage of all CIs except $CI^{\text{Robust}}_{\text{sup,cv+}}$ is slightly less than 95% at some points. As we can see from Figure 2, using splines works well for highly nonlinear designs and coverage of $CI^{\text{Robust}}_{\text{sup,cv}}$, $CI^{\text{Robust}}_{\text{sup,cv+}}$, and $CI^{\text{Robust}}_{\text{inf}}$ are also close to or above 95% at most points except boundary.[18]

In addition to the length of CIs, I provide the power of the different test statistics considered in Section 3. Figure 3 reports power functions of the five different tests for $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_0 + \delta$ where $\theta_0 = g_2(x)$. As the power depends on the different point of interest x, I consider two cases where (1) bias of series estimator for $g_2(x)$ is small ($x = 0.5$); (2) and bias is relatively large ($x = 0.4$). I plot the following rejection probability as a function of δ: (1) $P(\text{Sup}\, T_n(\theta_0 + \delta) > \hat{c}^{\text{sup}}_{1-\alpha})$; (2) $P(\text{Inf}\, T_n(\theta_0 + \delta) > \hat{c}^{\text{inf}}_{1-\alpha})$; (3) $P(|T_n(\hat{K}_{\text{cv}}, \theta_0 + \delta)| > \hat{c}^{\text{sup}}_{1-\alpha})$ with $\hat{K}_{\text{cv}}$ selected by cross-validation; (4) $P(|T_n(\hat{K}_{\text{cv}} + 2, \theta_0 + \delta)| > \hat{c}^{\text{sup}}_{1-\alpha})$; (5) $P(|T_n(\bar{K}, \theta_0 + \delta)| > z_{1-\alpha/2})$. Figure 3 shows the size and power trade-off of $\text{Sup}\, T_n(\theta)$ and $\text{Inf}\, T_n(\theta)$; the tests based on $\text{Sup}\, T_n(\theta)$ seems to have better power, but the size distortions can be substantial when there are large bias for some K (Figure 3-(b)) while the test based on $\text{Inf}\, T_n(\theta)$ control size or bound the size in such cases. $\text{Inf}\, T_n(\theta)$ has similar power to the test with $T_n(\hat{K}_{\text{cv}})$ and $\hat{c}^{\text{sup}}_{1-\alpha}$, and it has slightly better power than the standard test based on $T_n(\bar{K})$ in Figure 3-(a), but does not uniformly dominate as in Figure 3-(b).

7   Empirical application

In this section, I illustrate our inference procedures by revisiting a paper by Blomquist and Newey (2002). Understanding how tax policy affects individual labor supply has been central issues in labor economics (see Hausman (1985) and Blundell and MaCurdy (1999), among many others). Blomquist and Newey (2002) estimate conditional mean of hours of work given the individual nonlinear budget sets using nonparametric series estimation. They also estimate wage elasticity of the expected labor supply and find some evidence of possible misspecification of the usual parametric model such as maximum likelihood estimation (MLE).

Specifically, Blomquist and Newey (2002) consider the following model by exploiting additive structure from the utility maximization with piecewise linear budget sets. I use the similar notations on their paper,

$$h_i = g(x_i) + \varepsilon_i, \quad E(\varepsilon_i | x_i) = 0, \tag{7.1}$$

$$g(x_i) = g_1(y_J, w_J) + \sum_{j=1}^{J-1} [g_2(y_j, w_j, \ell_j) - g_2(y_{j+1}, w_{j+1}, \ell_j)], \tag{7.2}$$

where h_i is the hours of ith individual worked and $x_i = (y_1, \cdots, y_J, w_1, \cdots, w_J, \ell_1, \cdots, \ell_J, J)$ is the budget set that can be represented by the intercept y_j (non-labor income), slope w_j (marginal wage rates) and the end point ℓ_j of the jth segment in a piecewise linear budget with J segments. Here,

---

[18]Possible poor coverage property of standard kernel based CIs for $g_4(x)$ at single peak ($x = 0.5$) was described in Hall and Horowitz (2013, Figure 3).

equation (7.2) for the conditional mean function follows from Theorem 2.1 of Blomquist and Newey (2002), and this additive structure substantially reduces the dimensionality issues. They consider following power series for $g(x)$,

$$p_k(x) = (y_J^{p_1(k)} w_J^{q_1(k)}, \sum_{j=1}^{J-1} \ell_j^{m(k)}(y_j^{p_2(k)} w_j^{q_2(k)} - y_{j+1}^{p_2(k)} w_{j+1}^{q_2(k)})), \quad p_2(k) + q_2(k) \geq 1. \tag{7.3}$$

From the Swedish "Level of Living" survey in 1973, 1980 and 1990, they pool the data from three waves and use the data for married or cohabiting men of ages 20-60. Changes in tax system over three different time periods gives a large variation in the budget sets. Sample size is $n = 2321$. See Section 5 of Blomquist and Newey (2002) for more detail descriptions. They estimate wage elasticity of the expected labor supply

$$E_w = \bar{w}/\bar{h}[\frac{\partial g(w, \cdots, w, \bar{y}, \cdots, \bar{y})}{\partial w}]|_{w=\bar{w}}, \tag{7.4}$$

which is the regression derivative of $g(x)$ evaluated at the mean of the net wage rates $\bar{w}$, virtual income $\bar{y}$ and level of hours $\bar{h}$.

Table 1 is the same table as in Blomquist and Newey (2002, Table 1). They report estimates $\widehat{E}_w$ and standard errors $SE_{\widehat{E}_w}$ with a different number of series terms by adding additional series terms. For example, estimates in the second raw use the term in the first row $(1, y_J, w_J)$ with the additional terms $(\Delta y, \Delta w)$. Here, $\ell^m \Delta y^p w^q$ denotes approximating term $\sum_j \ell_j^m(y_j^p w_j^q - y_{j+1}^p w_{j+1}^q)$. Blomquist and Newey (2002) also report cross-validation criteria, CV, for each specification. In their formula, series terms are chosen to maximize CV, which minimizes asymptotic MSE. In addition to their original table, I add the standard 95% CI for each specification, i.e., $\widehat{E}_w \pm 1.96 SE_{\widehat{E}_w}$. From the table, it is ambiguous which large model (K) can be used for the inference, and we do not have compelling data-dependent methods to select one of the large K for the confidence interval to be reported. We also want to construct CIs that are robust to specification searches.

To construct proposed CIs, I exploit the covariance structure using the asymptotic distribution of the t-statistics under the homoskedastic error in pointwise inference setup: the variance-covariance matrix is only a function of the variance of series estimators. Therefore, construction of the critical values only requires estimated variances for different specifications that are already reported in the table of Blomquist and Newey (2002), and it is straightforward to construct proposed CIs without replication of the data sets. Based on 10000 simulation repetitions, we have $\widehat{c}_{1-\alpha}^{\text{sup}} = 2.4764, \widehat{c}_{1-\alpha}^{\text{inf}} = 0.9668$, respectively.

In Table 1, I report proposed robust confidence intervals $CI_{\text{sup, cv}}^{\text{Robust}}$ with $\widehat{K}_{\text{cv}}$ as well as $CI_{\text{inf}}^{\text{Robust}}$. Figure 4 also displays pointwise 95% uniform confidence interval $CI_{\text{sup}}^{\text{Robust}}$ for $K_m \in \{K_1, K_2, \cdots, K_{11}\}$ where K_m corresponds to each specification in Table 1 with increasing order of series term, along with the point estimates and standard 95% confidence interval using normal critical value. $CI_{\text{sup,cv}}^{\text{Robust}} = [0.0169, 0.0916]$ widens $CI^{\text{Naive}} = [0.0247, 0.0839]$, the standard CI with $\widehat{K}_{\text{cv}}$, by using $\widehat{c}_{1-\alpha}^{\text{sup}}$. From Figure 4, we can also reject a zero wage elasticity of the labor supply for almost all

models except $\bar{K}$. Moreover, given the table in Blomquist and Newey (2002), $CI_{\text{inf}}^{\text{Robust}} = [0.0271, 0.1111]$ is tighter than the standard CI with the largest number of series terms as well as close to the standard CI with some "large" K. For example, $CI_{\text{inf}}^{\text{Robust}}$ is substantially tighter than $CI_{\text{maxK}} = [0.0148, 0.1280]$ that uses the largest $\bar{K}$ as well as those based on the second largest series terms, $[0.0214, 0.1336]$.

8 Conclusion

This paper considers the inference methods given the specification searches over a range of the different number of series terms in the nonparametric series regression model. I provide methods of constructing uniform CIs by adjusting the conventional normal critical value to the critical value based on the supremum of the t-statistics. Then, I introduce tests based on the infimum of the t-statistics over different series terms and show that the tests control the asymptotic size or bound the size distortions with possibly large bias in our setup.

For this, I develop an asymptotic distribution of the t-statistics over the set of the number of series term, and I also provide an empirical process theory for the t-statistics in series estimation. Asymptotic local power results are also established to show the size-power trade-off of the tests. Pointwise confidence interval for the true regression function is obtained by test statistic inversion, and the critical values can be constructed using a simple Monte Carlo simulation method. I illustrate proposed CIs by revisiting empirical example of Blomquist and Newey (2002). Finally, I provide an extension of the proposed CIs in the partially linear model setup.

There are some potential directions to extend results established here. First, it would be of interest to develop confidence bands for $g_0(x)$ over all $x \in \mathcal{X}$. Recent results of Chernozukhov, Chetverikov, and Kato (2014b) can be used to construct confidence bands which are uniform in $K \in \mathcal{K}_n$. Second, an extension of the current theory for the NPIV setup would be desirable to deal with endogeneity issues. For example, pointwise CIs (or uniform confidence bands) that are uniform in pairs of $(K_n, J_n) \in \mathcal{K}_n \times \mathcal{J}_n$ with an additional dimension of the instrument sieve and the number of instruments $J = J_n$. This is a difficult problem, and it would require a distinct theory to deal with the ill-posed inverse problem as well as two-dimensional choices. Thus we leave it for future research.

References

ANDREWS, D. W. K. (1991a): "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models,"*Econometrica*, 59, 307-345.

ANDREWS, D. W. K. (1991b): "Asymptotic Optimality of Generalized C_L, Cross-Validation, and Generalized Cross-Validation in Regression with Heteroskedastic Errors," *Journal of Econometrics*, 47, 359-377.

ANDREWS, D. W. K. AND P. GUGGENBERGER (2009): "Validity of Subsampling and "Plug-in Asymptotic" Inference for Parameters Defined by Moment Inequalities," *Econometric Theory*, 25, 669-709.

ANDREWS, D. W. K. AND W. PLOBERGER (1994): "Optimal Tests When a Nuisance Parameter Is Present Only Under the Alternative,"*Econometrica*, 62(6), 1383-1414.

ARMSTRONG, T. B. AND M. KOLESÁR (2015): "A Simple Adjustment for Bandwidth Snooping," Working Paper.

BELLONI, A., V. CHERNOZHUKOV, D. CHETVERIKOV, AND K. KATO (2015): "Some New Asymptotic Theory for Least Squares Series: Pointwise and Uniform Results," *Journal of Econometrics*, 186, 345-366.

BELLONI, A., V. CHERNOZHUKOV, AND C. HANSEN (2014): "Inference on Treatment Effects after Selection among High-Dimensional Controls,"*Review of Economic Studies*, 81, 608-650.

BLOMQUIST, S. AND W. K. NEWEY (2002): "Nonparametric Estimation with Nonlinear Budget Sets," *Econometrica*, 70, 2455-2480.

BLUNDELL, R. AND T. E. MACURDY (1999): "Labor Supply: A Review of Alternative Approaches," *Handbook of Labor Economics*, In: O. Ashenfelter, D. Card (Eds.), vol. 3., Elsevier, Chapter 27.

CALONICO, S., M. D. CATTANEO, AND M. H. FARRELL (2015): "On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference," Working paper.

CATTANEO, M. D., M. JANSSON, AND W. K. NEWEY (2015a): "Alternative Asymptotics and the Partially Linear Model with Many Regressors," Working paper.

CATTANEO, M. D., M. JANSSON, AND W. K. NEWEY (2015b): "Treatment Effects With Many Covariates and Heteroskedasticity," Working paper.

CHAO, J. C., N. R. SWANSON, J. A. HAUSMAN, W. K. NEWEY, AND T. WOUTERSEN (2012): "Asymptotic Distribution of JIVE in a Heteroskedastic IV Regression with Many Instruments,"*Econometric Theory*, 28, 42-86.

CHEN, X. (2007): "Large Sample Sieve Estimation of Semi-nonparametric Models," *Handbook of Econometrics*, In: J.J. Heckman, E. Leamer (Eds.), vol. 6B., Elsevier, Chapter 76.

CHEN, X. AND T. CHRISTENSEN (2015a): "Optimal Sup-Norm Rates, Adaptivity and Inference in Nonparametric Instrumental Variables Estimation," Cowles Foundation Discussion Paper 1923.

CHEN, X. AND T. CHRISTENSEN (2015b): "Optimal Uniform Convergence Rates and Asymptotic Normality for Series Estimators Under Weak Dependence and Weak Conditions," *Journal of Econometrics*, 188, 447-465.

CHEN, X. AND Z. LIAO (2014): "Sieve M inference on irregular parameters," *Journal of Econometrics*,182, 70-86.

CHEN, X., Z. LIAO, AND Y. SUN (2014): "Sieve inference on possibly misspecified semi-nonparametric time series models," *Journal of Econometrics*,178, 639-658.

CHEN, X. AND X. SHEN (1998): "Sieve extremum estimates for weakly dependent data," *Econometrica*, 66 (2), 289-314.

CHERNOZHUKOV V, D. CHETVERIKOV, AND K. KATO (2013): "Gaussian Approximations and Multiplier Bootstrap for Maxima of Sums of High-Dimensional Random Vectors," *The Annals of Statistics*, 41(6), 2786-2819.

CHERNOZHUKOV V, D. CHETVERIKOV, AND K. KATO (2014a): "Anti-Concentration and Honest, Adaptive Confidence Bands," *The Annals of Statistics*, 42(5), 1787-1818.

CHERNOZHUKOV V, D. CHETVERIKOV, AND K. KATO (2014b): "Gaussian approximation of suprema of empirical processes," *The Annals of Statistics*, 42(4), 1564-1597.

CONLEY T. G., C. B. HANSEN, AND P.E. ROSSI (2012): "Plausibly Exogenouss," *Review of Economics and Statistics*, 94(1), 260-272.

DONALD, S. G. AND W. K. NEWEY (1994): "Series Estimation of Semilinear Models," *Journal of Multivariate Analysis*, 50, 30-40.

EASTWOOD, B. J. AND A.R. GALLANT, (1991): "Adaptive Rules for Seminonparametric Estimators That Achieve Asymptotic Normality,"*Econometric Theory*, 7, 307-340.

GINÉ, E. AND R. NICKL (2010): "Confidence bands in density estimation," *The Annals of Statistics*, 38, 1122-1170.

GINÉ, E. AND R. NICKL (2015): *Mathematical Foundations of Infinite-Dimensional Statistical Models*, Cambridge University Press .

HALL, P. AND J. HOROWITZ (2013): "A Simple Bootstrap Method for Constructing Nonparametric Confidence Bands for Functions," *The Annals of Statistics*, 41, 1892-1921.

HANSEN B. E. (2015): "The Integrated Mean Squared Error of Series Regression and a Rosenthal Hilbert-Space Inequality," *Econometric Theory*, 31, 337-361.

HANSEN, P.R. (2005): "A Test for Superior Predictive Ability," *Journal of Business and Economic Statistics*, 23, 365-380.

HÄRDLE, W. AND O. LINTON (1994): "Applied Nonparametric Methods," *Handbook of Econometrics*, In: R. F. Engle, D. F. McFadden (Eds.), vol. 4., Elsevier, Chapter 38.

HAUSMAN, J. A. (1985): "The Econometrics of Nonlinear Budget Sets", *Econometrica*, 53, 1255-1282.

HOROWITZ, J. L. (2014): "Adaptive Nonparametric Instrumental Variables Estimation: Empirical Choice of the Regularization Parameter," *Journal of Econometrics*, 180, 158-173.

HOROWITZ, J. L. AND S. LEE (2012): "Uniform Confidence Bands for Functions Estimated Nonparametrically with Instrumental Variables," *Journal of Econometrics*, 168, 175-188.

HUANG, J. Z. (2003a): "Asymptotics for Polynomial Spline Regression Under Weak Conditions," *Statistics & Probability Letters*, 65, 207-216.

HUANG, J. Z. (2003b): "Local Asymptotics for Polynomial Spline Regression," *The Annals of Statistics*, 31, 1600-1635.

LEAMER, E. E. (1983): "Let's Take the Con Out of Econometrics," *The American Economic Review*, 73, 31-43.

LEPSKI, O. V. (1990): "On a problem of adaptive estimation in Gaussian white noise," *Theory of Probability and its Applications*, 35, 454-466.

LEVINE, R.E. AND D. RENELT (1992): "A Sensitivity Analysis of Cross-Country Growth Regressions," *American Economic Review*, 82(4), 942-963.

LI, K. C. (1987): "Asymptotic Optimality for C_p, C_L, Cross-Validation and Generalized Cross-Validation: Discrete Index Set," *The Annals of Statistics*, 15, 958-975.

LI, QI, AND J. S. RACINE (2007): *Nonparametric Econometrics: Theory and Practice*, Princeton University Press.

LINTON, O. (1995): "Second order approximation in the partialy linear regression model," *Econometrica*, 63(5), 1079-1112.

NEWEY, W. K. (1994a): "Series Estimation of Regression Functionals," *Econometric Theory*, 10, 1-28.

NEWEY, W. K. (1994b): "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62, 1349-1382.

NEWEY, W. K. (1997): "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics*, 79, 147-168.

NEWEY, W. K. (2013): "Nonparametric Instrumental Variables Estimation," *American Economic Review: Papers & Proceedings*, 103, 550-556.

NEWEY, W. K. AND J. L. POWELL (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1565-1578.

NEWEY, W. K. AND J. L. POWELL, F. VELLA (1999): "Nonparametric Estimation of Triangular Simultaneous Equations Models," *Econometrica*, 67, 565-603.

ROBINSON, P. M. (1988): "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56(4), 931-954.

ROMANO, J. P. AND M. WOLF (2005): "Stepwise Multiple Testing as Formalized Data Snooping," *Econometrica*, 73, 1237-1282.

ROMANO, J. P. AND A.M. SHAIKH (2006): "Stepup procedures for control of generalizations of the familywise error rate," *Annals of Statistics*, 34, 1850-1873.

SCHENNACH, S. M. (2015): "A bias bound approach to nonparametric inference," *CEMMAP working paper CWP71/15*.

TROPP, J. A. (2015): *An Introduction to Matrix Concentration Inequalities*, Foundations and Trends in Machine Learning, Vol. 8: No.1-2, 1-230.

VAN DER VAART, A. W. AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes*, Springer.

WHITE, H. (2000): "A Reality Check for Data Snooping," *Econometrica*, 68, 1097-1126.

ZHOU, S., X. SHEN, AND D.A. WOLFE (1998): "Local Asymptotics for Regression Splines and Confidence Regions," *The Annals of Statistics*, 26, 1760-1782.

Appendix A   Proofs

A.1   Preliminaries and Useful Lemmas

In the Appendix, we define additional notations for the empirical process theory used in the proof of Theorem 2.2. Given measurable space $(S, \mathcal{S})$, let $\mathcal{F}$ as a class of measurable functions $f : S \to \mathbb{R}$. For any probability measure Q on $(S, \mathcal{S})$, we define $N(\epsilon, \mathcal{F}, L_2(Q))$ as covering numbers, which is the minimal number of the $L_2(Q)$ balls of radius ϵ to cover $\mathcal{F}$ with $L_2(Q)$ norms $||f||_{Q,2} = (\int |f|^2 dQ)^{1/2}$. Uniform entropy numbers relative to the $L_2(Q)$ norms are defined as $\sup_Q \log N(\epsilon ||F||_{Q,2}, \mathcal{F}, L_2(Q))$ where the supremum is over all discrete probability measures with an envelope function F.

Let the data $z_i = (\varepsilon_i, x_i)$ be i.i.d. random vectors defined on probability space $(\mathcal{Z} = \mathcal{E} \times \mathcal{X}, \mathcal{A}, P)$ with common probability distribution $P \equiv P_{\varepsilon,x}$. We think of $(\varepsilon_1, x_1), \cdots (\varepsilon_n, x_n)$ as the coordinates of the infinite product probability space. For notational convenience, we avoid discussing nonmeasurability issues and outer expectations (for the related issues, see van der Vaart and Wellner (1996)). Throughout the proofs, we denote $c, C > 0$ as a universal constant that does not depend on n.

For any sequence $\{K = K_n : n \geq 1\} \in \prod_{n=1}^{\infty} \mathcal{K}_n$, either under Assumption 2.1 or 2.3, we first define the orthonormalized vector of basis functions

$$\tilde{P}_K(x) \equiv Q_K^{-1/2} P_K(x) = E[P_{Ki} P_{Ki}']^{-1/2} P_K(x),$$
$$\tilde{P}_{Ki} = \tilde{P}_K(x_i), \tilde{P}^K = [\tilde{P}_{K1}, \cdots, \tilde{P}_{Kn}]'.$$

We observe that

$$\hat{g}_K(x) = P_K(x)'(P^{K'} P^K)^{-1} P^{K'} Y = \tilde{P}_K(x)'(\tilde{P}^{K'} \tilde{P}^K)^{-1} \tilde{P}^{K'} Y,$$
$$V_n(K) = P_K(x)' Q_K^{-1} \Omega_K Q_K^{-1} P_K(x) = \tilde{P}_K(x)' \tilde{\Omega}_K \tilde{P}_K(x),$$
$$\tilde{\Omega}_K = E(\tilde{P}_{Ki} \tilde{P}_{Ki}' \varepsilon_i^2).$$

Without loss of generality, we may impose normalization of $Q_{\bar{K}} = I_{\bar{K}}$ or $Q_K = E(P_{Ki} P_{Ki}') = I_K$ uniformly over $K \in \mathcal{K}_n$, since $\hat{g}_K(x)$ is invariant to nonsingular linear transformations of $P_K(x)$. However, we shall treat Q_K as unknown and deal with non-orthonormalized series terms here.

Next, we re-define pseudo-true value β_K, with abuse of notation, using orthonormalized series terms $\tilde{P}_{Ki}$. That is, $y_i = \tilde{P}_{Ki}' \beta_K + \varepsilon_{Ki}, E[\tilde{P}_{Ki} \varepsilon_{Ki}] = 0$ where $\varepsilon_{Ki} = r_{Ki} + \varepsilon_i$, $r_n(K) = r_n(K, x) = g_0(x) - \tilde{P}_K(x)' \beta_K, r_{Ki} = r_n(K, x_i)$, and $r_K \equiv (r_{K1}, \cdots r_{Kn})'$. We also define $\hat{Q}_K \equiv \frac{1}{n} \tilde{P}^{K'} \tilde{P}^K$, $\underline{\sigma}^2 \equiv \inf_x E[\varepsilon_i^2 | x_i = x], \bar{\sigma}^2 \equiv \sup_x E[\varepsilon_i^2 | x_i = x]$. We first provide useful lemmas which will be used in the proof of Theorem 2.1 and 2.2. Versions of proof of Lemma 1 are available in the literature, such as Belloni et al. (2015) and Chen and Christensen (2015b), among others. For completeness, we provide the results of Lemma 1. Note that different rate conditions of $K = K_n$ can be used in Assumption 2.2, such as Newey (1997), but lead to different bounds (A.1)-(A.3) in the following Lemma 1.

Lemma 1. *Suppose either Assumption 2.1 or Assumption 2.3 holds. Under the Assumption 2.2, following holds for any $K \in \mathcal{K}_n$,*

$$||\widehat{Q}_K - I_K|| = O_p(\sqrt{\frac{\lambda_K^2 \zeta_K^2 \log K}{n}}), \tag{A.1}$$

$$R_1(K) \equiv \sqrt{\frac{1}{nV_n(K)}} \tilde{P}_K(x)' \left(\widehat{Q}_K^{-1} - I_K\right) \tilde{P}^{K'}(\varepsilon + r_K) = O_p(\sqrt{\frac{\lambda_K^2 \zeta_K^2 \log K}{n}}(1 + \ell_K c_K \sqrt{K})), \tag{A.2}$$

$$R_2(K) \equiv \sqrt{\frac{1}{nV_n(K)}} \tilde{P}_K(x)' \tilde{P}^{K'} r_K = O_p(\ell_K c_K). \tag{A.3}$$

To provide (A.1) in Lemma 1, we first introduce matrix Bernstein inequality in Tropp (2015).

Lemma 2 (Theorem 6.1.1 of Tropp (2015)). *Consider a finite sequence $\{S_i\}$ of independent, random matrices with common dimension $d_1 \times d_2$. Assume that $ES_i = 0$, $||S_i|| \leq L$ for each i. Let $Z = \sum_i S_i$, and define $v(Z) = \max\{||E(ZZ')||, ||E(Z'Z)||\}$. Then,*

$$P(||Z|| \geq t) \leq (d_1 + d_2) \exp(\frac{-t^2/2}{v(Z)Lt/3}), \qquad \forall t \geq 0,$$

$$E||Z|| \leq \sqrt{2v(Z)\log(d_1 + d_2)} + \frac{1}{3}L\log(d_1 + d_2).$$

Proof of Lemma 1.
To provide bound in (A.1), we apply Lemma 2 by setting $S_i = \frac{1}{n}(\tilde{P}_{Ki}\tilde{P}'_{Ki} - E(\tilde{P}_{Ki}\tilde{P}'_{Ki}))$. Note that $\mathbb{E}\,S_i = 0$, $||S_i|| \leq L = \frac{1}{n}(\lambda_K^2 \zeta_K^2 + 1)$, and $v(Z) = \frac{1}{n}||E(\tilde{P}_{Ki}\tilde{P}'_{Ki}\tilde{P}_{Ki}\tilde{P}'_{Ki}) - E(\tilde{P}_{Ki}\tilde{P}'_{Ki})E(\tilde{P}_{Ki}\tilde{P}'_{Ki})|| \leq \frac{1}{n}(\lambda_K^2 \zeta_K^2 + 1)$ by definition of λ_K, ζ_K and $E(\tilde{P}_{Ki}\tilde{P}'_{Ki}) = I_K$. By Lemma 2, we have

$$E||\widehat{Q}_K - I_K|| = E||\sum_i \frac{1}{n}(\tilde{P}_{Ki}\tilde{P}'_{Ki} - I_K)|| \leq C(\sqrt{\lambda_K^2 \zeta_K^2 \log(K)/n} + \lambda_K^2 \zeta_K^2 \log(K)/n).$$

Then we have $||\widehat{Q}_K - I_K|| = O_P(\sqrt{\lambda_K^2 \zeta_K^2 \log(K)/n})$ by Markov inequality. For (A.2), we first look at the terms $\sqrt{\frac{1}{nV_n(K)}} \tilde{P}_K(x)' \left(\widehat{Q}_K^{-1} - I_K\right) \tilde{P}^{K'}\varepsilon$. Conditional on the sample $X = [x_1, \cdots, x_n]$, this term has mean zero and variance,

$$\frac{1}{nV_n(K)} \tilde{P}_K(x)' \left(\widehat{Q}_K^{-1} - I_K\right) \tilde{P}^{K'} E(\varepsilon\varepsilon'|X)\tilde{P}^K \left(\widehat{Q}_K^{-1} - I_K\right) \tilde{P}_K(x)$$

$$\leq \frac{\bar{\sigma}^2}{V_n(K)} \tilde{P}_K(x)' \left(\widehat{Q}_K^{-1} - I_K\right) \widehat{Q}_K \left(\widehat{Q}_K^{-1} - I_K\right) \tilde{P}_K(x)$$

$$= \frac{\bar{\sigma}^2}{V_n(K)} \tilde{P}_K(x)' \left(\widehat{Q}_K - I_K\right) \widehat{Q}_K^{-1} \left(\widehat{Q}_K - I_K\right) \tilde{P}_K(x)$$

$$\leq \frac{\bar{\sigma}^2 \tilde{P}_K(x)' \tilde{P}_K(x)}{V_n(K)} \lambda_{max}(\widehat{Q}_K^{-1})|| \left(\widehat{Q}_K - I_K\right) ||^2 = O_P(\lambda_K^2 \zeta_K^2 \log(K)/n)$$

where the first and the last inequality uses $V_n(K) \leq \bar{\sigma}^2 \tilde{P}_K(x)' \tilde{P}_K(x)$, $V_n(K) \geq \underline{\sigma}^2 \tilde{P}_K(x)' \tilde{P}_K(x)$ by

33

Assumption 2.2(ii), $||\widehat{Q}_K - I_K|| = O_P(\sqrt{\lambda_K^2 \zeta_K^2 \log(K)/n})$ by (A.1) and $\lambda_{max}(\widehat{Q}_K^{-1}) = (\lambda_{max}(\widehat{Q}_K))^{-1} = O_p(1)$ since all eigenvalues of $\widehat{Q}_K$ are bounded away from zero as $|\lambda_{min}(\widehat{Q}_K) - 1| \leq ||\widehat{Q}_K - I_K|| = o_p(1)$ by (A.1) and Assumption 2.2(iv)-(v). Then, by Chebyshev's inequality, we have that

$$\sqrt{\frac{1}{nV_n(K)}} \tilde{P}_K(x)' \left(\widehat{Q}_K^{-1} - I_K \right) \tilde{P}^{K'} \varepsilon = O_P(\sqrt{\lambda_K^2 \zeta_K^2 \log(K)/n}).$$

Next, consider the terms $\sqrt{\frac{1}{nV_n(K)}} \tilde{P}_K(x)' \left(\widehat{Q}_K^{-1} - I_K \right) \tilde{P}^{K'} r_K$. Observe that $||\frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{P}_{Ki} r_{Ki}|| = O_p(\ell_K c_K \sqrt{K})$ since

$$E[||\frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{P}_{Ki} r_{Ki}||^2] = E[\sum_{j=1}^K \tilde{P}_{ji}^2 r_{Ki}^2] \leq \ell_K^2 c_K^2 E[||\tilde{P}_{Ki}||^2] = \ell_K^2 c_K^2 K. \tag{A.4}$$

Combining (A.1) and (A.4) yields the results

$$|\sqrt{\frac{1}{nV_n(K)}} \tilde{P}_K(x)' \left(\widehat{Q}_K^{-1} - I_K \right) \tilde{P}^{K'} r_K| \leq C ||\widehat{Q}_K^{-1}|| \cdot || \left(\widehat{Q}_K - I_K \right) || || \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{P}_{Ki} r_{Ki}||$$

$$= O_p(\sqrt{\frac{\lambda_K^2 \zeta_K^2 \log(K)}{n}} \ell_K c_K \sqrt{K})$$

by $||\frac{\tilde{P}_K(x)}{V_n(K)^{1/2}}|| \asymp 1$ and $||\widehat{Q}_K^{-1}|| = O_p(1)$.

We now prove (A.3). Consider $\sqrt{\frac{1}{nV_n(K)}} \tilde{P}_K(x)' \tilde{P}^{K'} r_K$,

$$E[(\sqrt{\frac{1}{nV_n(K)}} \tilde{P}_K(x)' \tilde{P}^{K'} r_K)^2] = E[(\frac{\tilde{P}_K(x)' \tilde{P}_{Ki}}{V_n(K)^{1/2}} r_{Ki})^2] \leq (c_K \ell_K)^2$$

since $E[(\frac{\tilde{P}_K(x)' \tilde{P}_{Ki}}{V_n(K)^{1/2}})^2] \asymp 1$ by Assumption 2.2(ii) and $E(r_{Ki})^2 \leq (\ell_K c_K)^2$ by Assumption 2.2(iii). Therefore, we have (A.3) by Chebyshev's inequality and using $E[\tilde{P}_{Ki} r_{Ki}] = 0$ from the projection model. This completes the proof. ∎

A.2 Proofs of the Main Results

A.2.1 Proof of Theorem 2.1

Proof. For any $K \in \mathcal{K}_n$, we first show the decomposition of the t-statistic in equation (2.5)

$$
\begin{aligned}
T_n(K, \theta_0) &= \sqrt{\frac{n}{V_n(K)}} \tilde{P}_K(x)'(\hat{\beta}_K - \beta_K) - \sqrt{\frac{n}{V_n(K)}} r_n(K) \\
&= \sqrt{\frac{1}{nV_n(K)}} \tilde{P}_K(x)' \tilde{P}^{K'}(\varepsilon + r_K) \\
&\quad + \sqrt{\frac{1}{nV_n(K)}} \tilde{P}_K(x)' \left(\hat{Q}_K^{-1} - I_K \right) \tilde{P}^{K'}(\varepsilon + r_K) - \sqrt{\frac{n}{V_n(K)}} r_n(K) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\tilde{P}_K(x)' \tilde{P}_{Ki} \varepsilon_i}{V_n(K)^{1/2}} + R_1(K) + R_2(K) - \sqrt{n} V_n(K)^{-1/2} r_n(K)
\end{aligned}
$$

where $R_1(K), R_2(K)$ are defined in (A.2), (A.3). By Lemma 1, we have $R_1(K) = O_p(\sqrt{\frac{\lambda_K^2 \zeta_K^2 \log K}{n}}(1 + \ell_K c_K \sqrt{K})) = o_p(1)$, $R_2(K) = O_p(\ell_K c_K) = o_p(1)$ for any $K \in \mathcal{K}_n$ under Assumptions 2.2, thus we have following decompositions for any $m = 1, 2, \cdots M$,

$$
T_n(K_m, \theta_0) = t_n(m) + \nu_n(m) + o_p(1),
$$

where $t_n(m) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\tilde{P}_{K_m}(x)' \tilde{P}_{K_m i} \varepsilon_i}{V_n(K_m)^{1/2}}$ and $\nu_n(m) = -\sqrt{n} V_n(K_m)^{-1/2} r_n(K_m)$. If some elements of $\nu(m) = -\lim_{n \to \infty} \nu_n(m) = \infty$ under oversmoothing sequences, joint distribution of $(T_n(K_1, \theta_0), \cdots, T_n(K_M, \theta_0))'$ does not converge in distribution to a bounded random vector. To circumvent the issue, remaining proof use the same type of argument as in Theorem 1 of Andrews and Guggenberger (2009) in the moment inequality literature.

To obtain a joint asymptotic distribution of $t_n(m)$, by the Cramér-Wold device, it suffices to show that

$$
\delta' t_n \xrightarrow{d} N(0, \delta' \Sigma \delta), \quad \forall \delta \in \mathbb{R}^M \tag{A.5}
$$

where $t_n = (t_n(1), ..., t_n(M))', \Sigma(j, l) = \lim_{n \to \infty} \Sigma_n(j, l), \Sigma_n(j, l) \equiv \frac{\tilde{P}_{K_j}(x)' E(\tilde{P}_{K_j i} \tilde{P}_{K_l i}' \varepsilon_i^2) \tilde{P}_{K_l}(x)}{V_n(K_j)^{1/2} V_n(K_l)^{1/2}}$. To show (A.5) we will verify Lindberg's condition of the CLT for $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \omega_{ni} \xrightarrow{d} N(0, 1)$, where $\omega_{ni} = (\delta' \Sigma_n \delta)^{-1/2} \sum_{j=1}^{M} \delta_j \frac{\tilde{P}_{K_j}(x)' \tilde{P}_{K_j i} \varepsilon_i}{V_n(K_j)^{1/2}}$. Observe that $E\omega_{ni} = 0$, and $\frac{1}{n} \sum_{i=1}^{n} E[\omega_{ni}^2] = 1$, and

$$
\| \sum_{j=1}^{M} \frac{\tilde{P}_{K_j}(x)' \tilde{P}_{K_j}}{V_n(K_j)^{1/2}} \|_\infty \lesssim \sum_{j=1}^{M} \zeta_{K_j} \lambda_{K_j} \lesssim \zeta_{\bar{K}} \lambda_{\bar{K}}
$$

by Assumptions 2.1 and 2.2. Moreover, $(\delta' \Sigma_n \delta)^{-1} \lesssim 1$. Therefore, for any $a > 0$,

$$\frac{1}{n} \sum_{i=1}^n E(|\omega_{ni}|^2 1\{|\omega_{ni}| > a\sqrt{n}\})$$

$$\lesssim M \sum_{j=1}^M E[|\frac{\tilde{P}_{K_j}(x)' \tilde{P}_{K_j i} \varepsilon_i}{V_n(K_j)^{1/2}}|^2 1\{|\sum_{j=1}^M \delta_j \frac{\tilde{P}_{K_j}(x)' \tilde{P}_{K_j i}}{V_n(K_j)^{1/2}} \varepsilon_i| > a\sqrt{n}\}]$$

$$\leq M \sum_{j=1}^M E(|\frac{\tilde{P}_{K_j}(x)' \tilde{P}_{K_j i}}{V_n(K_j)^{1/2}}|^2) \sup_x E[\varepsilon_i^2 1\{|\varepsilon_i| > a(\sqrt{n}/(\zeta_{\bar{K}} \lambda_{\bar{K}})\}|x_i = x],$$

where the last term goes to 0 under $n \to \infty$ by Assumption 2.2(ii), since $E[(\frac{\tilde{P}_K(x)' \tilde{P}_{Ki}}{V_n(K)^{1/2}})^2] \asymp 1$ and $(\zeta_{\bar{K}} \lambda_{\bar{K}})/\sqrt{n} = o(1)$ by Assumption 2.2(iv). Thus, Lindberg condition is verified and therefore (A.5) holds by Lindberg-Feller CLT and Slutzky's Theorem.

Next, we let $G(\cdot)$ be a strictly increasing continuous distribution function on $\mathbb{R}$, for example, standard normal CDF $\Phi(\cdot)$. For any m,

$$G_{n,m} = G(T_n(K_m, \theta_0)) = G(t_n(m) + \nu_n(m) + o_p(1)).$$

If $|\nu(m)| < \infty$, then we have

$$G_{n,m} \xrightarrow{d} G(Z_m + \nu(m)) \tag{A.6}$$

by finite dimensional CLT and the continuous mapping theorem. If $\nu(m) = +\infty$,

$$G_{n,m} \xrightarrow{p} 1 \tag{A.7}$$

since $t_n(m) = O_p(1)$, and $G(x) \to 1$ as $x \to \infty$, and by CLT. Moreover, if $\nu(m) = -\infty$

$$G_{n,m} \xrightarrow{p} 0 \tag{A.8}$$

as $G(x) \to 0$ as $x \to -\infty$. Since (A.6), (A.7), and (A.8) holds jointly, following holds for any strictly increasing continuous distribution function on $\mathbb{R}$, $G(\cdot)$,

$$G_n \equiv (G_{n,1}, \cdots, G_{n,M})' \xrightarrow{d} G_\infty \equiv (G(Z_1 + \nu(1)), \cdots, G(Z_M + \nu(M)))' \tag{A.9}$$

where $G_{n,m} = G(T_n(K_m, \theta_0))$, and $G(Z_m + \nu(m))$ denotes $G(+\infty) = 1$ when $\nu(m) = +\infty$, and $G(-\infty) = 0$ when $\nu(m) = -\infty$.

Next, we define $G^{-1}(\cdot)$ as the inverse of $G(\cdot)$. For $t = (t_1, \cdots, t_M)' \in \mathcal{T} \subset \mathbb{R}_{[\pm\infty]}^M$, define $G_{(M)}(t) \equiv (G(t_1), \cdots, G(t_M))' \in \mathcal{Y} = \mathcal{Y}(\mathcal{T}) \subset [0, 1]^M$. For $y = (y_1, \cdots, y_M)' \in \mathcal{Y}$, define $G_{(M)}^{-1}(y) \equiv (G^{-1}(y_1), \cdots, G^{-1}(y_M))' \in \mathcal{T}$. Define also $S^*(y)$ for $y \in \mathcal{Y} \subset [0, 1]^M$, $S^*(y) \equiv S(G_{(M)}^{-1}(y))$. Note

that $S^*(y)$ is continuous at all $y \in \mathcal{Y}$ since $S(t)$ is continuous at all $t \in \mathcal{T}$. Then, we have

$$
\begin{aligned}
S(T_n(\theta_0)) &= S(G_{(M)}^{-1}(G_n)) \\
&= S^*(G_n) \\
&\xrightarrow{d} S^*(G_\infty) \\
&= S(G_{(M)}^{-1}(G_\infty)) = S(Z_\Sigma + \nu)
\end{aligned}
$$

where the first equality holds by the definition of $G_{(M)}^{-1}(\cdot)$, the second equality uses the definition of S^*. Convergence in the third line holds by (A.9), and the fourth and fifth equality uses the definition of S^*. ∎

A.2.2 Proof of Theorem 2.2

Proof. Using similar arguments to those used in the proof of Theorem 2.1, we have following decomposition of the t-statistic for any $\pi \in \Pi = [\underline{\pi}, 1]$ under Assumptions 2.2 and 2.3,

$$
T_n^*(\pi, \theta_0) = T_n(\lfloor \pi \bar{K} \rfloor, \theta_0) = t_n^*(\pi) - \sqrt{n} V_n(\pi)^{-1/2} r_n(\pi) + o_p(1), \tag{A.10}
$$

where

$$
t_n^*(\pi) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\tilde{P}_\pi(x)' \tilde{P}_\pi(x_i) \varepsilon_i}{V_n(\pi)^{1/2}}. \tag{A.11}
$$

Define $f_{n,\pi} : (\mathcal{E} \times \mathcal{X}) \mapsto \mathbb{R}$ for given $n \geq 1$, $\pi \in \Pi$,

$$
f_{n,\pi}(\varepsilon, t) = \frac{\tilde{P}_\pi(x)' \tilde{P}_\pi(t) \varepsilon}{V_n(\pi)^{1/2}} = \frac{\tilde{P}_{\lfloor \bar{K}\pi \rfloor}(x)' \tilde{P}_{\lfloor \bar{K}\pi \rfloor}(t) \varepsilon}{V_n(\lfloor \bar{K}\pi \rfloor, x)^{1/2}}, (\varepsilon, t) \in \mathcal{E} \times \mathcal{X}. \tag{A.12}
$$

Consider the class of measurable functions $\mathcal{F}_n = \{f_{n,\pi} : \pi \in \Pi\}$. Then, we consider following empirical process

$$
\left\{ t_n^*(\pi) : \pi \in \Pi \right\} = \left\{ n^{-1/2} \sum_{i=1}^n f_{n,\pi}(\varepsilon_i, x_i) : \pi \in \Pi \right\}
$$

which is indexed by classes of functions $\mathcal{F}_n$.

We want to show weak convergence of the empirical process $\{t_n^*(\cdot) : n \geq 1\}$ to a centered Gaussian process, $\mathbb{T}(\cdot)$ defined in the Theorem 2.2, in the space $\ell^\infty(\Pi)$ with totally bounded semimetric space (Π, ρ), where ρ is defined as $\rho(\pi_1, \pi_2) = |\pi_1 - \pi_2|$. Weak convergence results follow from marginal convergence to a multivariate normal distribution and asymptotic tightness. We closely follow Section 2.11.3 in van der Vaart and Wellner (1996) and verify conditions for the asymptotic tightness as in Theorem 2.11.22.

Note that the covariance kernel can be derived as follows

$$Ef_{n,\pi_1}f_{n,\pi_2} - Ef_{n,\pi_1}Ef_{n,\pi_2} = \frac{\tilde{P}_{\pi_1}(x)'E(\tilde{P}_{\pi_1}(x_i)\tilde{P}_{\pi_2}(x_i)'\varepsilon_i^2)\tilde{P}_{\pi_2}(x)}{V_n(\pi_1)^{1/2}V_n(\pi_2)^{1/2}}, \tag{A.13}$$

for any $\underline{\pi} \leq \pi_1 \leq \pi_2 \leq \bar{\pi} = 1$. This term converges to the claimed covariance kernel $\Sigma(\pi_1, \pi_2)$ under Assumption 2.4(i). This covariance kernel can be bounded below and above some constants $0 < C_1, C_2 < \infty$ for all n,

$$0 < C_1 \leq \underline{\sigma}^2 \frac{V_n(\pi_1)^{1/2}}{V_n(\pi_2)^{1/2}} \leq \frac{\tilde{P}_{\pi_1}(x)'E(\tilde{P}_{\pi_1}(x_i)\tilde{P}_{\pi_2}(x_i)'\varepsilon_i^2)\tilde{P}_{\pi_2}(x)}{V_n(\pi_1)^{1/2}V_n(\pi_2)^{1/2}} \leq \bar{\sigma}^2 \frac{V_n(\pi_1)^{1/2}}{V_n(\pi_2)^{1/2}} \leq C_2 \tag{A.14}$$

by using $\underline{\sigma}^2\tilde{P}_\pi(x)'\tilde{P}_\pi(x) \leq V_n(\pi) \leq \bar{\sigma}^2\tilde{P}_\pi(x)'\tilde{P}_\pi(x)$ from Assumption 2.2(ii). We also use $V_n(\pi_1)^{1/2} \asymp V_n(\pi_2)^{1/2} \asymp ||\tilde{P}_{\bar{\pi}}(x)||$ for any $\pi_1, \pi_2 \in \Pi$ under Assumption 2.4(ii).

We prove the finite dimensional convergence using similar arguments to those used in the proof of Theorem 2.1, but we repeat this here as Theorem 2.2 imposes assumption on $\mathcal{K}_n$. To show the finite dimensional convergence, it suffices to show that for any $0 < \underline{\pi} \leq \pi_1 < \cdots < \pi_M \leq 1$,

$$\delta' t_n^* \xrightarrow{d} N(0, \delta'\Sigma\delta) \quad \forall \delta \in \mathbb{R}^M \tag{A.15}$$

where $t_n^* = (t_n^*(\pi_1), ..., t_n^*(\pi_M))'$, $\Sigma(j,l) = \lim_{n\to\infty}\Sigma_n(j,l)$, $\Sigma_n(j,l) \equiv \frac{\tilde{P}_{\pi_j}(x)'E(\tilde{P}_{\pi_j i}\tilde{P}_{\pi_l i}'\varepsilon_i^2)\tilde{P}_{\pi_l}(x)}{V_n(\pi_j)^{1/2}V_n(\pi_l)^{1/2}}$. Similar to the proof of Theorem 2.1, we define $\omega_{ni} = (\delta'\Sigma_n\delta)^{-1/2}\sum_{j=1}^M \delta_j \frac{\tilde{P}_{\pi_j}(x)'\tilde{P}_{\pi_j i}\varepsilon_i}{V_n(\pi_j)^{1/2}}$. Note that $E\omega_{ni} = 0$, and $\frac{1}{n}\sum_{i=1}^n E[\omega_{ni}^2] = 1$, since $E[\omega_{ni}^2] = (\delta'\Sigma_n\delta)^{-1}\delta'Var(f_n(\varepsilon_i, x_i))\delta = 1$, where $f_n(\varepsilon_i, x_i) = (f_{n,\pi_1}(\varepsilon_i, x_i), ..., f_{n,\pi_M}(\varepsilon_i, x_i))'$. By Assumption 2.2, we have $||\sum_{j=1}^M \delta_j \frac{\tilde{P}_{\pi_j}(x)'\tilde{P}_{\pi_j i}}{V_n(\pi_j)^{1/2}}||_\infty \lesssim \zeta_{\bar{K}}\lambda_{\bar{K}}$. Lindberg's condition can be verified similarly as in the proof of Theorem 2.1, and finite-dimensional convergence holds by Lindberg-Feller CLT and Slutzky's Theorem.

Now, we only need to show stochastic equicontinuity. Define $\alpha(x,\pi) \equiv \tilde{P}_\pi(x)/V_n(\pi)^{1/2} = \tilde{P}_\pi(x)/||\Omega_\pi^{1/2}\tilde{P}_\pi(x)||$. Note that $|f_{n,\pi}(\varepsilon,t)| = |\alpha(x,\pi)'P_\pi(t)\varepsilon| \leq C|f_{n,\bar{\pi}}(\varepsilon,t)| \leq C|\varepsilon|\zeta_{\bar{K}}\lambda_{\bar{K}}$. We define envelope function $F_n(\varepsilon,t) \equiv |f_{n,\bar{\pi}}(\varepsilon,t)| \vee 1$. Without loss of generality, we assume that $F_n \geq 1$. Note that $Ef_{n,\pi}^2 = 1$ for any π, thus $EF_n^2 = O(1)$. Moreover, Lindeberg conditions can be verified easily as follows. For any $a > 0$,

$$E(F_n^2 1\{F_n > a\sqrt{n}\}) = E[(\frac{\tilde{P}_{\bar{\pi}}(x)'\tilde{P}_{\bar{\pi}}(x_i)}{V_n(\underline{\pi})^{1/2}}\varepsilon_i)^2 1\{|\varepsilon_i| > a(\sqrt{n}/(\zeta_{\bar{K}}\lambda_{\bar{K}}))\}] \tag{A.16}$$

$$\leq \sup_x E[\varepsilon_i^2 1\{|\varepsilon_i| > a(\sqrt{n}/(\zeta_{\bar{K}}\lambda_{\bar{K}}))\}|X_i = x] = o(1) \tag{A.17}$$

since $(\zeta_{\bar{K}}\lambda_{\bar{K}})/\sqrt{n} = o(1)$ and Assumption 2.2(ii). Moreover, for every $\delta_n \to 0$,

$$\sup_{\rho(\pi_1,\pi_2)<\delta_n} E(f_{n,\pi_1} - f_{n,\pi_2})^2 \to 0 \tag{A.18}$$

since $Ef_{n,\pi_1}f_{n,\pi_2} \to 1$ as $\rho(\pi_1, \pi_2) \to 0$.

38

Define $\kappa_{1,n} \equiv \sup_{\pi \neq \pi'} \frac{||\tilde{P}_{\pi'-\pi}(x)||}{||\pi'-\pi||}$ where $\tilde{P}_{\pi'-\pi}(x) = (\tilde{p}_{\lfloor \bar{K}\pi \rfloor+1}(x),...,\tilde{p}_{\lfloor \bar{K}\pi' \rfloor}(x))'$. For sufficiently large n, $\kappa_{1,n} \lesssim ||\tilde{P}_{\pi'-\pi}(x)|| \lesssim V_n(\pi'-\pi)^{1/2}$ under Assumption 2.2 and 2.4(ii). Also define $\kappa_{2,n} \equiv \sup_{\pi \neq \pi'} \frac{|V_n(\pi')-V_n(\pi)|}{||\pi'-\pi||}$. Then, for any $\pi, \pi' \in \Pi$ such that $\pi < \pi'$, following holds for sufficiently large n,

$$
\begin{aligned}
|\alpha(x,\pi')'P_{\pi'}(t) - \alpha(x,\pi)'P_\pi(t)| &= |\frac{\tilde{P}_{\pi'}(x)'\tilde{P}_{\pi'}(t)}{V_n(\pi')^{1/2}} - \frac{\tilde{P}_\pi(x)'\tilde{P}_\pi(t)}{V_n(\pi)^{1/2}}| \\
&\leq |\frac{\tilde{P}_{\pi'}(x)'\tilde{P}_{\pi'}(t) - \tilde{P}_\pi(x)'\tilde{P}_\pi(t)}{V_n(\pi')^{1/2}}| + |\tilde{P}_\pi(x)'\tilde{P}_\pi(t)(\frac{1}{V_n(\pi')^{1/2}} - \frac{1}{V_n(\pi)^{1/2}})| \\
&\leq (\sup_\pi \frac{1}{|V_n(\pi)^{1/2}|})|\tilde{P}_{\pi'-\pi}(x)'\tilde{P}_{\pi'-\pi}(t)| + |\frac{\tilde{P}_\pi(x)'\tilde{P}_\pi(t)}{V_n(\pi)^{1/2}}(\frac{V_n(\pi') - V_n(\pi)}{V_n(\pi')^{1/2}(V_n(\pi)^{1/2} + V_n(\pi')^{1/2})})| \\
&\leq C_1(\sup_\pi \frac{1}{|V_n(\pi)^{1/2}|})\kappa_{1,n}\zeta_{\bar{K}}\lambda_{\bar{K}}||\pi'-\pi|| + C_2\zeta_{\bar{K}}\lambda_{\bar{K}}\frac{1}{\inf_\pi |V_n(\pi)|}\kappa_{2,n}||\pi'-\pi|| \\
&\leq C_3\zeta_{\bar{K}}\lambda_{\bar{K}}||\pi'-\pi|| + C_4\zeta_{\bar{K}}\lambda_{\bar{K}}||\pi'-\pi|| = A\zeta_{\bar{K}}\lambda_{\bar{K}}||\pi'-\pi||
\end{aligned}
$$

where C_1, C_2, C_3, C_4, A are some constants do not depend on n. The third inequality uses the definition of $\kappa_{1,n}, \kappa_{2,n}$, $|\tilde{P}_{\pi'-\pi}(t)| \lesssim \zeta_{\bar{K}}\lambda_{\bar{K}}$ and $|\frac{\tilde{P}_\pi(x)'\tilde{P}_\pi(t)}{V_n(\pi)^{1/2}}| \lesssim \zeta_{\bar{K}}\lambda_{\bar{K}}$ under Assumption 2.2 and 2.3. The last inequality uses $\kappa_{1,n} \lesssim V_n(\pi'-\pi)^{1/2}$, $\kappa_{2,n} \lesssim \sup_\pi V_n(\pi)$, and $V_n(\pi) \asymp V_n(\pi')$ for any $\pi, \pi' \in \Pi$ under Assumption 2.4(ii).

From this, we have

$$
|f_{n,\pi'} - f_{n,\pi}| = |\varepsilon\alpha(x,\pi')'P_{\pi'}(t) - \varepsilon\alpha(x,\pi)'P_\pi(t)| \leq |\varepsilon|A\zeta_{\bar{K}}\lambda_{\bar{K}}||\pi'-\pi||. \tag{A.19}
$$

Therefore, the class of functions $\mathcal{F}_n = \{f_{n,\pi} : \pi \in \Pi\}$ satisfies Lipschitz conditions for each n, and this implies that there are constants $A, V > 0$ such that

$$
\sup_Q N(\epsilon||F_n||_{L^2(Q)}, \mathcal{F}_n, L^2(Q)) \leq (A/\epsilon)^V, 0 < \forall\epsilon \leq 1 \tag{A.20}
$$

for each n. Then, following uniform-entropy condition holds for every $\delta_n \to 0$.

$$
J(\delta_n, \mathcal{F}_n, L^2(Q)) = \int_0^{\delta_n} \sqrt{\log \sup_Q N(\epsilon||F_n||_{L^2(Q)}, \mathcal{F}_n, L^2(Q))} \longrightarrow 0. \tag{A.21}
$$

Thus, by the Theorem 2.11.22 in van der Vaart and Wellner (1996), we have shown that the sequence $\{t_n^*(\pi) : \pi \in \Pi\}$ is asymptotically tight in $\ell^\infty(\Pi)$. Together with the definition of $\nu(\pi) = \lim_{n\to\infty} -\sqrt{n}V_n(\pi)^{-1/2}r_n(\pi)$, we have $T_n^*(\pi, \theta_0) \Rightarrow \mathbb{T}(\pi)+\nu(\pi)$ for $\pi \in \Pi$. In addition, if Assumption 2.5 holds, then $|\sqrt{n}V_n(\pi)^{-1/2}r_n(\pi)| = o(1)$ for any $\pi \in \Pi$. Therefore, $T_n^*(\pi, \theta_0) \Rightarrow \mathbb{T}(\pi)$. This completes the proof. ∎

A.2.3 Proof of Corollary 3.1

Proof. First consider $S(t) = \sup_m |t_m|$ for $t = (t_1, \cdots, t_M) \in \mathbb{R}^M$. Since $S(t)$ is continuous at all $t \in \mathcal{T} = \mathbb{R}^M$ under $\sup_m |\nu(m)| < +\infty$, $\text{Sup}\, T_n(\theta_0) = \sup_{K_m \in \mathcal{K}_n} |T_n(K_m, \theta_0)| = S(T_n(\theta_0)) \xrightarrow{d} S(Z_\Sigma + \nu) = \sup_m |Z_m + \nu(m)|$ follows from Theorem 2.1. If $|\nu(m)| = +\infty$ for some m, $|T_n(K_m, \theta_0)| \xrightarrow{p} +\infty$ and thus $\text{Sup}\, T_n(\theta_0) \xrightarrow{p} +\infty$.

For the second part of Corollary, we consider $S(t) \equiv \inf_m |t_m|$ for $t = (t_1, \cdots, t_M) \in \mathcal{T} = \mathbb{R}^{M-1}_{[\pm\infty]} \times \mathbb{R}$. Note that $S(t)$ is continuous at all $t \in \mathcal{T} = \mathbb{R}^{M-1}_{[\pm\infty]} \times \mathbb{R}$ under Assumption 2.1 by $\inf |\nu(m)| = O(1)$. Then, we have

$$\text{Inf}\, T_n(\theta_0) = S(T_n(\theta_0)) \xrightarrow{d} S(Z_\Sigma + \nu) = \inf_m |Z_m + \nu(m)| \tag{A.22}$$

by Theorem 2.1. This completes the proof of Corollary 3.1. ∎

A.2.4 Proof of Corollary 3.2

Proof. From Corollary 3.1, we have

$$\lim_{n\to\infty} P(\text{Sup}\, T_n(\theta_0) > c^{\text{sup}}_{1-\alpha}(\nu)) = P(\xi_{\text{sup}}(\Sigma, \nu) > c^{\text{sup}}_{1-\alpha}(\nu)) = \alpha$$

by the definition of $c^{\text{sup}}_{1-\alpha}(\nu)$ in (3.4). For the second part of Corollary, we have

$$\limsup_{n\to\infty} P(|T_n(\widehat{K}, \theta_0)| > c^{\text{sup}}_{1-\alpha}(\nu)) \leq \limsup_{n\to\infty} P(\text{Sup}\, T_n(\theta_0) > c^{\text{sup}}_{1-\alpha}(\nu))$$

$$= \lim_{n\to\infty} P(\text{Sup}\, T_n(\theta_0) > c^{\text{sup}}_{1-\alpha}(\nu)) = \alpha$$

where the first inequality uses $|T_n(\widehat{K}, \theta)| \leq \text{Sup}\, T_n(\theta)$ for any $\widehat{K} \in \mathcal{K}_n$ and the second equality holds holds under subsequence $\{u_n\}$ of $\{n\}$ by the definition of $\limsup$. ∎

A.2.5 Proof of Corollary 3.3

Proof. This follows from Corollary 3.1 and using the similar arguments as in the proof of Corollary 3.2. ∎

A.2.6 Proof of Corollary 3.4

Proof. Using the same decomposition of the t-statistic as in the proof of Theorem 2.1, for any $m = 1, 2, \cdots M$, we have

$$
\begin{aligned}
T_n(K_m, \theta_n) &= \sqrt{\frac{n}{V_n(K_m)}} (\widehat{\theta}_n(K_m) - \theta_n) = T_n(K_m, \theta_0) - \sqrt{\frac{n}{V_n(K_m)}} \frac{\bar{\mu}}{n^\gamma} \\
&= t_n(m) + \nu_n(m) + \mu_n(m) + o_p(1),
\end{aligned}
$$

where $t_n(m) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\tilde{P}_{K_m}(x)' \tilde{P}_{K_m i} \varepsilon_i}{V_n(K_m)^{1/2}}$, $\nu_n(m) = -\sqrt{n} \frac{r_n(K_m)}{V_n(K_m)^{1/2}}$, and $\mu_n(m) = -\sqrt{n} \frac{\bar{\mu} n^{-\gamma}}{V_n(K_m)^{1/2}}$.

Under Assumption 3.1 and $\sup_m |\nu(m)| < +\infty$, $\operatorname{Sup} T_n(\theta_n) \xrightarrow{p} +\infty$ for any $\gamma < \delta_1$ as $\mu_n(m) \asymp n^{\delta_m - \gamma}$. From the null limiting distribution in Corollary 3.1, $c_{1-\alpha}^{\sup}(\nu)$ is bounded in probability, thus we conclude that the test based on $\operatorname{Sup} T_n(\theta)$ is consistent against all $n^{-\gamma}$-local alternatives $\gamma < \delta_1$.

For $\gamma = \delta_1$,

$$\operatorname{Sup} T_n(\theta_n) \xrightarrow{d} \sup_{m=1,\cdots,M} |Z_m + \nu(m) + \mu(m)|, \tag{A.23}$$

where $\mu = (\mu(1), \cdots, \mu(M))' = (\mu(1), 0'_{M-1})'$ since $\mu_n(m) = o(1)$ for all $m \neq 1$, and $\mu_n(1) = O(1)$. From Corollary 3.1, we also conclude that the test based on $\operatorname{Sup} T_n(\theta)$ has nontrivial asymptotic local power against local alternatives $\gamma = \delta_1$, but not against $\gamma > \delta_1$. ∎

A.2.7 Proof of Corollary 3.5

Proof. Under Assumption 3.1, for any $\gamma < \delta_M$, $\operatorname{Inf} T_n(\theta_n) \xrightarrow{p} +\infty$ because $\inf_m |\mu_n(m)| \to +\infty$ as $n \to \infty$. From the Corollary 3.1, $c_{1-\alpha}^{\inf}(\nu)$ is bounded in probability, thus

$$\lim_{n \to \infty} P(\operatorname{Inf} T_n(\theta_n) > c_{1-\alpha}^{\inf}(\nu)) = 1$$

for any $\gamma < \delta_M$.

From Corollary 3.1 and the assumption $|\nu(m)| = +\infty$ for all $m < p + 1$, we have the behavior of the test statistic under the null,

$$\operatorname{Inf} T_n(\theta_0) \xrightarrow{d} \inf_{m=p+1,\cdots,M} |Z_m + \nu(m)| \tag{A.24}$$

since corresponding elements of $|Z_m + \nu(m)| = +\infty$ for $m < p + 1$.

Using the same arguments in the proof of Theorem 2.1, we can show that for any $\gamma \geq \delta_M$ and any continuous function $S(t)$ at all $t \in \mathcal{T} \subset \mathbb{R}_{[\pm\infty]}^{M-1} \times \mathbb{R}$

$$S(T_n(\theta_n)) \xrightarrow{d} S(Z_\Sigma + \nu + \mu). \tag{A.25}$$

Therefore, following holds for any $\gamma \in [\delta_M, \delta_{p+1}]$,

$$\operatorname{Inf} T_n(\theta_n) \xrightarrow{d} \inf_{m=p+1,\cdots,M} |Z_m + \nu(m) + \mu(m)| \tag{A.26}$$

where $Z_\Sigma = (Z_1, \cdots, Z_M)' \sim N(0, \Sigma)$ and $\nu = (\nu(1), \cdots, \nu(M))'$ are defined in Theorem 2.1, and $\mu = (\mu(1), \cdots, \mu(M))'$ is defined in (3.8). Note that if either $\nu(m) = \pm\infty$ or $\mu(m) = \pm\infty$, then the corresponding element of $Z_\Sigma + \nu + \mu$ equals $\pm\infty$. Technically, we exclude the possibility $|\nu(m)| = +\infty$ and $|\mu(m)| = +\infty$, but $|\nu(m) + \mu(m)| = 0$. This may not be restrictive, as we only impose rate conditions $\nu_n(m) \asymp n^{\delta_m} r_n(K_m)$ and $\mu_n(m) \asymp n^{\delta_m - \gamma}$.

To show the test based on the $\operatorname{Inf} T_n(\theta_n)$ has nontrivial local power against some alternatives $\gamma \in \{\delta_{p+1}, \cdots, \delta_M\}$ and has power no less than the size, we need to consider different cases. For

the first case $\gamma = \delta_{p+1}$, we have $|\mu(m)| = 0$ for all $m > p + 1$ and $0 < |\mu(p+1)| < +\infty$, thus the test has nontrivial asymptotic local power and the power is an monotonically increasing function in $|\mu(p+1)|$.

For the next case, $\delta_{p+2} < \gamma < \delta_{p+1}$, following holds

$$\text{Inf } T_n(\theta_n) \xrightarrow{d} \inf_{m=p+2,\cdots,M} |Z_m + \nu(m)| \tag{A.27}$$

as $|\mu(p+1)| = +\infty$ and $|\mu(m)| = 0$ for all $m > p + 1$. Note that $\inf_{m=p+2,\cdots,M} |Z_m + \nu(m)| \geq \inf_{m=p+1,\cdots,M} |Z_m + \nu(m)|$ holds for the null limiting distribution in (A.24). Thus, $\lim_{n\to\infty} P(\text{Inf } T_n(\theta_n) > c_{1-\alpha}^{\text{inf}}(\nu)) \geq \alpha$. Using the similar arguments, we can show that the test based on $\text{Inf } T_n(\theta)$ has nontrivial local power against $\gamma \in \{\delta_{p+1}, \cdots, \delta_M\}$ and has power no less than α for all other $\gamma \in [\delta_M, \delta_{p+1}]$, but does not have power against $\gamma > \delta_{p+1}$ as $|\mu(m)| = 0$ for all $m \geq p + 1$. ∎

A.2.8 Proof of Corollary 4.1

Proof. Under the Assumptions 2.1, 2.2, 2.5, and 4.1, following holds

$$(T_{n,\widehat{V}}(K_1,\theta_0),\cdots,T_{n,\widehat{V}}(K_M,\theta_0))' = AT_n(\theta) \xrightarrow{d} Z_\Sigma \sim N(0,\Sigma) \tag{A.28}$$

where $T_{n,\widehat{V}}(K,\theta) = \frac{\sqrt{n}(\widehat{\theta}_n(K)-\theta_0)}{\widehat{V}_n(K)^{1/2}} = \frac{V_n(K)^{1/2}}{\widehat{V}_n(K)^{1/2}}T_n(K,\theta)$, $A \equiv diag\{\frac{V_n(K_1)^{1/2}}{\widehat{V}_n(K_1)^{1/2}},\cdots,\frac{V_n(K_M)^{1/2}}{\widehat{V}_n(K_M)^{1/2}}\}$ by Theorem 2.1, $A \xrightarrow{p} I_M$ and Slutzky Theorem.

Next consider $\widehat{c}_{1-\alpha}^{\text{sup}}$, which is $(1-\alpha)$ quantile of $\sup_{m=1,\cdots,M} |Z_{m,\widehat{\Sigma}}|$ defined in (4.5),

$$\widehat{c}_{1-\alpha}^{\text{sup}} = \inf\{x \in \mathbb{R} : P(\sup_{m=1,\cdots,M} |Z_{m,\widehat{\Sigma}}| \leq x) \geq 1 - \alpha\}$$

where $Z_{\widehat{\Sigma}} = (Z_{1,\widehat{\Sigma}},\cdots,Z_{M,\widehat{\Sigma}})' \sim N(0,\widehat{\Sigma})$, $\widehat{\Sigma}(j,j) = 1, \widehat{\Sigma}(j,l) = \widehat{V}_n(K_j,K_l)/\widehat{V}_n(K_j)^{1/2}\widehat{V}_n(K_l)^{1/2}$. Note that for any $j,l \in \{1,2,\cdots,M\}$,

$$\widehat{\Sigma}(j,l) = \frac{\widehat{V}_n(K_j,K_l)}{\widehat{V}_n(K_j)^{1/2}\widehat{V}_n(K_l)^{1/2}} = \frac{\widehat{V}_n(K_j,K_l)}{V_n(K_j,K_l)} \frac{V_n(K_j,K_l)}{V_n(K_j)^{1/2}V_n(K_l)^{1/2}} \frac{V_n(K_j)^{1/2}}{\widehat{V}_n(K_j)^{1/2}} \frac{V_n(K_l)^{1/2}}{\widehat{V}_n(K_l)^{1/2}} \xrightarrow{p} \Sigma(j,l) \tag{A.29}$$

by Assumption 4.1. Therefore, $\widehat{\Sigma} \xrightarrow{p} \Sigma$, $Z_{\widehat{\Sigma}} \xrightarrow{d} Z_\Sigma$, and $\sup_{m=1,\cdots,M} |Z_{m,\widehat{\Sigma}}| \xrightarrow{d} \sup_{m=1,\cdots,M} |Z_m|$ hold, and thus, $\widehat{c}_{1-\alpha}^{\text{sup}} \xrightarrow{p} c_{1-\alpha}^{\text{sup}}(0_M)$. Similarly, we can show $\widehat{c}_{1-\alpha}^{\text{inf}} \xrightarrow{p} c_{1-\alpha}^{\text{inf}}(0_M)$.

Then we have,

$$\lim_{n\to\infty} P(\theta_0 \in [\widehat{\theta}_n(K) \pm \widehat{c}_{1-\alpha}^{\text{sup}} s(\widehat{\theta}_n(K))] \quad \forall K \in \mathcal{K}_n) = \lim_{n\to\infty} P(\sup_{K_m \in \mathcal{K}_n} |T_{n,\widehat{V}}(K_m,\theta_0)| \leq c_{1-\alpha}^{sup}(0_M) + o_p(1))$$

$$= P(\sup_m |Z_m| \leq c_{1-\alpha}^{sup}(0_M)) = 1 - \alpha.$$

Coverage property of the CI, $[\widehat{\theta}_n(\widehat{K}) \pm \widehat{c}_{1-\alpha}^{\mathrm{sup}} s(\widehat{\theta}_n(\widehat{K}))]$ for $\widehat{K} \in \mathcal{K}_n$ immediately follows from the above equation and the coverage of $CI_{\mathrm{inf}}^{\mathrm{Robust}}$ can be similarly derived. ∎

A.2.9 Proof of Theorem 5.1

Proof. Conditional on $X = [x_1, \cdots, x_n]'$, following decomposition holds for any single sequence $K \in \mathcal{K}_n$

$$\sqrt{n}(\widehat{\theta}_n(K) - \theta_0) = \widehat{\Gamma}_K^{-1} S_K,$$

$$\widehat{\Gamma}_K = \frac{1}{n}(W' M_K W), \quad S_K = \frac{1}{\sqrt{n}} W' M_K (g + \varepsilon)$$

where $g = [g_1, \cdots, g_n]', g_i = g_0(x_i), g_w = [g_{w1}, \cdots, g_{wn}]', g_{wi} = g_{w0}(x_i) = E[w_i|x_i], v = [v_1, \cdots, v_n]$. All remaining proofs contain conditional expectations (conditioning on X) hold almost surely (a.s.).

Under Assumption 5.2 and $E[v_i^2|x_i] = E[v_i^2]$,

$$\widehat{\Gamma}_K = \Gamma_K + o_p(1), \quad \Gamma_K = (1 - K/n)E[v_i^2] \tag{A.30}$$

by Lemma 1 of Cattaneo, Jansson, and Newey (2015a). Moreover,

$$S_K = \frac{1}{\sqrt{n}} v' M_K \varepsilon + \frac{1}{\sqrt{n}} g_w' M_K g + \frac{1}{\sqrt{n}}(v' M_K g + g_w' M_K \varepsilon) \tag{A.31}$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^n M_{K,ii} v_i \varepsilon_i - \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{j=1, j<i}^n P_{K,ij}(v_i \varepsilon_j + v_j \varepsilon_i) + o_p(1) \tag{A.32}$$

since $M_{K,ij} = -P_{K,ij}$ for $j < i$, $\frac{1}{\sqrt{n}} g_w' M_K g = O_p(\sqrt{n} \bar{K}^{-\gamma_g - \gamma_{g_w}}) = o_p(1)$, $\frac{1}{\sqrt{n}}(v' M_K g + g_w' M_K \varepsilon) = O_p(\bar{K}^{-\gamma_g} + \bar{K}^{-\gamma_{g_w}}) = o_p(1)$ by Lemma 2 of Cattaneo, Jansson and Newey (2015a) under Assumption 5.2. Under $E[\varepsilon_i^2|w_i, x_i] = \sigma_\varepsilon^2$ following holds

$$T_n(K, \theta_0) = \sqrt{n} V_n(K)^{-1/2}(\widehat{\theta}_n(K) - \theta_0) = V_n(K)^{-1/2} \Gamma_K^{-1} \frac{1}{\sqrt{n}} v' M_K \varepsilon + o_p(1) \xrightarrow{d} N(0,1)$$

by Theorem 1 of Cattaneo, Jansson and Newey (2015a) which follows from Lemma A2 in Chao, Swanson, Hausman, Newey and Woutersen (2012).

For simplicity, here we only show the joint convergence of bivariate t-statistics, but the proof can be easily extended to multivariate case. For any $K_1 < K_2$ in $\mathcal{K}_n$, we show

$$\delta_1 T_n(K_1, \theta_0) + \delta_2 T_n(K_2, \theta_0) \xrightarrow{d} N(0, (\delta_1^2 + \delta_2^2 + 2\delta_1 \delta_2 v_{12})), \quad \forall(\delta_1, \delta_2) \in \mathbb{R}^2 \tag{A.33}$$

where $v_{12} = \lim_{n \to \infty} V_n(K_1)^{1/2}/V_n(K_2)^{1/2}$. We closely follows the proof of Lemma A2 in Chao et

al. (2012). Define $Y_n, Y_{1,n}$ and $Y_{2,n}$ as follows

$$Y_n = \delta_1 Y_{1,n} + \delta_2 Y_{2,n}, \tag{A.34}$$

$$Y_{1,n} = \omega_{1,1n} + \sum_{i=2}^{n} y_{1,in}, \quad y_{1,in} = \omega_{1,in} + \bar{y}_{1,in}, \tag{A.35}$$

$$Y_{2,n} = \omega_{2,1n} + \sum_{i=2}^{n} y_{2,in}, \quad y_{2,in} = \omega_{2,in} + \bar{y}_{2,in}, \tag{A.36}$$

where $\omega_{1,in} = V_n(K_1)^{-1/2}\Gamma_{K_1}^{-1}M_{K_1,ii}/\sqrt{n}$, $\bar{y}_{1,in} = \sum_{j<i}(u_{1,j}P_{K_1,ij}\varepsilon_i + u_{1,i}P_{K_1,ij}\varepsilon_j)/\sqrt{n}, u_{1,i} = V_n(K_1)^{-1/2}\Gamma_{K_1}^{-1}v_i$ and $\omega_{2,in}, \bar{y}_{2,in}$ are similarly defined with appropriate terms $P_{K_2}, V_n(K_2), \Gamma_{K_2}$ with K_2. Similar to the proof of Lemma A2 in Chao et al. (2012), $\omega_{1,1n} = o_p(1), \omega_{2,1n} = o_p(1)$. Thus, we only need to show that following holds conditional on X with probability one

$$\sum_{i=2}^{n}(\delta_1 y_{1,in} + \delta_2 y_{2,in}) \xrightarrow{d} N(0, \delta_1^2 + \delta_2^2 + 2\delta_1\delta_2 v_{12}). \tag{A.37}$$

It remains to provide Lindeberg-Feller condition.

$$E[(\sum_{i=2}^{n}\delta_1 y_{1,in} + \delta_2 y_{2,in})^2|X] = \delta_1^2 E[(\sum_{i=2}^{n}y_{1,in})^2|X] + \delta_2^2 E[(\sum_{i=2}^{n}y_{2,in})^2|X]$$

$$+2\delta_1\delta_2 E[\sum_{i=2}^{n}\sum_{j=2}^{n}y_{1,in}y_{2,in}|X], \tag{A.38}$$

where the first and second terms in (A.38) goes to δ_1^2, δ_2^2 a.s., respectively, as in the proof of Lemma A.2 in Chao et al. (2012). Note that $E[\omega_{1,in}\bar{y}_{2,in}|X] = 0, E[\omega_{2,in}\bar{y}_{1,in}|X] = 0$, and $E[\omega_{1,1n}\omega_{2,in}|X] = 0, E[\omega_{2,1n}\omega_{1,in}|X] = 0$ for any $i > 1$. Followings are the key calculations for the asymptotic variance of leading terms in Y_n.

$$E[Y_{1,n}Y_{2,n}|X] = \frac{1}{n}V_n(K_1)^{-1/2}\Gamma_{K_1}^{-1}E[v'M_{K_1}\varepsilon\varepsilon'M_{K_2}v|X]\Gamma_{K_2}^{-1}V_n(K_2)^{-1/2} \tag{A.39}$$

$$= \frac{1}{n}V_n(K_1)^{-1/2}\Gamma_{K_1}^{-1}\sigma_\varepsilon^2 E[v'M_{K_2}v|X]\Gamma_{K_2}^{-1}V_n(K_2)^{-1/2} \tag{A.40}$$

$$= V_n(K_1)^{-1/2}\Gamma_{K_1}^{-1}\sigma_\varepsilon^2\Gamma_{K_2}\Gamma_{K_2}^{-1}V_n(K_2)^{-1/2} \tag{A.41}$$

$$= V_n(K_1)^{1/2}/V_n(K_2)^{1/2} \tag{A.42}$$

where the second equality uses conditional homoskedasticity $E[\varepsilon\varepsilon'|X, W] = \sigma_\varepsilon^2 I$ and $M_{K_1}M_{K_2} = M_{K_2}$, the third equality uses $tr(M_{K_2}) = n - K_2$ and $E[v^2|X] = E[v^2]$, and the last equality uses

$V_n(K_1) = \sigma_\varepsilon^2 \Gamma_{K_1}^{-1}$. Therefore, we calculate components of last terms in (A.38) as follows

$$E[\sum_{i=2}^{n} \sum_{j=2}^{n} y_{1,in} y_{2,in}|X] = E[Y_{1,n} Y_{2,n}|X] - \sum_{i=2}^{n} E[\omega_{1,1n} y_{2,in}|X]$$

$$- \sum_{i=2}^{n} E[\omega_{2,1n} y_{1,in}|X] - E[\omega_{1,1n} \omega_{2,1n}|X] \tag{A.43}$$

$$= V_n(K_1)^{1/2}/V_n(K_2)^{1/2} - E[\omega_{1,1n} \omega_{2,1n}|X] \to v_{12} \quad a.s. \tag{A.44}$$

As in the proof of Lemma A.2 of Chao et al. (2012), we have

$$\sum_{i=2}^{n} E[(\delta_1 y_{1,in} + \delta_2 y_{2,in})^4|X] \lesssim \sum_{i=2}^{n} E[(y_{1,in})^4|X] + \sum_{i=2}^{n} E[(y_{2,in})^4|X] \to 0 \quad a.s. \tag{A.45}$$

Thus, by similar arguments following the proof of Lemma A.2 in Chao et al. (2012), we can apply the martingale central limit theorem. Then, by Slutzky theorem, joint convergence holds with the claimed covariance. In addition, if $\sup_{K \in \mathcal{K}_n} |\frac{\widehat{V}_n(K)}{V_n(K)} - 1| = o_p(1)$ as $n, K \to \infty$ holds, we can show the coverage results using similar arguments to those used in the proof of Corollary 4.1. This completes the proof. ∎
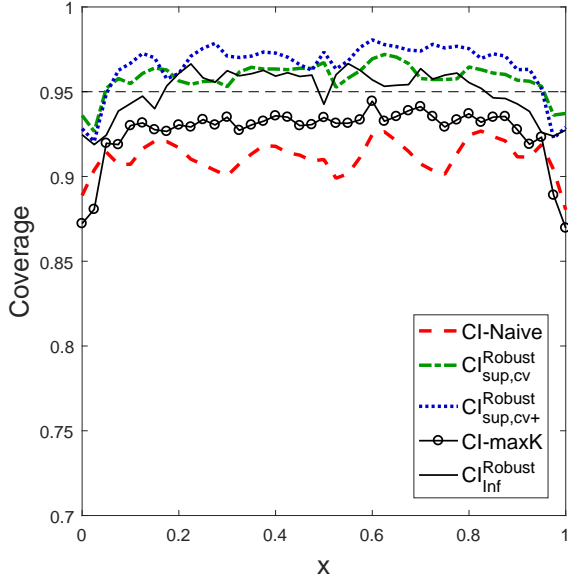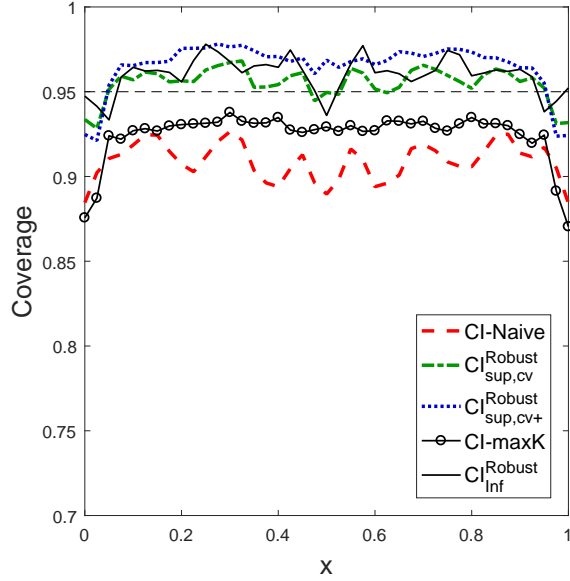
Appendix B Figures and Tables

Figure 1: Coverage and Length of CIs - Polynomials
Coverage and average length of nominal 95% CIs for $g(x)$:
(1) $CI^{\texttt{Naive}}$ with $\widehat{K}_{\texttt{cv}}$ (2) $CI^{\texttt{Robust}}_{\texttt{sup}}$ with $\widehat{K}_{\texttt{cv}}$ (3) $CI^{\texttt{Robust}}_{\texttt{sup}}$ with $\widehat{K}_{\texttt{cv}} + 2$ (4) $CI_{\texttt{maxK}}$ with $\bar{K}$ (5) $CI^{\texttt{Robust}}_{\texttt{inf}}$
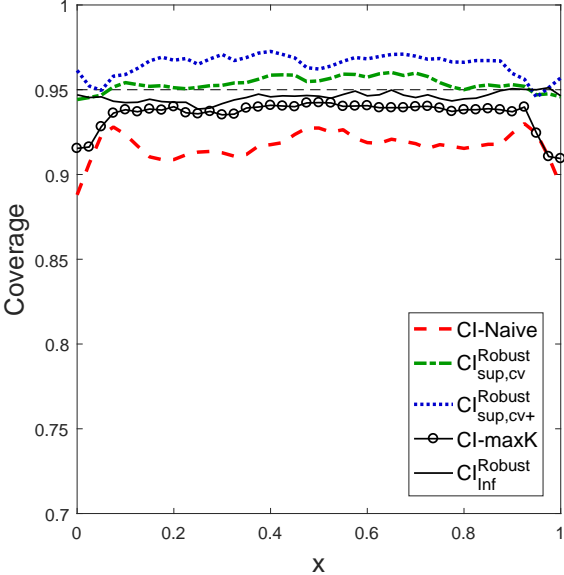
$g_1(x) = 4x - 1$
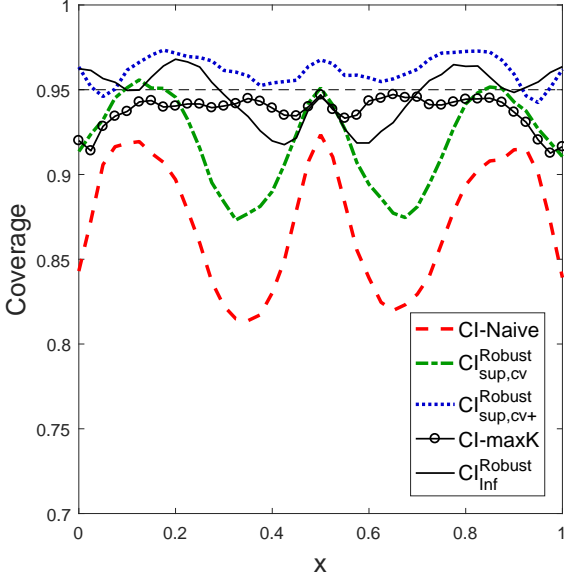
$g_2(x) = \ln(|6x - 3| + 1)sgn(x - 1/2)$
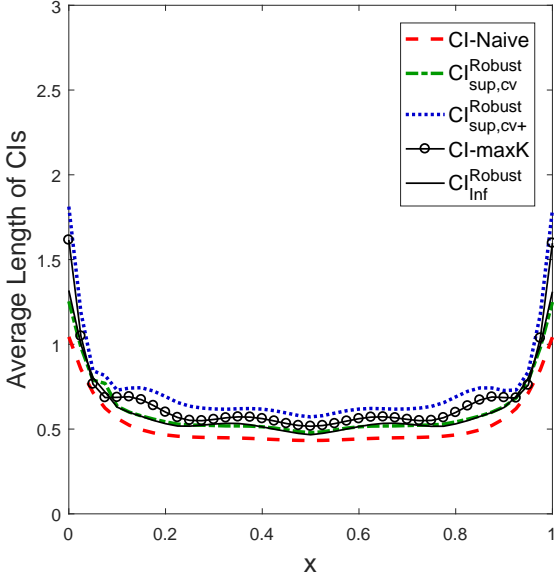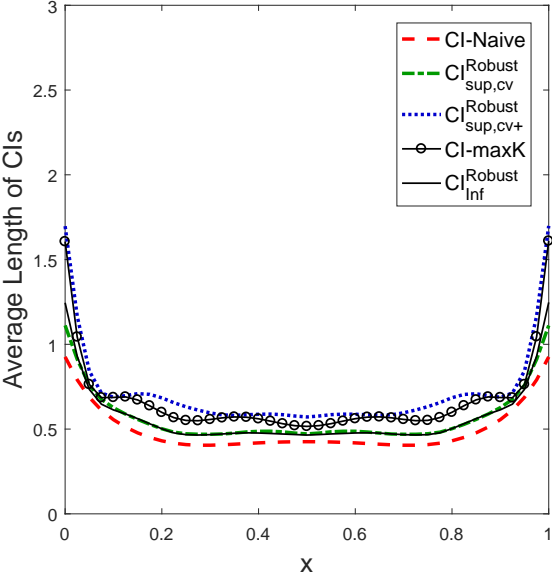
Figure 2: Coverage and Length of CIs - Quadratic Splines
Coverage and average length of nominal 95% CIs for $g(x)$:
(1) $CI^{\texttt{Naive}}$ with $\widehat{K}_{\texttt{cv}}$ (2) $CI^{\texttt{Robust}}_{\texttt{sup}}$ with $\widehat{K}_{\texttt{cv}}$ (3) $CI^{\texttt{Robust}}_{\texttt{sup}}$ with $\widehat{K}_{\texttt{cv}} + 2$ (4) $CI_{\texttt{maxK}}$ with $\bar{K}$ (5) $CI^{\texttt{Robust}}_{\texttt{inf}}$

$$g_3(x) = \frac{\sin(7\pi x/2)}{1 + 2x^2(sgn(x)+1)}$$

$$g_4(x) = x - 1/2 + 5\phi(10(x - 1/2))$$
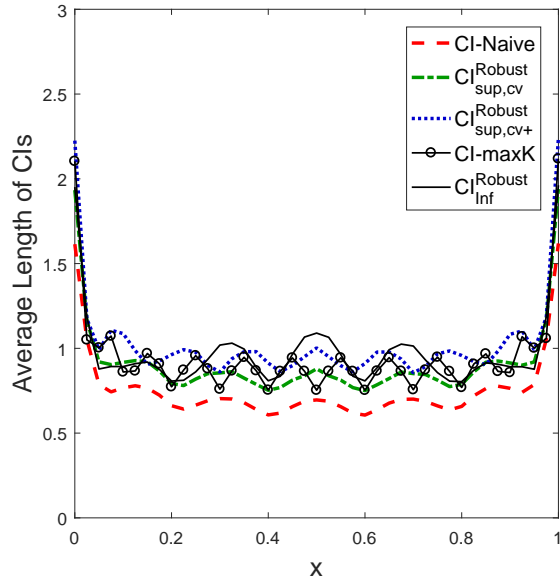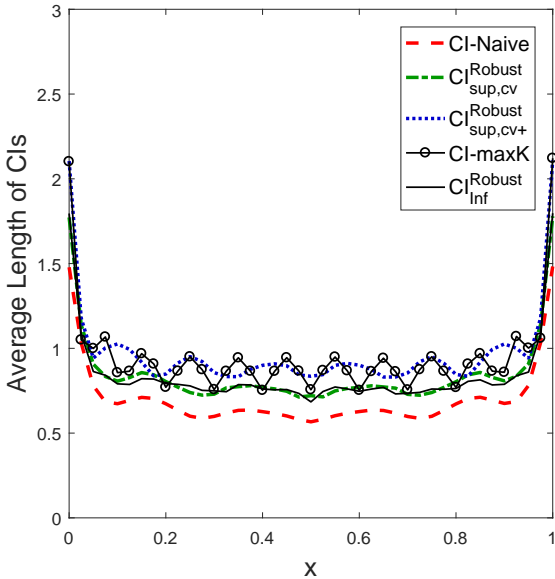
Figure 3: Power function against fixed alternatives.
$H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_0 + \delta$, where $\theta_0 = g_2(x)$ at $x = 0.5$ for figure (a) and $x = 0.4$ for figure (b). $g_2(x) = \ln(|6x - 3| + 1)sgn(x - 1/2)$.
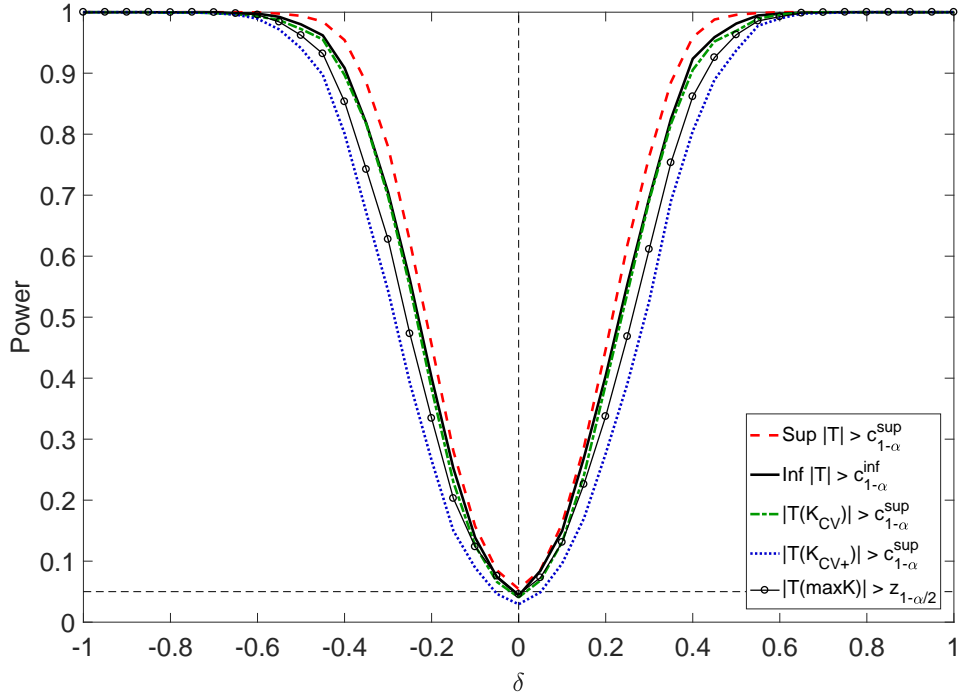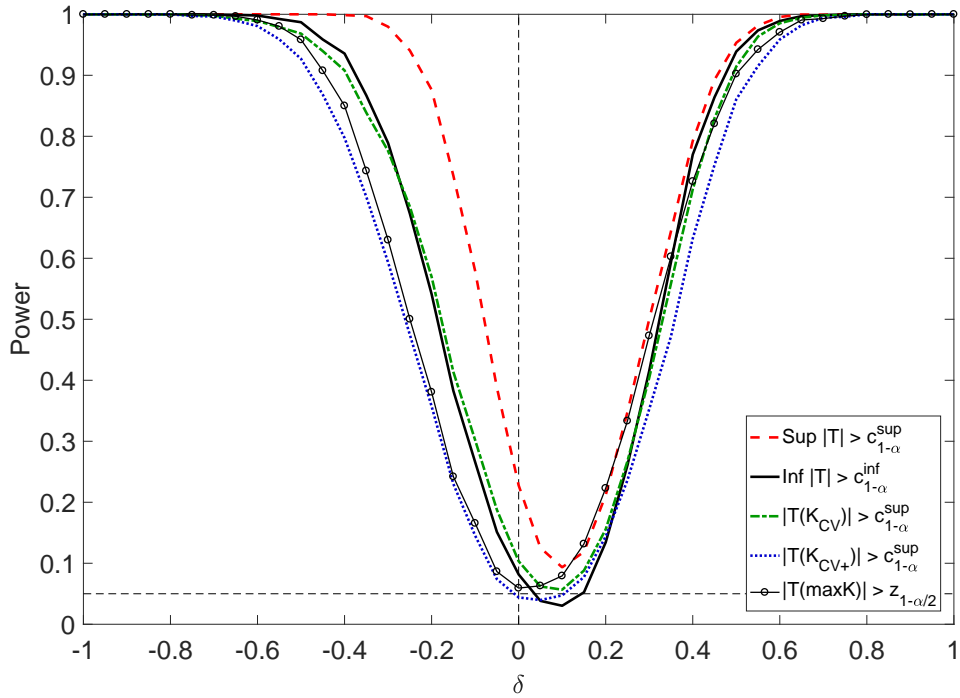
Table 1: Nonparametric Wage Elasticity of Hours of Work Estimates in Blomquist and Newey (Table 1, 2002). Wage elasticity evaluated at the mean net wage rates, virtual income, and level of hours.

Additional Terms[1]	CV^2	$\widehat{E}_w$	$SE_{\widehat{E}_w}$	$CI_{\widehat{E}_w}$
$1, y_J, w_J$	0.00472	0.0372	0.0104	[0.0168, 0.0576]
$\Delta y \Delta w$	0.0313	0.0761	0.0128	[0.0510, 0.1012]
$\ell \Delta y$	0.0305	0.0760	0.0127	[0.0511, 0.1009]
y_J^2, w_J^2	0.0323	0.0763	0.0129	[0.0510, 0.1016]
$\Delta y^2, \Delta w^2$	0.0369	0.0543	0.0151	[0.0247, 0.0839]
$y_J w_J$	0.0364	0.0659	0.0197	[0.0273, 0.1045]
Δyw	0.0350	0.0628	0.0223	[0.0191, 0.1065]
$\ell^2 \Delta y$	0.0364	0.0636	0.0223	[0.0199, 0.1073]
y_J^3, w_J^3	0.0331	0.0845	0.0275	[0.0306, 0.1384]
$\ell \Delta y^2, \ell \Delta w^2, \ell \Delta yw$	0.0263	0.0775	0.0286	[0.0214, 0.1336]
$y_J^2 w_J, y_J w_J^2$	0.0252	0.0714	0.0289	[0.0148, 0.1280]
MLE estimates		0.123	0.0137	

critical values: $\widehat{c}_{1-\alpha}^{\text{inf}} = 0.9668$, $\widehat{c}_{1-\alpha}^{\text{sup}} = 2.4764$
$CI_{\text{inf}}^{\text{Robust}} = [0.0271, 0.1111]$ $CI_{\text{sup, cv}}^{\text{Robust}} = [0.0169, 0.0916]$

[1] y : non-labor income, w : marginal wage rates, ℓ: the end point of the segment in a piecewise linear budget set. $\ell^m \Delta y^p w^q$ denotes $\sum_j \ell_j^m (y_j^p w_j^q - y_{j+1}^p w_{j+1}^q)$.

[2] CV denotes cross-validation criteria defined in Blomquist and Newey (2002, p.2464).

Figure 4: Nonparametric Wage Elasticity of Hours of Work Estimates in Blomquist and Newey (Table 1, 2002).
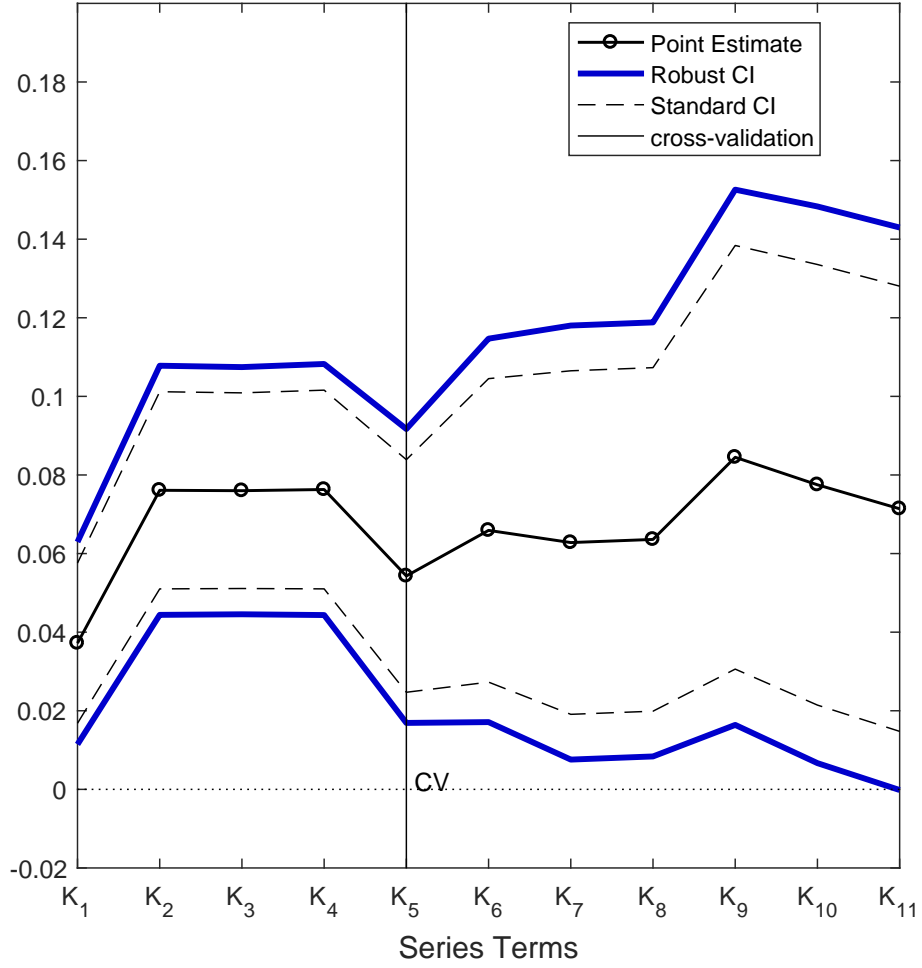
Figure plots wage elasticity estimates of the expected labor supply same as in Table 1, with standard pointwise 95% CIs as well as uniform (in $K_m \in \mathcal{K}_n$) CIs constructed by using critical value $\widehat{c}_{1-\alpha}^{\mathrm{sup}}$.