Inference in Nonparametric Series Estimation with Data-Dependent Undersmoothing

Byunghoon Kang*

Department of Economics, Lancaster University

First version December 9, 2014; Revised February 26, 2017

Abstract

Existing asymptotic theory for inference in nonparametric series estimation typically imposes an undersmoothing condition that the number of series terms is sufficiently large to make bias asymptotically negligible. However, there is no formally justified data-dependent method for this in practice. This paper constructs inference methods for nonparametric series regression models and introduces tests based on the infimum of t-statistics over different series terms. First, I provide a uniform asymptotic theory for the t-statistic process indexed by the number of series terms. Using this result, I show that test based on the infimum of the t-statistics and its asymptotic critical value controls asymptotic size with undersmoothing condition. Using this test, we can construct a valid confidence interval (CI) by test statistic inversion that has correct asymptotic coverage probability. Even when asymptotic bias terms are present without the undersmoothing condition, I show that CI based on the infimum of the t-statistics bounds coverage distortions. In an illustrative example, nonparametric estimation of wage elasticity of the expected labor supply from Blomquist and Newey (2002), proposed CI is close to or tighter than those based on the standard CI with possibly ad hoc choice of series terms.

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

JEL classification: C12, C14.

*This paper is a revised version of my first chapter in Ph.D. thesis at UW-Madison. I am deeply indebted to Bruce Hansen for his continuous guidance and suggestions. I am also grateful to Jack Porter, Xiaoxia Shi and Joachim Freyberger for thoughtful discussions. Thanks to Michal Kolesár, Denis Chetverikov, Yixiao Sun, Andres Santos, Patrik Guggenberger, Federico Bugni, Joris Pinkse, and Liangjun Su for helpful conversations and criticism. I acknowledge support by Kwanjeong Educational Foundation Graduate Research Fellowship and Leon Mears Dissertation Fellowship from UW-Madison. All errors are my own. Email: b.kang1@lancaster.ac.uk, Homepage: https://sites.google.com/site/davidbhkang

1 Introduction

Nonparametric series estimation has received attention in both theoretical econometrics and applied economics. I consider the following nonparametric regression model;

$$\begin{aligned} y_i &= g_0(x_i) + \varepsilon_i, \\ E(\varepsilon_i | x_i) &= 0 \end{aligned} \tag{1.1}$$

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. Examples falling into the model (1.1) include nonparametric estimation of the Mincer equation, gasoline demand, and labor supply function (see, among many others, Heckman, Lochner and Todd (2006), Hausman and Newey (1995), Blomquist and Newey (2002), Blundell and MaCurdy (1999), and references therein). Addressing potential misspecification of the parametric model, nonparametric series methods have several advantages, as they can easily impose shape restrictions such as additive separability or concavity, and implementation is easy because the estimation method is least squares. However, implementation in practice requires a choice of *the number of series terms*, K. Estimation and inference may largely depend on its choice in finite samples. Moreover, required K may vary with different data sets to accommodate the smoothness and nonlinearity of unknown function and different sample sizes, as well as whether the goal is estimation or inference.

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 to make bias asymptotically negligible relative to variance. The undersmoothing condition has been imposed, particularly for valid inference, in many nonparametric series methods both in theory and in practice, as there is no theory for bias-corrections available to date. Ignoring asymptotic bias with the undersmoothing assumption, one can apply the conventional confidence interval (CI) using the standard normal critical value with estimates and standard errors based on some choice of "sufficiently large" K. However, the asymptotic theory does not provide specific guidelines for choosing a "large" number of series terms to make bias small in practice. With given sample sizes n, some possibly ad hoc methods in practice select $\widehat{K} = \widetilde{K} \cdot n^{\gamma}$ with some pre-selected $\widetilde{K}$ and a specific rate of γ that satisfies the undersmoothing level. However, there is no formally justified data-dependent method to choose K that gives the desired level of undersmoothing in series regression literature.

Due to these unsatisfactory results for the inference procedure both in theory and practice, a specification search seems necessary, i.e., search over different series terms $K \in [\underline{K}, \bar{K}]$.

For example, a researcher may use quadratic, cubic, or quartic terms in polynomial regression, or try a different number of knots in regression spline to see how the estimate and standard error change. Moreover, some data-dependent selection rules that are valid for estimation (such as cross-validation or Akaike information criterion (AIC)) and some rule-of-thumb methods that are suggested for inference, also require evaluating estimates with different Ks. If researchers evaluate different specifications with different number of series terms and/or select one specific specification as a baseline model, it is not clear how this randomness affects standard inference.

In this paper, I construct inference methods in nonparametric series regression given the range of different series terms. I consider the testing problem for a regression function at a point and introduce tests based on *infimum of the studentized t-statistics* over different series terms. To describe intuition heuristically, we may decompose infimum t-statistic as follows

$$\inf_K |T_n(K)| \approx \inf_K |N(0,1) + \frac{Bias(K)}{S.E.(K)}|$$

where $T_n(K), Bias(K), S.E.(K)$ denote t-statistic, bias and standard error of the series estimator using K terms, respectively. The test based on infimum t-statistics and searching for small t-statistics have a similar motivation to the one on which the undersmoothing condition is theoretically based: using faster rates of K than the optimal MSE rate (using "large" K that has a small bias and large variance), so that make the second term, $\frac{Bias(K)}{S.E.(K)}$, small. Many papers in nonparametric series estimation literature typically suggested to increase the number of series terms and include additional terms than those cross-validation chooses, especially for inference (for example, see Newey (2013), Newey, Powell and Vella (2003)). I formally justify this conventional wisdom by introducing the infimum test statistic, and provide an inference method based on its asymptotic distribution, although I do not consider data-dependent methods that satisfy desired undersmoothing rates in this paper.

For this, I first provide a uniform asymptotic theory for the t-statistic process indexed by the number of series terms. Existing asymptotic normality of the t-statistic in the literature holds under a deterministic sequence of $K \to \infty$ as the sample size increases. The main contribution of this paper is to derive the asymptotic distribution theory for the entire sequences of t-statistics over a range of K.

Using this result, I show that test based on the infimum of the t-statistics and its asymptotic critical value control the asymptotic size (null rejection probability) with the undersmoothing condition for all Ks in a set. Allowing asymptotic bias without the undersmoothing condition, I also analyze the effect of bias on the asymptotic size of the test. Even when asymptotic bias terms are present, the test based on the infimum t-statistic bound the size distortions, in the sense that the asymptotic size is bounded above by the asymptotic size of a test with single t-statistic that has the smallest bias. The infimum t-statistic is less sensitive to the asymptotic bias; it naturally excludes small K with large bias and selects among some large Ks under the null.

I also construct a valid pointwise confidence interval for the true parameter that has nominal asymptotic coverage probability by test statistic inversion. The proposed CI based on infimum test statistic can be easily constructed using estimates and standard errors for the set of Ks. It is obtained as the union of all CIs by replacing the standard normal critical value with the critical value from the asymptotic distribution of the infimum t-statistic. We can approximate the asymptotic critical value using a simple Monte Carlo or weighted bootstrap method. I find that our proposed CI performs well in Monte Carlo experiments; coverage probability of the CI based on the infimum t-statistic is close to the nominal level in various simulation setups. I also find that proposed CI bounds the coverage distortions even when asymptotic bias is present similar to the asymptotic size results. As an illustrative example, I revisit nonparametric estimation of wage elasticity of the expected labor supply, as in Blomquist and Newey (2002).

As a by-product of the joint asymptotic distribution results, this paper also provides a valid CI after selecting the number of series terms. By adjusting the conventional normal critical value to the critical value from supremum of the t-statistics over all series terms, this gives a valid post-selection CI that has a correct coverage with any choice of $\widehat{K}$ among some ranges. By enlarging the CI with critical values larger than the normal critical value, this post-selection CI can accommodate bias, although it does not explicitly deal with bias problems. I expect this lead to a tighter CI than those based on the Bonferroni-type critical value, as I incorporate the dependence structure of the t-statistics from our asymptotic distribution theory.

I also investigate inference methods in partially linear model setup. Focusing on the common parametric part, choice problems also occur for the number of approximating terms or the number of covariates in estimating the nonparametric part. Unlike the nonparametric object of interest that has a slower convergence than $n^{1/2}$ rate (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, in which K increases much slower than the sample size n. To fully account for dependency of the t-statistics with the different sequences of Ks in the partially linear model setup, this requires a different approximation theory than standard first order approximation results. Using the recent results of Cattaneo, Jansson, and Newey (2015a), I develop a joint asymptotic distribution of the studentized t-statistics over a different number of series terms. By focusing on the faster rate of K that grows as fast as the sample size n and using larger variance than the standard variance formula, we are able to account for the dependency of t-statistics with different Ks. I also propose methods to construct CIs that are similar to the nonparametric regression setup and provide their asymptotic coverage properties. Potential empirical applications include, but are not limited to, estimation of the treatment effect model with series approximations.

1.1 Related literature

The literature on nonparametric series estimation is vast, but data-dependent series term selection and its impact on estimation or inference is comparatively less developed. Perhaps the most widely used data-dependent rule in practice is cross-validation. Asymptotic optimality results have been developed (see, for example, Li (1987), Andrews (1991b), Hansen (2015)) in terms of asymptotic equivalence between integrated mean squared error (IMSE) of the nonparametric estimator with $\widehat{K}_{\texttt{cv}}$ selected by minimizing the cross-validation criterion and IMSE of the infeasible optimal estimator. However, there are two problems with cross-validation selected $\widehat{K}_{\texttt{cv}}$ for the valid inference. First, it is asymptotically equivalent to selecting K to minimize IMSE, and thus it does not satisfy the undersmoothing condition needed for asymptotic normality without bias terms. Therefore, a t-statistic based on $\widehat{K}_{\texttt{cv}}$ will be asymptotically invalid. Second, $\widehat{K}_{\texttt{cv}}$ selected by cross-validation will itself be random and not deterministic. Thus, it is not clear whether the t-statistic based on $\widehat{K}_{\texttt{cv}}$ has a standard asymptotic normal distribution which is derived under a deterministic sequence of K.

Important recent papers by Horowitz (2014), Chen and Christensen (2015a) develop the state-of-the-art data-dependent methods in the nonparametric instrumental variables (NPIV) estimation (see also other references therein). They develop data-driven methods for choosing sieve dimension in that resulting NPIV estimators attain the optimal sup-norm or L^2 norm rates adaptive to the unknown smoothness of g_0. In this paper, we focus on the inference problem rather than estimation with the similar issues arising from cross-validation.

Moreover, this paper is also closely related to the previous methods that conceptually require increasing K until t-statistic is "small enough". For example, among many others, Newey (2013) suggested increasing K until standard errors are large relative to small changes in objects of interest, and Horowitz and Lee (2012) suggested increasing K until variance suddenly increases. They discuss these methods work well in practice and simulation for the inference. Using similar ideas, we account the randomness introduced in the first step specification search by providing formal inference methods based on asymptotic distribution results of the infimum test statistic.

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, Chernozhukov, Chetverikov, and Kato (2015), and Chen and Christensen (2015b), among many others. Under i.i.d. or weakly dependent data, they focused on Sup/L^2-norm convergence rates, asymptotic normality of series estimators, and pointwise/uniform inference on linear/nonlinear functionals under a deterministic sequence of K. 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, and focuses on a pointwise inference on irregular (i.e., slower than $n^{1/2}$ rate) and linear functional of $g_0(x)$ under i.i.d. data.

For the kernel-based density or regression estimation, the data-dependent bandwidth selection problem is well known. Several rule-of-thumb methods and plug-in optimal bandwidths have been proposed (see Härdle and Linton (1994), Li and Racine (2007) for references). Recent paper by Calonico, Cattaneo and Farrell (2015) compared higher-order coverage properties of undersmoothing and explicit bias-corrections, and derived coverage optimal bandwidth choices in kernel estimation. See also Hall and Horowitz (2013), Schennach (2015) and references therein for various recent work on related bias issues and nonparametric inference for the kernel estimator. Unlike the kernel-based methods, little is known about the statistical properties of data-dependent selection rules (e.g., rates of $\widehat{K}_{\texttt{cv}}$) in series estimation and asymptotic distribution of nonparametric estimators with data-dependent methods. In general, the main technical difficulty arises from the lack of an explicit asymptotic bias formula for the series estimator (see Zhou, Shen, and Wolfe (1998) and Huang (2003b) for exceptions with some specific sieves). Thus, it is difficult to derive an asymptotic theory for the bias-correction or some plug-in formula compare with kernel estimation.

A novel recent paper that is concurrent with this paper, Armstrong and Kolesár (2015) considered inference methods in kernel estimation with bandwidth snooping. Focusing on the supremum of the t-statistics over the bandwidths, they developed confidence intervals that are uniform in bandwidths. Considering supremum statistic is motivated by the sensitivity analysis as a usual correction for the multiple testing problem. Moreover, considering different bandwidths and the test based on the supremum of the studentized t-statistics has been used to achieve adaptive inference procedures when smoothness of the function is unknown (See Horowitz and Spokoiny (2001), and also Armstrong (2015)). Although this paper has analogous results with Armstrong and Kolesár (2015) in series estimation considering supremum of the t-statistics (see Appendix C), the main focus of this paper are asymptotic bias and the undersmoothing condition and their effect on the coverage of CIs, which may be crucial in series estimation.[1]

The outline of the paper is as follows. I first introduce basic nonparametric series regression setup in Section 2. In Section 3, I provide an empirical process theory for the t-statistic sequences over a set. Section 4 introduces infimum of the t-statistic and describes the asymptotic null distributions of the test statistic. Then, I provide the asymptotic size results of the test and implementation procedure for the critical value. Section 5 introduces CIs based on the infimum test statistic and provides their coverage properties. Section 6 analyzes valid post-model selection inference in this setup. Section 7 extends our inference methods to the partially linear model setup. Section 8 includes Monte Carlo experiments in various setups. Section 9 illustrates proposed inference methods using the nonparametric estimation of wage elasticity of the expected labor supply, as in Blomquist and Newey (2002), then Section 10 concludes. Appendix A and B include all proofs, figures and tables. Appendix C discuss inference procedures based on the supremum of the t-statistics.

1.2 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 a 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 the bounded real valued functions f on the support $\mathcal{X}$. Let also $\mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$, $\mathbb{R}_{+,\infty} = \mathbb{R}_+ \cup \{+\infty\}$, $\mathbb{R}_{[+\infty]} = \mathbb{R} \cup \{+\infty\}$ and $\mathbb{R}_{[\pm\infty]} = \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$.

[1] We may also consider other types of test statistics that is robust to the bias (for example, "median" of the t-statistics). Any types of test statistics that are continuous transformation of joint t-statistics with its appropriate critical value leads to the tests that control the asymptotic size with undersmoothing.

2 Nonparametric series regression

I first introduce the nonparametric series regression setup in 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 pointwise inference in x and I will omit the dependence on x if there is no confusion.

We consider sequence of approximating model indexed by number of series terms $K \equiv K(n)$. Let $\widehat{g}_K(x)$ be an estimator of $g_0(x)$ using the first K vectors of approximating functions $P_K(x) = (p_1(x), \cdots, p_K(x))'$ 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. Basis functions may come from set of large number of potential regressors and/or their nonlinear transformations.

Series estimator $\widehat{g}_K(x)$ is then obtained by standard least square (LS) estimation of y_i on regressors P_{Ki}

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

where $P_{Ki} \equiv P_K(x_i) = (p_1(x_i), p_2(x_i), \cdots, p_K(x_i))', P^K = [P_{K1}, \cdots, P_{Kn}]', Y = (y_1, \cdots y_n)'$. We can think of $\widehat{g}_K(x)$ as an estimator of the best linear L^2 approximation for $g_0(x)$, i.e., $P_K(x)'\beta_K$ where β_K can be defined as the best linear projection coefficients $\beta_K \equiv (E[P_{Ki}P_{Ki}'])^{-1}E[P_{Ki}y_i]$. For some $x \in \mathcal{X}$, define the approximation error using K series terms as $r_K(x) = g_0(x) - P_K(x)'\beta_K$. Also define $r_{Ki} \equiv r_K(x_i)$, $p_i \equiv p(x_i) = (p_{1i}, p_{2i}, \cdots,)'$. We can write the model using K approximating terms as the following projection model

$$y_i = P_{Ki}'\beta_K + \varepsilon_{Ki}, \quad E[P_{Ki}\varepsilon_{Ki}] = 0 \tag{2.2}$$

where $\varepsilon_{Ki} \equiv r_{Ki} + \varepsilon_i$.

For simplicity of notation, I define the true regression function at a point as $\theta_0 \equiv g_0(x)$. Let $\widehat{\theta}_K \equiv \widehat{g}_K(x)$ and $\theta_K \equiv P_K(x)'\beta_K$. Define the asymptotic variance

$$\begin{aligned} &V_K \equiv V_K(x) = P_K(x)'Q_K^{-1}\Omega_K Q_K^{-1}P_K(x), \\ &Q_K = E(P_{Ki}P_{Ki}'), \quad \Omega_K = E(P_{Ki}P_{Ki}'\varepsilon_i^2) \end{aligned} \tag{2.3}$$

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

To account specification search in nonparametric regression model, we use notion of testing setup. We consider two-sided testing for θ

$$H_0 : \theta = \theta_0, \quad H_1 : \theta \neq \theta_0. \tag{2.4}$$

The studentized t-statistic for H_0 is

$$T_n(K, \theta_0) \equiv \frac{\sqrt{n}(\widehat{g}_K(x) - g_0(x))}{V_K^{1/2}} = \frac{\sqrt{n}(\widehat{\theta}_K - \theta_0)}{V_K^{1/2}}. \tag{2.5}$$

Under standard regularity conditions (will be discussed in Section 3) including an undersmoothing rate for *deterministic sequence* $K \to \infty$ as $n \to \infty$, the asymptotic distribution of the t-statistic is well known

$$T_n(K, \theta_0) \xrightarrow{d} N(0, 1). \tag{2.6}$$

See, for example, Andrews (1991a), Newey (1997), Belloni et al. (2015), Chen and Christensen (2015b) among many others. In the next section, I formally develop an asymptotic distribution theory of $T_n(K, \theta_0)$ over a set $\mathcal{K}_n$.

3 Asymptotic distribution of the joint t-statistics

3.1 Weak convergence of t-statistic process

In this section, I provide asymptotic distribution theory of the joint t-statistics over a set. First, I introduce following set $\mathcal{K}_n$ to construct empirical process theory of the t-statistics over $K \in \mathcal{K}_n$ that can be indexed by the continuous parameter π, which is a 'fraction' of the largest series terms $\bar{K}$.

Assumption 3.1. *(Set of number of series terms) Let $\mathcal{K}_n$ as*

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

where $K \equiv \lfloor \pi \bar{K} \rfloor$ for $\pi \in \Pi = [\underline{\pi}, 1]$, $\underline{\pi} > 0$.

The standard inference methods in this nonparametric regression setup typically consider singleton set $\mathcal{K}_n = \{K\}$. Assumption 3.1 considers range of number of series terms and considers infinite sequence of models indexed by $\pi \in \Pi$ using $K = \lfloor \pi \bar{K} \rfloor$ series terms. Note that $\mathcal{K}_n$ is indexed by sample size n, as I will impose rate conditions for the largest $\bar{K} \equiv \bar{K}(n)$ in the next Assumption 3.2. Together with the Assumption 3.2 below, set $\mathcal{K}_n$ in Assumption

3.1 considers the sequence of models that has the same rate of K, i.e., $K \asymp K'$ for any $K, K' \in \mathcal{K}_n$.

Under Assumption 3.1, I define the *t-statistic process*, $T_n^*(\pi, \theta)$, as

$$T_n^*(\pi, \theta) \equiv T_n(\lfloor \pi \bar{K} \rfloor, \theta) \quad \pi \in \Pi, \tag{3.1}$$

where $T_n(K, \theta)$ is defined in (2.5). Note that $T_n^*(\pi, \theta)$ is a t-statistic using $K = \lfloor \pi \bar{K} \rfloor$ number of series terms and is a step function of π.

In addition, I impose mild regularity conditions that are standard in nonparametric series regression literature and are satisfied by well-known basis functions. I closely follow assumptions in the recent paper by Belloni et al. (2015), Chen and Christensen (2015b) and impose rate conditions of K uniformly over $\mathcal{K}_n$. Other standard regularity conditions in the literature (e.g., Newey (1997)) can also be imposed here with different rate conditions of K.

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 3.2. *(Regularity conditions)*

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

(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 3.2(ii) imposes moment conditions and standard uniform integrability conditions. ζ_K, c_K, ℓ_K in Assumption 3.2(iii)-(v) are satisfied with various basis functions. For example, if the support $\mathcal{X}$ is a cartesian product of compact connected intervals (e.g. $\mathcal{X} = [0, 1]^{d_x}$), 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. c_K and ℓ_K in Assumption 3.2(iii) vary with different basis and can be replaced by series specific bounds. For example, if $g_0(x)$ belongs to the Hölder space of smoothness p, then $c_K \lesssim K^{-p/d_x}, \ell_K \lesssim K$ for power series, $c_K \lesssim K^{-(p \wedge s_0)/d_x}, \ell_K \lesssim 1$ for spline and wavelet series of order s_0 (see Newey (1997), Chen

(2007), Belloni et al. (2015), and Chen and Christensen (2015b) for more discussions on c_K, ℓ_K, ζ_K with various sieve bases).

When the probability density function of x_i is uniformly bounded above and bounded away from zero over compact support $\mathcal{X}$ and orthonormal basis is used, then we have $\lambda_K \lesssim 1$ (see, for example, Proposition 2.1 in Belloni et al. (2015) and Remark 2.2 in Chen and Christensen (2015b)). The rate conditions in Assumption 3.2(v) can be replaced by the specific bounds of ζ_K, c_K, ℓ_K. For example, for the power series, Assumption 3.2(v) reduced to $\sup_{K\in\mathcal{K}_n}\sqrt{K^2(\log K)/n}(1+K^{3/2-p/d_x})+K^{1-p/d_x} = \sqrt{\bar{K}^2(\log \bar{K})/n}(1+\bar{K}^{3/2-p/d_x})+\bar{K}^{1-p/d_x} \to 0$ with the Assumption 3.1.

For notational simplicity, it is convenient to define $P_\pi(x) \equiv P_{\lfloor \bar{K}\pi \rfloor}(x)$, $P_{\pi i} \equiv P_\pi(x_i) = P_{\lfloor \bar{K}\pi \rfloor i}$ and $r_\pi \equiv r_\pi(x) = r_{\lfloor \bar{K}\pi \rfloor}(x)$. Asymptotic variance can be defined as $V_\pi \equiv V_\pi(x) = ||\Omega_\pi^{1/2} Q_\pi^{-1} P_\pi(x)||^2$, where $\Omega_\pi = E(P_{\pi i}P_{\pi i}'\varepsilon_i^2), Q_\pi = E(P_{\pi i}P_{\pi i}')$. Under Assumptions 3.1 and 3.2, the t-statistic process under H_0 can be decomposed as follows

$$T_n^*(\pi,\theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{P_\pi(x)'P_{\pi i}\varepsilon_i}{V_\pi^{1/2}} - \sqrt{n}V_\pi^{-1/2}r_\pi + o_p(1), \quad \pi \in \Pi \tag{3.2}$$

where $\sqrt{n}V_\pi^{-1/2}r_\pi$ is a bias term due to approximation errors. I define the asymptotic bias for the sequence of models indexed by π as the limit of the second term

$$\nu(\pi) \equiv \lim_{n\to\infty} -\sqrt{n}V_\pi^{-1/2}r_\pi. \tag{3.3}$$

Under the following undersmoothing condition, asymptotic bias $\nu(\pi)$ is 0 for all $\pi \in \Pi$. To assess the effect of bias on inference, we will consider distinction between results imposing undersmoothing condition or not.

Assumption 3.3. *(Undersmoothing)* $\sup_{K\in\mathcal{K}_n} |\sqrt{n}V_K^{-1/2}\ell_K c_K| \to 0$ *as* $n \to \infty$.

A sufficient condition for Assumption 3.3 is $\sup_{K\in\mathcal{K}_n} \sqrt{n}V_K^{-1/2}K^{-p/d_x} = o(1)$ when we use explicit bounds $c_K\ell_K \lesssim K^{-p/d_x}$ for spline or wavelet series. When we consider a point $x \in \mathcal{X}$ where standard error $V_K^{1/2} \propto K^{1/2}$, Assumption 3.1 and 3.3 together imply that Assumption 3.3 is provided by $\sqrt{n}\bar{K}^{1/2-p/d_x} \to 0$ for power series.

Next theorem is our first main result which provides uniform central limit theorem of the t-statistic process for nonparametric LS series estimation.

Theorem 3.1. *Under Assumptions 3.1, 3.2 and* $\sup_\pi |\nu(\pi)| < \infty$*,*

$$T_n^*(\pi, \theta_0) \Rightarrow \mathbb{T}(\pi) + \nu(\pi) \tag{3.4}$$

where $\mathbb{T}(\pi)$ *is a mean zero Gaussian process on* $\ell^\infty(\Pi)$ *with covariance function* $\Sigma(\pi_1, \pi_2) = \lim_{n\to\infty} \Sigma_n(\pi_1, \pi_2)$*, where*

$$\Sigma_n(\pi_1, \pi_2) = \frac{P_{\pi_1}(x)' E(P_{\pi_1 i} P'_{\pi_2 i} \varepsilon_i^2) P_{\pi_2}(x)}{V_{\pi_1 \wedge \pi_2}^{1/2} V_{\pi_1 \vee \pi_2}^{1/2}} \tag{3.5}$$

for any $\pi_1, \pi_2 \in \Pi$*, and* $\nu(\pi)$ *is defined in* (3.3)*. In addition, if Assumption 3.3 is satisfied, then*

$$T_n^*(\pi, \theta_0) \Rightarrow \mathbb{T}(\pi), \quad \pi \in \Pi. \tag{3.6}$$

Theorem 3.1 provides weak convergence of the t-statistic process $T_n^*(\pi, \theta_0), \pi \in \Pi$. This is an asymptotic theory for the entire sequence of t-statistics $T_n(K, \theta_0), K \in \mathcal{K}_n$. The asymptotic null distribution of the t-statistic process in (3.4) is equal to a mean zero Gaussian process $\mathbb{T}(\pi)$ plus the asymptotic bias $\nu(\pi)$.

Under conditional homoskedasticity, $E(\varepsilon_i^2 | x_i = x) = \sigma^2$, the covariance function of the limiting Gaussian process reduces to the simple form

$$\Sigma(\pi_1, \pi_2) = \lim_{n\to\infty} \frac{V_{\pi_1 \wedge \pi_2}^{1/2}}{V_{\pi_1 \vee \pi_2}^{1/2}} \tag{3.7}$$

for any $\pi_1, \pi_2 \in \Pi$. For example, if we further assume $V_\pi^{1/2} \propto (\bar{K}\pi)^{1/2}$, then $\Sigma(\pi_1, \pi_2) \propto (\pi_1/\pi_2)^{1/2}$.

Remark 3.1 (Rate conditions)**.** Note that the asymptotic bias $|\nu(\pi)| = 0$ if $\bar{K}$ increases faster than the optimal MSE rate (undersmoothing). $0 < |\nu(\pi)| < \infty$ if $\bar{K}$ increases at the optimal MSE rate, and $|\nu(\pi)| = \infty$ if $\bar{K}$ increases slower than the optimal MSE rate (oversmoothing). Theorem 3.1 does not allow oversmoothing rates as we require $\sup_\pi |\nu(\pi)| < \infty$. Assumption 3.1 does not consider all different sequences of K satisfying asymptotic normality of series estimators, however, these are the class of sequences to be able to provide uniform central limit theorem of the t-statistic process. As studentized t-statistic is normalized by variance terms V_K which may increase differently with different rates of K, two t-statistics with different rates of K can be asymptotically independent, thus hard to incorporate dependency (see also discussions in Section 3.2 with alternative $\mathcal{K}_n$ allowing different rates of

K).

Remark 3.2 (Other functionals)**.** Here, I focus on the leading example, where $\theta_0 = g_0(x)$ for some fixed point $x \in \mathcal{X}$, but I may 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 with estimators $\widehat{\theta} = a(\widehat{g}_K(x)) = a_K(x)'\widehat{\beta}_K$ and appropriate transformation of basis $a_K(x) = (a(p_1(x), \cdots, a(p_K(x)))'$. 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 analogous results for the partially linear model in Section 7.

3.2 Alternative set with different rates

Next, we provide different approximations to the sequence of t-statistics with an alternative set $\mathcal{K}_n$ constructed to allow different rates of Ks. Alternative set assumption allows for optimal mean squared error rates of K as well as oversmoothing rates which increases slower than the optimal rate.

Assumption 3.4. *(Alternative set with different rates) Let $\mathcal{K}_n$ as*

> $\mathcal{K}_n = \{\underline{K} = K_1, \cdots, K_m, \cdots, \bar{K} = K_M\}$ *where* $K_m \equiv \tau n^{\phi_m}$ *for constant* $\tau > 0$, $0 < \phi_1 < \phi_2 < \cdots < \phi_M$, *and fixed* M. *Define asymptotic bias for the sequence of models as* $\nu(m) \equiv - \lim_{n\to\infty} \sqrt{n} V_{K_m}^{-1/2} r_{K_m}$. *Assume that the largest model* $\bar{K}$ *satisfies* $\sqrt{n} V_{\bar{K}}^{-1/2} \ell_{\bar{K}} c_{\bar{K}} \to 0$ *as* $n \to \infty$.

Assumption 3.4 can consider rates of K from oversmoothing ($|\nu(m)| = \infty$) to undersmoothing ($|\nu(m)| = 0$) with different ϕ_m. Here, $\underline{K}$ can increase slower than the optimal MSE rates and $\bar{K}$ satisfies undersmoothing rates. As this alternative set assumption allows slower than optimal MSE rates, this condition considers broader range of Ks than the Assumption 3.1. Together with Assumption 3.2(v), there exists rate restrictions on ϕ_m uniformly over m. Undersmoothing assumption for the $\bar{K}$, i.e. $\nu(M) = 0$, can be restrictive. However, this is merely a modeling device considering broad range of K and taking some large enough $\bar{K}$ so that satisfy undersmoothing.

Under Assumption 3.4, if we impose $\sup_m |\nu(m)| < \infty$, then the joint t-statistics converge in distribution to a normal distribution plus the asymptotic bias terms similar to Theorem 3.1. However, joint t-statistics do not converge in distribution to a bounded random vector if any of the elements $|\nu(m)| = \infty$ with oversmoothing sequences. If $\nu(m) = \pm\infty$ for some m, then it can be shown that 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 that is some continuous transformation of the joint t-statistics because continuous mapping theorem cannot be directly applied.

To obtain the asymptotic distribution even 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). 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 3.2. *Under Assumptions 3.2 and 3.4, following holds for any continuous function $S(t)$ at all $t \in \mathbb{R}^{M-1}_{[\pm\infty]} \times \mathbb{R}$,*

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

where $T_n(\theta) = (T_n(K_1, \theta), \cdots, T_n(K_M, \theta))', Z = (Z_1, \cdots, Z_M)' \sim N(0, \Sigma)$, $\nu = (\nu(1), \cdots, \nu(M))'$ are $M \times 1$ vectors, $\Sigma_{jl} = \lim_{n\to\infty} \Sigma_{jl,n}$, and $\Sigma_{jl,n} = \frac{P_{K_j}(x)' E(P_{K_j i} P'_{K_l i} \varepsilon_i^2) P_{K_l}(x)}{V_{K_j}^{1/2} V_{K_l}^{1/2}}$. If $\nu(m) = \pm\infty$, then the corresponding element of $Z + \nu$ equals $\pm\infty$.

Note that we do not require either Assumption 3.3 (undersmoothing) or $\sup_m |\nu(m)| < \infty$ in Theorem 3.2. Variance-covariance matrix Σ is similarly defined as in Theorem 3.1. Moreover, if $V_K^{1/2} \asymp \sqrt{K}$ at some point x, then for any $j < l$, $\Sigma_{jl,n} \leq C \frac{V_{K_j}^{1/2}}{V_{K_l}^{1/2}}$ for some constant $C > 0$ by Assumption 3.2(ii) and the latter term converges to 0 as $n \to 0$ by Assumption 3.4, thus $\Sigma_{jl,n} \to 0$, and $\Sigma = I_M$.

Remark 3.3 (Rate conditions (continued))**.** Note that Assumption 3.4 only considers finite K sequences, i.e., $|\mathcal{K}_n| = M$. In finite samples, we only consider finite set $\mathcal{K}_n$, so the difference between Assumption 3.1 and 3.4 only matters in large samples. Assumption 3.4 is useful to consider the effect of bias on inference problems allowing broader range of K, i.e., oversmoothing rates (see Section 4 for formal results). On the other hand, Assumption 3.1 only consider sequences of K with the same growth rates which only differ in constant π. Thus, Theorem 3.1 gives the joint asymptotic distribution of t-statistics that has either zero bias for all $K \in \mathcal{K}_n$ or non-zero bounded bias for all $K \in \mathcal{K}_n$.

4 Test statistic

In this section, I introduce an *infimum* test statistic and analyze its asymptotic null distribution based on Theorem 3.1 and 3.2. Then, I provide the asymptotic size result of the tests, and methods to obtain critical values for our inference procedures.

I consider following test statistic

$$\operatorname{Inf} T_n(\theta) \equiv \inf_{K \in \mathcal{K}_n} |T_n(K, \theta)|. \tag{4.1}$$

As I denoted in the introduction, there are several reasons to consider $\operatorname{Inf} T_n(\theta)$ in series regression context. First of all, small t-statistic centered at the true value corresponds to the approximation with certain choice of series terms that has a small bias and large variance, which is good for the coverage as what undersmoothing assumption does for eliminating asymptotic bias, theoretically. This is also closely related to some rule-of-thumb methods suggested by several papers to choose undersmoothed K (see, for example, Newey (2013), Newey, Powell and Vella (2003)).

4.1 Asymptotic distribution of the test statistic

Asymptotic null limiting distribution of the infimum test statistic follows immediately from Theorem 3.1 and 3.2.

Corollary 4.1.
1. *Under Assumptions 3.1, 3.2 and* $\sup_\pi |\nu(\pi)| < \infty$, $\operatorname{Inf} T_n(\theta_0) \xrightarrow{d} \inf_{\pi \in [\underline{\pi}, 1]} |\mathbb{T}(\pi) + \nu(\pi)|$, *where* $\mathbb{T}(\pi)$ *is the mean zero Gaussian process defined in Theorem 3.1. In addition, if Assumption 3.3 holds, then* $\operatorname{Inf} T_n(\theta_0) \xrightarrow{d} \xi_{\inf} \equiv \inf_{\pi \in [\underline{\pi}, 1]} |\mathbb{T}(\pi)|$.
2. *Under Assumptions 3.2 and 3.4,* $\operatorname{Inf} T_n(\theta_0) \xrightarrow{d} \inf_{m=1,\cdots,M} |Z_m + \nu(m)|$, *where* Z_m *is an element of* $M \times 1$ *normal vector* $Z \sim N(0, \Sigma)$ *and* $\nu = (\nu(1), \cdots, \nu(M))'$ *is defined in Theorem 3.2.*

Corollary 4.1.1 derives the asymptotic null limiting distribution of $\operatorname{Inf} T_n(\theta)$ under $\mathcal{K}_n$ with same rates of K (Assumption 3.1) and Corollary 4.1.2 provides the asymptotic distribution under alternative $\mathcal{K}_n$ with different rates of K (Assumption 3.4).

Whether some asymptotic bias $|\nu(m)|$ are unbounded or not, Corollary 4.1.2 shows that $\operatorname{Inf} T_n(\theta_0)$ converge in distribution to the bounded random variable. Under H_0, $\operatorname{Inf} T_n(\theta)$ exclude all small Ks corresponding to oversmoothing (where the bias is of larger order than the standard error) and select among large Ks with optimal MSE rates and undersmoothing rates (where the bias is of smaller order), asymptotically. Using this Corollary, I discuss the effect of asymptotic bias on the inference in Section 4.2 (for the size results) and Section 5 (for the coverage results).

4.2 Asymptotic size of the test

I start by defining critical value $c^{\inf}_{1-\alpha}$ as $(1-\alpha)$ quantile of the asymptotic null distribution $\xi_{\inf} = \inf_{\pi\in[\underline{\pi},1]} |\mathbb{T}(\pi)|$ in Corollary 4.1.1, i.e., solves

$$P(\xi_{\inf} > c^{\inf}_{1-\alpha}) = \alpha \tag{4.2}$$

for $0 < \alpha < 1/2$.[2] The asymptotic null distribution, $\xi_{\inf}$, can be completely defined by covariance kernel of the limiting Gaussian process $\mathbb{T}(\pi)$ in Theorem 3.1. Since the limiting process can not be written as some transformation of Brownian motion process, the asymptotic critical value cannot be tabulated, in general. However, critical value can be obtained by standard Monte Carlo method or by the weighted bootstrap method. I will discuss approximation of the critical value in Section 4.3. With abuse of notation, I also use $c^{\inf}_{1-\alpha}$ as $(1-\alpha)$ quantile of the $\inf_{m=1,\cdots,M} |Z_m|$ if Corollary 4.1.2 is used under alternative set Assumption 3.4.

Next, we define $z_{1-\alpha/2}$ as $(1-\alpha/2)$ quantile of standard normal distribution function, which solves $P(|Z| > z_{1-\alpha/2}) = \alpha$ where $Z \sim N(0,1)$. Next Corollary provides the asymptotic size of the tests based on $\text{Inf}\, T_n(\theta)$ follow from the Corollary 4.1.

Corollary 4.2. 1. *Under Assumptions 3.1, 3.2 and 3.3, following holds with critical value $c^{\inf}_{1-\alpha}$ defined in* (4.2) *and the normal critical value $z_{1-\alpha/2}$,*

$$\limsup_{n\to\infty} P(\text{Inf}\, T_n(\theta_0) > c^{\inf}_{1-\alpha}) = \alpha, \qquad \limsup_{n\to\infty} P(\text{Inf}\, T_n(\theta_0) > z_{1-\alpha/2}) \leq \alpha. \tag{4.3}$$

2. *Suppose Assumptions 3.1 and 3.2 hold. If $\sup_\pi |\nu_\pi| < \infty$, then following inequality holds*

$$\limsup_{n\to\infty} P(\text{Inf}\, T_n(\theta_0) > c^{\inf}_{1-\alpha}) \leq F(c^{\inf}_{1-\alpha}, \inf_\pi |\nu(\pi)|), \tag{4.4}$$

$$\limsup_{n\to\infty} P(\text{Inf}\, T_n(\theta_0) > z_{1-\alpha/2}) \leq F(z_{1-\alpha/2}, \inf_\pi |\nu(\pi)|), \tag{4.5}$$

where $F(c, |\nu|) = 1 - \Phi(c - |\nu|) + \Phi(-c - |\nu|)$ with standard normal cumulative distribution function $\Phi(\cdot)$.

[2] Without imposing the undersmoothing assumption, asymptotic distribution of $\text{Inf}\, T_n(\theta_0)$ in Corollary 4.1.1 also depend on asymptotic bias $\nu(\pi)$ as well. If $\nu(\pi)$ can be replaced by some estimates $\widehat{\nu}(\pi)$, then the critical value from $\inf_{\pi\in\Pi} |\mathbb{T}(\pi) + \widehat{\nu}(\pi)|$ can be used. We do not pursue this approach as it is a difficult problem beyond the scope of this paper.

3. Under Assumptions 3.2 and 3.4, following holds

$$\limsup_{n\to\infty} P(\operatorname{Inf} T_n(\theta_0) > c^{\inf}_{1-\alpha}) \leq F(c^{\inf}_{1-\alpha}, 0), \tag{4.6}$$

$$\limsup_{n\to\infty} P(\operatorname{Inf} T_n(\theta_0) > z_{1-\alpha/2}) \leq \alpha. \tag{4.7}$$

Corollary 4.2.1 shows that the tests based on the infimum test statistic asymptotically control size with all the sequences $K \in \mathcal{K}_n$ satisfy the undersmoothing condition. As $\operatorname{Inf} T_n(\theta_0) \leq |T_n(K,\theta_0)|$ and $|T_n(K,\theta_0)| \xrightarrow{d} |N(0,1)|$ for any single $K \in \mathcal{K}_n$, the test based on $\operatorname{Inf} T_n(\theta)$ using normal critical value also controls the asymptotic size, but conservative.

Without undersmoothing assumption, Corollary 4.2.2 derives the upper bounds of the asymptotic null rejection probability of the tests based on $\operatorname{Inf} T_n(\theta)$. Equations (4.4) and (4.5) show that the asymptotic size is bounded above by the asymptotic size of a single t-statistic with the smallest asymptotic bias $\inf_\pi |\nu(\pi)|$. Note that $F(c, |\nu|)$ is a monotone decreasing in c and increasing in $|\nu|$. Typically $c^{\inf}_{1-\alpha}$ is strictly less than $z_{1-\alpha/2}$, so that $F(z_{1-\alpha/2}, 0) = \alpha < F(c^{\inf}_{1-\alpha}, 0)$. Moreover, upper bounds of the size can be small if the smallest bias $\inf_\pi |\nu(\pi)|$ is small. For example, when $c^{\inf}_{1-\alpha} = 1.5$, $F(c^{\inf}_{1-\alpha}, |\nu|) = 0.31$ for $|\nu| = 1$ and $F(c^{\inf}_{1-\alpha}, |\nu|) = 0.13$ for $|\nu| = 0$. (4.5) also show that the test based on $\operatorname{Inf} T_n(\theta_0)$ with normal critical value controls size asymptotically if the smallest bias is 0, i.e., $\inf_\pi |\nu(\pi)| = 0$ (see also Hall and Horowitz (2013), Hansen (2014) for the similar function and Figure 2 for the plots of $F(\cdot,\cdot)$ as a function of $|\nu|$ with some different c).

Corollary 4.2.3 shows similar asymptotic size results under the alternative set assumption. Furthermore, this can give useful information about the effect of bias on the asymptotic size by allowing 'large' bias $|\nu(m)| = \infty$ ('small' K_m). If asymptotic variance of joint distribution $\Sigma = I_M$ under Theorem 3.2 (refer to discussions below Theorem 3.2), we can get asymptotic size of the test as $\prod_{m=1}^{M} F(c^{\inf}_{1-\alpha}, |\nu(m)|)$. Thus, the asymptotic size is not affected by K_m such that $|\nu(m)| = \infty$ as $F(c, \infty) = 1$ for any constant $c > 0$. Further, suppose that the last M_1 number of Ks satisfy undersmoothing conditions and the others satisfy oversmoothing rates, i.e., $|\nu(m)| = \infty$ for $m = 1, \cdots, M - M_1$ and $|\nu(m)| = 0$ for the others. Then, the asymptotic size is equal to $\alpha^{M_1/M}$, as $c^{\inf}_{1-\alpha} = z_{1-\alpha^{1/M}/2}$ follows from Theorem 3.2 and $\Sigma = I_M$. In this special case, the asymptotic size is a decreasing function of the fraction of number of undersmoothing sequences M_1/M, and is equal to α when $|\nu(m)| = 0$ for all m, similar to Corollary 4.2.1.

Remark 4.1 (Largest K)**.** Note that the asymptotic size result in (4.7) relies on the inequality $\operatorname{Inf} T_n(\theta_0) \leq |T_n(\bar{K}, \theta_0)|$ and the fact that $T_n(\bar{K}, \theta_0) \xrightarrow{d} N(0,1)$ under Assumption 3.4. But, our theory still provides the bound of the asymptotic size in (4.7) without any undersmoothing conditions on $K \in \mathcal{K}_n$, as $F(z_{1-\alpha/2}, \inf_m |\nu(m)|)$. Asymptotic distribution result in Corollary 4.1.2 is still valid, as long as at least one $|\nu(m)|$ is bounded.

If we know (a priori) that $\bar{K}$ satisfies undersmoothing condition and others not, then there's no point of searching over different K; we may just use $\bar{K}$ for the inference. This may work well if $\bar{K}$ coincides with some infeasible size-optimal sequence $K^*(n)$ that minimizes $|P(T_n(K,\theta_0) > z_{1-\alpha/2}) - \alpha|$. Formal justifications or data-dependent results for the range of $\mathcal{K}_n = [\underline{K}, \bar{K}]$ are beyond the scope of this paper, but choice of $\bar{K}$ can be ad hoc in practice. Heuristically, if we use too large $\bar{K}$, then the power of the test based on $T_n(\bar{K},\theta)$ with the normal critical value can be low, as $T_n(\bar{K},\theta)$ can be very small with large variance $V_{\bar{K}}$ under alternatives. Nevertheless, the test based on $\operatorname{Inf} T_n(\theta_0)$ and its asymptotic critical value $c^{\inf}_{1-\alpha}$ may have better power, as this test compare with the smaller critical value than the normal critical value.

Remark 4.2 (Asymptotic power of the test)**.** Although $\operatorname{Inf} T_n(\theta)$ leads to the tests that control the asymptotic size or bound the size distortions, one concern is that possible low power property of the test compare with the other statistics (e.g., the supremum of the t-statistics). Investigating local power comparisons of the level α test based on several different statistics, and the effect of asymptotic bias on subsequent power function are very important, but these are beyond the scope of this paper. I discuss the length of CIs based on the infimum test statistic in Section 5. Furthermore, I calibrate the length of CIs and report power functions of the different test statistics in various simulation setup in Section 8.

In this paper, we focus on controlling size of the test allowing large asymptotic bias and I want to emphasize that bias issues can severely affect commonly used inference procedures (i.e., coverage of standard CI) in series estimation. For example, high-order polynomials are not popular choices because it can be highly sensitive to the choice of series terms. Using low-order polynomials or regression splines can help to reduce bias issues, but does not solve bias problem completely. Moreoever, test based on the other transformation of the t-statistics can be sensitive to the bias problems, thus may lead to over-rejection of the tests (i.e., under-coverage of the CIs). See Appendix C for the inference based on the supremum test statistic, for example.

4.3 Critical values

In this section, I discuss detail descriptions to approximate critical value defined in (4.2). Here, I suggest using simple Monte Carlo method to obtain critical value. To make implementation procedures simple and feasible, I impose following set assumption and conditional homoskedasticity.

Assumption 4.1. *(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* $K_m = \lfloor \pi_m \bar{K} \rfloor$ *for constant* π_m, $0 < \underline{\pi} = \pi_1 < \pi_2 < \cdots < \pi_M = 1$, *and fixed* M.

Assumption 4.2. *(Conditional homoskedasticity)* $E(\varepsilon_i^2 | x_i = x) = \sigma^2$.

Assumption 4.1 is a finite dimensional version of Assumption 3.1, and is different with an alternative set (Assumption 3.4) that considers different rate of Ks. Conditional homoskedasticity assumption is only for a simpler implementation. Based on the general covariance function defined in (3.5), we can construct variance-covariance matrix under general heteroskedastic error.

By Theorem 3.1, following finite dimensional convergence of the t-statistics holds under the Assumptions 3.2, 3.3, 4.1 and 4.2

$$(T_n(K_1, \theta_0), \cdots, T_n(K_M, \theta_0))' \xrightarrow{d} Z = (Z_1, \cdots, Z_M)', \quad Z \sim N(0, \Sigma), \tag{4.8}$$

where Σ is a variance-covariance matrix defined in (3.7), $\Sigma_{jl} = \lim_{n\to\infty} V_{K_j}^{1/2}/V_{K_l}^{1/2}$ for any $j < l$. (4.8) also holds under same assumptions as in Theorem 3.2. Note that the limiting distribution does not depend on θ_0 and variance-covariance matrix Σ can be consistently estimated by its sample counterparts. This requires estimators of the variance V_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 $\widehat{V}_K$ as the simple pluig-in estimator for V_K

$$\begin{aligned} &\widehat{V}_K = P_K(x)'\widehat{Q}_K^{-1}\widehat{\Omega}_K\widehat{Q}_K^{-1}P_K(x), \\ &\widehat{Q}_K = \frac{1}{n}\sum_{i=1}^n P_{Ki}P_{Ki}', \quad \widehat{\Omega}_K = \frac{1}{n}\sum_{i=1}^n P_{Ki}P_{Ki}'\widehat{\varepsilon}_{Ki}^2. \end{aligned} \tag{4.9}$$

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

$$\begin{aligned} &\widehat{c}_{1-\alpha}^{\inf} \equiv (1-\alpha) \text{ quantile of } \inf_{m=1,\cdots,M} |Z_{m,\widehat{\Sigma}}|, \\ &\text{where } Z_{\widehat{\Sigma}} = (Z_{1,\widehat{\Sigma}}, \cdots, Z_{M,\widehat{\Sigma}})' \sim N(0, \widehat{\Sigma}), \quad \widehat{\Sigma}_{jj} = 1, \widehat{\Sigma}_{jl} = \widehat{V}_{K_j}^{1/2}/\widehat{V}_{K_l}^{1/2}. \end{aligned} \tag{4.10}$$

One can compute $\widehat{c}_{1-\alpha}^{\inf}$ 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 $\{\operatorname{Inf} T_n^b = \inf_m |Z_{m,\widehat{\Sigma}}^b| : b = 1, \cdots, B\}$.[3]

[3]Under heteroskedastic error terms, we can construct $\widehat{\Sigma}_{j,l} = \frac{\widehat{V}_{K_{jl}}}{\widehat{V}_{K_j}^{1/2}\widehat{V}_{K_l}^{1/2}}$ for any $j < l$, where $\widehat{V}_{K_{jl}}$ is an

I impose following assumption on consistency of variance estimator $\widehat{V}_K$ uniformly in $K \in \mathcal{K}_n$.

Assumption 4.3. $\sup_{K \in \mathcal{K}_n} |\frac{\widehat{V}_K}{V_K} - 1| = o_p(1)$ *as* $n, K \to \infty$.

Assumption 4.3 is satisfied under same regularity conditions (Assumption 3.1 and 3.2) with an additional assumption. For example, if we further assume $\sup_{K \in \mathcal{K}_n} || \sum_{i=1}^n \tilde{P}_{Ki} \tilde{P}'_{Ki} \varepsilon_i^2 - E[\tilde{P}_{Ki} \tilde{P}'_{Ki} \varepsilon_i^2] || = o_p(1)$ with an orthonormalized vector of basis functions $\tilde{P}_K(x) \equiv Q_K^{-1/2} P_K(x)$, then Assumption 4.3 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.

Next, we consider t-statistic $T_{n,\widehat{V}}(K, \theta) = \sqrt{\frac{n}{\widehat{V}_K}}(\widehat{\theta}_K - \theta_0)$ replacing variance of the series estimator V_K with $\widehat{V}_K$ and joint asymptotic distribution for $K \in \mathcal{K}_n$. Following Corollary provides the validity of Monte Carlo critical values $\widehat{c}_{1-\alpha}^{\inf}$ defined in (4.10).

Corollary 4.3. *Under Assumptions 3.2, 3.3, 4.1, 4.2 and 4.3,* $\widehat{c}_{1-\alpha}^{\inf} \xrightarrow{p} c_{1-\alpha}^{\inf}$ *holds where* $\widehat{c}_{1-\alpha}^{\inf}$ *are defined in* (4.10) *and* $c_{1-\alpha}^{\inf}$ *are the* $(1-\alpha)$ *quantile of the asymptotic null distribution* $\inf_{m=1,\cdots,M} |Z_m|$ *with* $Z = (Z_1, \cdots, Z_M)' \sim N(0, \Sigma)$, Σ *defined in* (3.7). *This also holds under the Assumptions 3.2, 3.3, 3.4, and 4.3.*

Remark 4.3 (Weighted bootstrap)**.** Alternatively, we can use weighted bootstrap method to approximate asymptotic critical value. Implementation of the weighted bootstrap method is as follows. First, generate i.i.d draws from exponential random variables $\{\omega_i\}_{i=1}^n$, independent of the data. Then, for each draw, calculate LS estimator weighted by $\omega_1, \cdots, \omega_n$ for each $K \in \mathcal{K}_n$ and construct weighted bootstrap t-statistic as follows

$$\begin{aligned} \hat{\beta}_K^b &= \arg\min_b \frac{1}{n} \sum_{i=1}^n \omega_i (y_i - P'_{Ki} b)^2, \quad \hat{g}_K^b(x) = P_K(x)' \hat{\beta}_K^b, \\ T_n^b(K) &= \frac{\sqrt{n}(\hat{g}_K^b(x) - \hat{g}_K(x))}{\widehat{V}_K^{1/2}}. \end{aligned} \tag{4.11}$$

Then, construct $\text{Inf}\, T_n^b = \inf_K |T_n^b(K)|$. Repeat this B times (1000 or 5000) and define $\widehat{c}_{1-\alpha}^{\inf,WB}$ as conditional $1-\alpha$ quantile of $\{\text{Inf}\, T_n^b : b = 1, \cdots, B\}$ given the data. The idea behind the weighted bootstrap methods may work is as follows; if the limiting distribution of weighted

sample analog estimator of $P_{K_j}(x)' E(P_{K_j i} P'_{K_l i} \varepsilon_i^2) P_{K_l}(x)$ and $\widehat{V}_{K_j}, \widehat{V}_{K_l}$ are estimator of the variance V_{K_j}, V_{K_l}, respectively.

bootstrap process is equal to the original process conditional on the data, then the weighted bootstrap process $\operatorname{Inf} T_n^b$ also approximate the original limiting distribution $\inf_{\pi\in[\underline{\pi},1]} \mathbb{T}(\pi)$. However, validity of the weighted bootstrap is beyond the scope of this paper and will be pursued for the future work.

5 Confidence intervals

Now, I introduce CIs for $\theta_0 = g_0(x)$ and provide their coverage properties. We consider a confidence interval based on inverting a test statistic for $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$. Define $CI_{\mathtt{inf}}^{\mathtt{Robust}}$ as the nominal level $1-\alpha$ CI for θ based on $\operatorname{Inf} T_n(\theta)$,

$$\begin{aligned} CI_{\mathtt{inf}}^{\mathtt{Robust}} &\equiv \{\theta : \inf_{K\in\mathcal{K}_n} |T_{n,\widehat{V}}(K,\theta)| \leq \widehat{c}_{1-\alpha}^{\inf}\} \\ &= \{\theta : |T_{n,\widehat{V}}(K,\theta)| > \widehat{c}_{1-\alpha}^{\inf}, \forall K\}^C = \bigcup_{K\in\mathcal{K}_n} \{\theta : |T_{n,\widehat{V}}(K,\theta)| \leq \widehat{c}_{1-\alpha}^{\inf}\} \\ &= [\inf_K(\widehat{\theta}_K - \widehat{c}_{1-\alpha}^{\inf} s(\widehat{\theta}_K)), \sup_K(\widehat{\theta}_K + \widehat{c}_{1-\alpha}^{\inf} s(\widehat{\theta}_K))] \end{aligned} \tag{5.1}$$

where $\widehat{c}_{1-\alpha}^{\inf}$ is a critical value defined in Section 4.3, $s(\widehat{\theta}_K) \equiv \sqrt{\widehat{V}_K/n}$ is a standard error of series estimator $\widehat{\theta}_K$ using K series terms, and A^C denotes the complement of a set A. Note that $CI_{\mathtt{inf}}^{\mathtt{Robust}}$ can be easily obtained by using estimates $\widehat{\theta}_K$, standard errors $s(\widehat{\theta}_K)$, and critical value $\widehat{c}_{1-\alpha}^{\inf}$. $CI_{\mathtt{inf}}^{\mathtt{Robust}}$ can be constructed as the lower and the upper end point of confidence intervals for all $K \in \mathcal{K}_n$ using critical value $\widehat{c}_{1-\alpha}^{\inf}$.

Similarly, I define $CI_{\mathtt{inf}}$ based on $\operatorname{Inf} T_n(\theta)$ and the normal critical value $z_{1-\alpha/2}$ as follows,

$$\begin{aligned} CI_{\mathtt{inf}} &\equiv \{\theta : \inf_{K\in\mathcal{K}_n} |T_{n,\widehat{V}}(K,\theta)| \leq z_{1-\alpha/2}\} \\ &= [\inf_K(\widehat{\theta}_K - z_{1-\alpha/2} s(\widehat{\theta}_K)), \sup_K(\widehat{\theta}_K + z_{1-\alpha/2} s(\widehat{\theta}_K))] \end{aligned} \tag{5.2}$$

Note that $CI_{\mathtt{inf}}$ is the union of all standard confidence intervals for $K \in \mathcal{K}_n$ using conventional normal critical value $z_{1-\alpha/2}$.

Next Corollary shows valid coverage property of the above CIs, and it follows from Corollary 4.2 and 4.3.

Corollary 5.1. *1. Under Assumptions 3.2, 3.3, 4.1, 4.2, and 4.3,*

$$\liminf_{n\to\infty} P(\theta_0 \in CI_{inf}^{Robust}) = 1-\alpha, \qquad \liminf_{n\to\infty} P(\theta_0 \in CI_{inf}) \geq 1-\alpha. \tag{5.3}$$

2. *Under Assumptions 3.2, 4.1, 4.2, 4.3, and* $\sup_m |\nu(m)| < \infty$,

$$\liminf_{n\to\infty} P(\theta_0 \in CI_{\mathit{inf}}^{\mathit{Robust}}) \geq 1 - F(c_{1-\alpha}^{\inf}, \inf_m |\nu(m)|), \tag{5.4}$$

$$\liminf_{n\to\infty} P(\theta_0 \in CI_{\mathit{inf}}) \geq 1 - F(z_{1-\alpha/2}, \inf_m |\nu(m)|). \tag{5.5}$$

3. *Under Assumptions 3.2, 3.4, and 4.3,*

$$\liminf_{n\to\infty} P(\theta_0 \in CI_{\mathit{inf}}^{\mathit{Robust}}) \geq 1 - F(c_{1-\alpha}^{\inf}, 0), \tag{5.6}$$

$$\liminf_{n\to\infty} P(\theta_0 \in CI_{\mathit{inf}}) \geq 1 - \alpha. \tag{5.7}$$

Corollary 5.1.1 shows the validity of $CI_{\texttt{inf}}^{\texttt{Robust}}$ and $CI_{\texttt{inf}}$, i.e., asymptotic coverage of CIs are greater than or equal to $1-\alpha$. Note that the Corollary 5.1.1 requires undersmoothing condition, i.e., no asymptotic bias for all Ks in $\mathcal{K}_n$.

Without undersmoothing condition, Corollary 5.1.2 and 5.1.3 show that the coverage probability of $CI_{\texttt{inf}}^{\texttt{Robust}}$ and $CI_{\texttt{inf}}$ are bounded below by the coverage of single K with smallest bias, similar to the asymptotic size results in Corollary 4.2. For example, the lower bound in (5.6) is 0.87 when $c_{1-\alpha}^{\inf} = 1.5$. Furthermore, (5.7) shows that $CI_{\texttt{inf}}$ using normal critical value achieve nominal coverage probability $1-\alpha$. $CI_{\texttt{inf}}$ and $CI_{\texttt{inf}}^{\texttt{Robust}}$ bound coverage distortions even when large asymptotic bias terms ($|\nu(m)| = \infty$) are present for several Ks in $\mathcal{K}_n$. In this sense $CI_{\texttt{inf}}$ and $CI_{\texttt{inf}}^{\texttt{Robust}}$ are robust to the bias problems.

Although $CI_{\texttt{inf}}$ gives formally valid coverage allowing asymptotic bias, coverage property of the $CI_{\texttt{inf}}$ in (5.3) and (5.7) holds with inequality, thus it can be conservative. As the variance of series estimator increases with K, we expect $CI_{\texttt{inf}}$ can be comparable to the standard CI using normal critical values with some large K around the largest $\bar{K}$. In contrast, $CI_{\texttt{inf}}^{\texttt{Robust}}$ have shorter length by using smaller critical value than the normal critical value.

Remark 5.1 (Length of the interval)**.** Note that potential large length of the $CI_{\texttt{inf}}^{\texttt{Robust}}$ is also related to the possible low power property of the test. The net effect of increasing largest $\bar{K}$ on the length of $CI_{\texttt{inf}}^{\texttt{Robust}}$ is not clear as it may decrease critical values $c_{1-\alpha}^{\inf}$ as well.

Also note that the last equality from the definition of $CI_{\texttt{inf}}^{\texttt{Robust}}$ in (5.1) holds only when there is no dislocated CI, i.e., intersection is nonempty at least for some two CIs using $\widehat{c}_{1-\alpha}^{\inf}$. Otherwise, using the superset widens the length of CI. As the variance of series estimator increases with K, we expect that the union of all confidence intervals may only be determined by some large Ks so that there is no dislocated CI. In general, dislocated confidence interval may show some evidence of bias for the specific model, but there is no guarantee that the union of the confidence intervals are connected in practice. Although this paper does not consider data-dependent choice of $\mathcal{K}_n$, possible large length of CI can be avoidable if $\underline{K}$ is reasonably large and this is exactly the condition needed in Corollary 5.1 to have a correct coverage.

6 Post-model selection inference

In this section, I provide methods to construct a valid CI that gives correct coverage after selecting the number of series terms. I first consider the 'post-model selection' t-statistic

$$|T_n(\widehat{K}, \theta)|, \quad \widehat{K} \in \mathcal{K}_n \tag{6.1}$$

where $\widehat{K}$ is a possibly data-dependent rule chosen from $\mathcal{K}_n$. Then, we can define following 'naive' post-selection CI with $\widehat{K}$ using the normal critical value $z_{1-\alpha/2}$,

$$CI_{\texttt{pms}}^{\texttt{Naive}} \equiv \{\theta : |T_n(\widehat{K}, \theta)| \leq z_{1-\alpha/2}\} = [\widehat{\theta}_{\widehat{K}} - z_{1-\alpha/2} s(\widehat{\theta}_{\widehat{K}}), \widehat{\theta}_{\widehat{K}} + z_{1-\alpha/2} s(\widehat{\theta}_{\widehat{K}})]. \tag{6.2}$$

Conventional method of using normal critical value in (6.2) 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 the asymptotic normality of the t-statistic $T_n(\widehat{K}, \theta_0) \xrightarrow{d} N(0, 1)$ holds with some random sequence of $\widehat{K}$. Even if we assume the asymptotic bias is negligible, the variability of $\widehat{K}$ introduced by some selection rules can affect the variance of the asymptotic distribution. Thus, it is not clear whether naive inference using standard normal critical value is valid. If the post-model selection t-statistic, $T_n(\widehat{K}, \theta_0)$ with some $\widehat{K}$, has non-normal asymptotic distribution, then the naive confidence interval $CI_{\texttt{pms}}^{\texttt{Naive}}$ may have coverage probability less than the nominal level $1 - \alpha$.

Furthermore, $\widehat{K}$ with some data-dependent rules may not satisfy the undersmoothing rate conditions which ensure the asymptotic normality without bias terms. For example, suppose a researcher uses $\widehat{K} = \widehat{K}_{\texttt{cv}}$ selected by cross-validation (or other asymptotically equivalent criteria such as AIC). It is well known that the $\widehat{K}_{\texttt{cv}}$ is typically too 'small', so that lead to a large bias by violating undersmoothing assumption needed to ensure asymptotic normality and the valid inference. If $\widehat{K}$ increases not sufficiently fast as undersmoothing condition does, then the asymptotic distribution may have bias terms and resulting naive CI may have large coverage distortions.

Here, I suggest constructing a valid post-selection CI with $\widehat{K} \in \mathcal{K}_n$ by adjusting standard normal critical value to critical value from a 'supremum' test statistic,

$$\operatorname{Sup} T_n(\theta) \equiv \sup_{K \in \mathcal{K}_n} |T_n(K, \theta)|. \tag{6.3}$$

Note that $|T_n(\widehat{K}, \theta_0)| \leq \operatorname{Sup} T_n(\theta_0)$ for any choice of $\widehat{K} \in \mathcal{K}_n$, and $\operatorname{Sup} T_n(\theta_0) \xrightarrow{d} \xi_{\text{sup}} \equiv \sup_{\pi \in [\underline{\pi}, 1]} |\mathbb{T}(\pi)|$ under the same assumptions as in Corollary 4.1. Therefore, inference based on $|T_n(\widehat{K}, \theta_0)|$ using asymptotic critical values from the limiting distribution of $\operatorname{Sup} T_n(\theta_0)$ will be valid, but conservative. Similar to $c^{\text{inf}}_{1-\alpha}$ defined in (4.2), I define asymptotic critical value $c^{\text{sup}}_{1-\alpha}$ as $1-\alpha$ quantile of ξ_{sup}. We can approximate this critical value by using Monte Carlo simulation based method similarly as in Section 4.3. To be specifically, I define

$$\widehat{c}^{\text{sup}}_{1-\alpha} \equiv (1-\alpha) \text{ quantile of } \sup_{m=1,\cdots,M} |Z_{m,\widehat{\Sigma}}|, \tag{6.4}$$

where $Z_{\widehat{\Sigma}} = (Z_{1,\widehat{\Sigma}}, \cdots, Z_{M,\widehat{\Sigma}})' \sim N(0, \widehat{\Sigma})$ and $\widehat{\Sigma}$ are defined in (4.10). Under the same assumptions as in Corollary 4.3, we can also verify $\widehat{c}^{\text{sup}}_{1-\alpha} \xrightarrow{p} c^{\text{sup}}_{1-\alpha}$.

Next, I define the following robust post-selection CI using the critical value $\widehat{c}^{\text{sup}}_{1-\alpha}$ rather than the normal critical value $z_{1-\alpha/2}$ compare to $CI^{\texttt{Naive}}_{\texttt{pms}}$,

$$CI^{\texttt{Robust}}_{\texttt{pms}} \equiv [\widehat{\theta}_{\widehat{K}} - \widehat{c}^{\text{sup}}_{1-\alpha} s(\widehat{\theta}_{\widehat{K}}), \widehat{\theta}_{\widehat{K}} + \widehat{c}^{\text{sup}}_{1-\alpha} s(\widehat{\theta}_{\widehat{K}})], \quad \widehat{K} \in \mathcal{K}_n. \tag{6.5}$$

Next Corollary shows that the robust post-selection $CI^{\texttt{Robust}}_{\texttt{pms}}$ guarantees asymptotic coverage as $1-\alpha$. Even though Corollary 6.1 does not implicitly use randomness of the specific data-dependent selection rules of $\widehat{K}$, $CI^{\texttt{Robust}}_{\texttt{pms}}$ can be useful as it can be applied to any selection rules that researchers might want to use among $\mathcal{K}_n$.

Corollary 6.1. *Under Assumptions 3.2, 3.3, 4.1, 4.2, and 4.3,*

$$\liminf_{n \to \infty} P(\theta_0 \in CI^{\textit{Robust}}_{\textit{pms}}) \geq 1 - \alpha. \tag{6.6}$$

(6.6) *also hold with Assumptions 3.2, 3.3, 3.4, and 4.3 as in Theorem 3.2.*

In Corollary 6.1, I impose an undersmoothing (Assumption 3.3) and optimal MSE rates ($\widehat{K}_{\texttt{cv}}$) are not allowed, thus $CI^{\texttt{Robust}}_{\texttt{pms}}$ does not dealing with the bias problem explicitly. However, it accommodates bias by enlarging confidence interval using larger critical values $\widehat{c}^{\text{sup}}_{1-\alpha}$ than the normal critical value. Moreover, we also expect $\widehat{c}^{\text{sup}}_{1-\alpha}$ is smaller than the usual Bonferroni-type critical value. Bonferroni corrections use normal critical value $z_{1-\frac{\alpha}{2M}}$ replacing α with α/M. However, Bonferroni critical value can be too large especially when $|\mathcal{K}_n| = M$ is large, as it ignores dependence structure of the t-statistics.

7 Extension: partially linear model setup

In this section, I provide inference methods for the partially linear model (PLM) similar to the 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

$$y_i = \theta_0 w_i + g_0(x_i) + \varepsilon_i, \qquad E(\varepsilon_i | w_i, x_i) = 0. \tag{7.1}$$

We are interested in inference on treatment/policy effect θ_0 after approximating unknown function $g_0(x_i)$ by series terms/regressors $p(x_i)$ among a set of potential control variables. Number of regressors could be large if there are many available control variables, i.e., $p(x_i) = x_i$ or if there are large number of transformations of $p(x_i)$ are available such as polynomials and interactions of x_i. Parametric part w_i is always included in the model, however, we are unsure which covariates/transformations of x_i should be used for $g_0(x_i)$.

For notational simplicity, I use the similar notation as defined in nonparametric regression setup. Suppose we use K regressors $P_{Ki} = P_K(x_i)$, where $P_K(x) = (p_1(x), \cdots, p_K(x))'$ from the basis functions $p(x)$. The approximating model can be written as

$$y_i = \theta_0 w_i + P_{Ki}' \beta_K + \varepsilon_{Ki}, \tag{7.2}$$

where the error term $\varepsilon_{Ki} = r_{Ki} + \varepsilon_i$ and approximation error r_{Ki} are defined similarly as in Section 2. Then, series estimator $\widehat{\theta}_K$ for θ_0 using the first K approximating functions is obtained by standard LS estimation of y_i on w_i and P_{Ki}

$$\widehat{\theta}_K = (W' M_K W)^{-1} W' M_K Y \tag{7.3}$$

where $W = (w_1, \cdots, w_n)', M_K = I_K - P^K (P^{K\prime} P^K)^{-1} P^{K\prime}, P^K = [P_{K1}, \cdots, P_{Kn}]', Y = (y_1, \cdots, y_n)'$. Estimator for β_K is given by $(\widehat{\theta}_K, \widehat{\beta}_K')' = (H^{K\prime} H^{K\prime})^{-1} H^{K\prime} Y$ where $H^K = [W, P^K]$. Similar to nonparametric regression model, we are interested in testing for $H_0 : \theta = \theta_0$, $H_1 : \theta \neq \theta_0$.

The asymptotic normality and valid inference for the partially linear model has been developed in the literature. Donald and Newey (1994) derived the asymptotic normality of $\widehat{\theta}_K$ under standard rate conditions where $K/n \to 0$. See also Robinson (1988), Linton (1995) and references therein for the related results of the kernel estimators. 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 under some form of sparsity condition. Recent paper by Cattaneo, Jansson, and Newey (2015a) provided a valid approximation theory for $\widehat{\theta}_K$ even when K grows at the same rate of n.

Different approximation theory using faster rate of K is particularly useful for our purpose to better reflect the choice/search of smoothing parameters in practice than the first order approximation. By using the higher order approximation theory that allows the number of series can grow as fast as sample size n, we can construct a joint distribution of the t-statistics with different sequence of models.

Under $K/n \to c$ for $c > 0$, the limiting normal distribution has a larger variance than the standard asymptotic variance derived under $K/n \to 0$, and the adjusted variance depends on the number of terms K. Unlike the nonparametric object of interest in fully nonparametric model where variance term increases with K, $\widehat{\theta}_K$ has parametric ($n^{1/2}$) convergence rate and variances are same as the semiparametric efficiency bound for all sequences under $K/n \to 0$, i.e., all estimators $\widehat{\theta}_K$ with different rate of Ks satisfying $K/n \to 0$, are all asymptotically equivalent. 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/or sequences of smoothing parameter sequences in certain ranges (see Newey (1994b)). I provide an approximation theory that fully account the dependency of the t-statistics with different Ks.

I impose an assumption that are same as in Cattaneo, Jansson, and Newey (2015a) uniformly over the model $K \in \mathcal{K}_n$, where $\mathcal{K}_n$ is same as in the Assumption 4.1. Let $v_i \equiv w_i - g_{w0}(x_i)$ where $g_{w0}(x_i) \equiv E[w_i|x_i]$.

Assumption 7.1. *(Regularity conditions for Partially Linear Model: Assumption PLM in Cattaneo, Jansson, and Newey (2015a))*

(i) $\{y_i, w_i, x_i\}$ *are i.i.d random variables satisfying the model* (7.1)*.*

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

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

(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 7.1 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 3.2(iii) in nonparametric setup, Assumption 7.1(iv) holds for the polynomials and splines basis. For example, this holds with $\gamma_g = p_g/d_x, \gamma_{g_w} = p_w/d_x$ when $\mathcal{X}$ is compact and unknown functions $g_0(x), g_{w0}(x)$ has p_g, p_w continuous derivates, respectively.

From the results in Cattaneo, Jansson, and Newey (2015a), we have following decomposition for any $K \in \mathcal{K}_n$ under Assumptions 4.1, 7.1 and H_0,

$$
\begin{aligned}
\sqrt{n}(\widehat{\theta}_K - \theta_0) &= (\frac{1}{n}W'M_KW)^{-1}\frac{1}{\sqrt{n}}W'M_KY \\
&= \widehat{\Gamma}_K^{-1}(\frac{1}{\sqrt{n}}\sum_i v_i M_{ii}^K \varepsilon_i + \frac{1}{\sqrt{n}}\sum_{i=1}^n \sum_{j=1, j\neq i}^n v_i M_{ij}^K \varepsilon_j) + o_p(1)
\end{aligned} \tag{7.4}
$$

where $\widehat{\Gamma}_K = W'M_KW/n$. Note that the t-statistics $T_K(\theta_0)$ are asymptotically normal with variance $V = \sigma_\varepsilon^2 E[v_i^2]^{-1}$ under any sequences $K \to \infty$ satisfying the standard rate conditions $K/n \to 0$. However, under the faster rate conditions on K imposed here, the second term is not negligible and converges to bounded random variables. Cattaneo, Jansson, and Newey (2015a) apply 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).

Now, consider the sequence of t-statistics $T_n(K, \theta), K \in \mathcal{K}_n$ for testing H_0. Under Assumptions 4.1, 7.1 and undersmoothing condition $nK^{-2(\gamma_g + \gamma_{gw})} \to 0$, we get following asymptotic null limiting distributions for all deterministic sequence of $K \in \mathcal{K}_n$ assuming conditional homoskedasticity:

$$
\begin{aligned}
&T_n(K, \theta_0) = \sqrt{n}V_K^{-1/2}(\widehat{\theta}_K - \theta_0) \xrightarrow{d} N(0,1), \\
&V_K = (1 - K/n)^{-1}V, \quad V = \sigma_\varepsilon^2 E[v_i^2]^{-1},
\end{aligned}
$$

where V_K coincides with the standard asymptotic variance formula V 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 that are assumed 'small' with the classical condition $K/n \to 0$. Note that the adjusted variance V_K is always greater than V when $K/n \nrightarrow 0$ and is an increasing function of K.

Next theorem is the main result for the partially linear model setup, analogous to nonparametric setup. Theorem 7.1 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 5 and 6.[4]

Theorem 7.1. *Suppose Assumptions 4.1 and 7.1 hold. Also,* $n\bar{K}^{-2(\gamma_g+\gamma_{gw})} \to 0$ *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*

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

with variance-covariance matrix Σ *where* $\Sigma_{jl} \equiv \lim_{n\to\infty} V_{K_{j\wedge l}}^{1/2}/V_{K_{j\vee l}}^{1/2}$ *for* $j \neq l$*, and* $\Sigma_{jl} = 1$ *for* $j = l$*. Moreover, under Assumptions 4.1, 4.3 and 7.1, coverage probability holds for the following CIs*

$$\liminf_{n\to\infty} P(\theta_0 \in CI_{inf}^{Robust}) = 1 - \alpha, \quad \liminf_{n\to\infty} P(\theta_0 \in CI_{inf}) \geq 1 - \alpha \tag{7.5}$$

$$\liminf_{n\to\infty} P(\theta_0 \in CI_{pms}^{Robust}) \geq 1 - \alpha \tag{7.6}$$

where $CI_{inf}^{Robust}, CI_{inf}$*, and* CI_{pms}^{Robust} *are similarly defined as in Section 5 and 6 with PLM estimator* $\widehat{\theta}_K$ *and variance estimator* $\widehat{V}_K$*, and the critical values* $\widehat{c}_{1-\alpha}^{\inf}, \widehat{c}_{1-\alpha}^{\sup}$*.*

Theorem 7.1 derives the joint asymptotic distribution of the $T_n(K, \theta_0)$ over $K \in \mathcal{K}_n$ for the parametric part in partially linear model. Note that the variance-covariance matrix Σ is same as in nonparametric model setup (see equation (3.7) or (4.8)). Variance-covariance matrix Σ_{jl} for any $j \neq l$ can be reduced under the condition $\bar{K}/n \to c$,

$$\Sigma_{jl} = \lim_{n\to\infty} \frac{V_{K_{j\wedge l}}^{1/2}}{V_{K_{j\vee l}^{1/2}}} = \lim_{n\to\infty} \frac{(1 - K_{j\wedge l}/n)^{-1/2} V^{1/2}}{(1 - K_{j\vee l}/n)^{-1/2} V^{1/2}} = \lim_{n\to\infty} \frac{(1 - \pi_{j\wedge l}\bar{K}/n)^{-1/2}}{(1 - \pi_{j\vee l}\bar{K}/n)^{-1/2}} = (\frac{1 - c\pi_{j\vee l}}{1 - c\pi_{j\wedge l}})^{1/2}. \tag{7.7}$$

Remark 7.1 (Variance Estimation)**.** Note that construction of CIs also requires consistent variance estimators $\widehat{V}_K$,

$$\widehat{V}_K = s^2 \widehat{\Gamma}_K^{-1}, \quad s^2 = \frac{1}{n - 1 - K} \sum_{i=1}^{n} \widehat{\varepsilon}_i^2, \quad \widehat{\varepsilon}_i^2 = \sum_{j=1}^{n} M_{K,ij}(y_j - \widehat{\theta}_K w_j). \tag{7.8}$$

For consistent variance estimation results and more discussions, see section 3.2 (Theorem 2) of Cattaneo, Jansson, and Newey (2015a) and also Cattaneo, Jansson, and Newey (2015b) even under heteroskedasticity.

[4]Theorem 7.1 also shows the asymptotic coverage property of CIs similar to Corollary 5.1 in the nonparametric setup. The lower bounds of the asymptotic coverage for $CI_{\texttt{inf}}^{\texttt{Robust}}, CI_{\texttt{inf}}$ can be also derived without undersmoothing assumption ($n\bar{K}^{-2(\gamma_g+\gamma_{gw})} \to 0$), but omitted here for the simplicity.

8 Simulations

This section investigates the small sample performance of the proposed methods in Sections 5 and 6. We are mainly interested in empirical coverage of CIs for the true value of $g(x)$ over the support of x for various functions $g(x)$ and different basis.

I consider the following data generating process similar to Newey and Powell (2003), Chen and Christensen (2015a),

$$y_i = g(x_i) + \varepsilon_i,$$
$$x_i = \Phi(x_i^*), \begin{pmatrix} x_i^* \\ \varepsilon_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & \sigma^2 \end{pmatrix}\right) \tag{8.1}$$

where $\Phi(\cdot)$ is the standard normal cdf 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 standard normal pdf and $sgn(\cdot)$ is sign function. The functions $g_1(x)$ and $g_2(x)$ are used in Newey and Powell (2003), Chen and Christensen (2015) and we label them as linear and nonlinear designs. $g_3(x)$ and $g_4(x)$ are rescaled version of Hall and Horowitz (2013), and we denote these as highly nonlinear designs. See Figure 1 for the shape of all functions on the support $\mathcal{X} = [0,1]$. In addition, I set $\sigma^2 = 1$ for all simulations results below. Results for $\sigma^2 = 0.5, 0.1$ show similar patterns from my experience.

I generate 5000 simulation replications for each different design with sample size $n = 100$. Then, I implement nonparametric series estimators using both power series bases with different orders and quadratic splines with evenly placed knots. In either case, K denotes the number of estimated coefficients. I also set $\mathcal{K}_n = [2, 10]$ for the polynomials and $\mathcal{K}_n = [3, 13]$ for the splines. Then, I calculate pointwise coverage properties of various CIs for all 40 grid points of x on [0,1]. To calculate critical values, 1000 additional Monte Carlo replications are also performed on each simulation iteration. Results for different sample sizes $n = 200, 400$ and results for the cubic spline regressions show similar patterns, thus omitted for brevity.

As a benchmark, I first consider post-selection CI with $\widehat{K}_{\texttt{cv}} \in \mathcal{K}_n$ selected to minimize leave-one-out cross-validation and using (naive) normal critical value, $CI^{\texttt{Naive}}_{\texttt{pms}} = [\widehat{\theta}_{\widehat{K}_{\texttt{cv}}} - z_{1-\alpha/2}s(\widehat{\theta}_{\widehat{K}_{\texttt{cv}}}), \widehat{\theta}_{\widehat{K}_{\texttt{cv}}} + z_{1-\alpha/2}s(\widehat{\theta}_{\widehat{K}_{\texttt{cv}}})]$. I also report coverage of $CI_{\texttt{maxK}} = [\widehat{\theta}_{\bar{K}} - z_{1-\alpha/2}s(\widehat{\theta}_{\bar{K}}), \widehat{\theta}_{\bar{K}} + z_{1-\alpha/2}s(\widehat{\theta}_{\bar{K}})]$ using the largest number of series terms $\bar{K}$. Next, I consider new CIs proposed in this paper, $CI^{\texttt{Robust}}_{\texttt{inf}}$ and $CI_{\texttt{inf}}$, based on the test statistics $\text{Inf}\, T_n(\theta)$ defined in Section 5. Finally, I examine robust post-selection CI, $CI^{\texttt{Robust}}_{\texttt{pms}}$ with $\widehat{K}_{\texttt{cv}}$, defined in Section 6. The critical values, $\widehat{c}^{\,\inf}_{1-\alpha}$ and $\widehat{c}^{\,\sup}_{1-\alpha}$ are constructed using the Monte-Carlo method described in Sections 4.3 and Section 6.

Figure 3 reports nominal 95% coverage probability of all five CIs. Overall, $CI^{\texttt{Robust}}_{\texttt{inf}}$ performs very well across the different simulation designs. Its empirical coverage is close to the nominal 95% level at many points over the support. $CI_{\texttt{inf}}$ using normal critical value also performs well, as coverage is no less than the nominal level at almost all points. However, $CI_{\texttt{inf}}$ seems quite conservative. $CI^{\texttt{Naive}}_{\texttt{pms}}$ using cross-validation selected series terms undercovers most of the cases: $\widehat{K}_{\texttt{cv}}$ is small and $CI^{\texttt{Naive}}_{\texttt{pms}}$ is somewhat narrow to cover the true value. $CI_{\texttt{maxK}}$ slightly undercovers at many points, and works quite poorly especially at the boundary. $CI^{\texttt{Robust}}_{\texttt{pms}}$ with the adjustment of using larger critical value $\widehat{c}^{\sup}_{1-\alpha}$ than normal critical value seems also work well, but does not solve bias problem completely (for example, see coverage probability of $g_2(x = 0.4)$).

For the linear function $g_1(x)$, polynomials should approximate unknown function very well for all K, i.e., finite sample bias is expected to be very small over $K \in \mathcal{K}_n$. In this setup, coverage of $CI^{\texttt{Robust}}_{\texttt{inf}}, CI_{\texttt{maxK}}$ are expected to be close to 95 % and $CI_{\texttt{inf}}, CI^{\texttt{Robust}}_{\texttt{pms}}$ are expected to be conservative. Slightly undercover results in Figure 3-(a) for $CI_{\texttt{maxK}}$ are mostly due to the small sample size. However, given the small sample size, coverage $CI^{\texttt{Robust}}_{\texttt{inf}}$ is still fairly close to 95%.

For the slightly nonlinear function $g_2(x)$, coverage of all confidence intervals except $CI_{\texttt{inf}}$ is less than 0.95 at some points. For example, at $x = 0.4$ and 0.6, the coverage of $CI^{\texttt{Naive}}_{\texttt{pms}}$, $CI^{\texttt{Robust}}_{\texttt{pms}}$ are 0.77, 0.87, respectively. Although it is slightly below than 0.95, coverage of $CI^{\texttt{Robust}}_{\texttt{inf}}$ is 0.93, and this is consistent with our theory that $CI^{\texttt{Robust}}_{\texttt{inf}}$ bounds the size distortions even when there are large biases for some polynomial approximations over $K \in \mathcal{K}_n$. In highly nonlinear function $g_4(x)$, $CI^{\texttt{Robust}}_{\texttt{inf}}$ does not achieve nominal coverage at point $x = 0.5$. At this single peak at $x = 0.5$, every polynomial approximation has large bias. Possibly poor coverage property at this point was also described in Hall and Horowitz (2013, Figure 3). In this case, regression spline seems much better for approximating this local point. Figure 4 shows the coverage probability of CIs using quadratic splines with different number of knots. As we can see from Figure 4, $CI^{\texttt{Robust}}_{\texttt{inf}}$ with splines works better to achieve correct coverage for $g_2(x = 0.4), g_4(x = 0.5)$, and for other different functions as well.

In Figure 5, I compare the length of the five CIs for the polynomial series. In the linear and nonlinear designs, rank of the length in a narrower order is (roughly) as follows; $CI^{\texttt{Naive}}_{\texttt{pms}} < CI^{\texttt{Robust}}_{\texttt{pms}} \leq CI^{\texttt{Robust}}_{\texttt{inf}} < CI_{\texttt{maxK}} < CI_{\texttt{inf}}$. This is what we expected as $CI^{\texttt{Naive}}_{\texttt{pms}}$ is too narrow, $CI_{\texttt{maxK}}$ is somewhat wide because of large variance using $\bar{K}$. For the highly nonlinear design, $CI^{\texttt{Robust}}_{\texttt{inf}}$ and $CI_{\texttt{inf}}$ become wider at some points where estimates are relatively sensitive across K. Length of $CI_{\texttt{maxK}}$ is similar for $g_3(x)$ or shorter for $g_4(x)$ compare than $CI^{\texttt{Robust}}_{\texttt{inf}}$. Figure 6 compares the length of CIs for the splines, and it shows similar patterns with polynomial approximation. Given that $CI^{\texttt{Robust}}_{\texttt{inf}}$ has a similar or only a slightly wider length than the others, we want to highlight that it has better or similar coverage probability at most points than $CI_{\mathtt{maxK}}, CI_{\mathtt{pms}}^{\mathtt{Naive}}$ and $CI_{\mathtt{pms}}^{\mathtt{Robust}}$, as in Figure 4.

We expect that the coverage probability of $CI_{\mathtt{maxK}}$ can be better when $\bar{K}$ coincides with coverage optimal K^* that minimizes the distance $|P(\theta_0 \in CI(K)) - (1-\alpha)|$, where $CI(K)$ is a standard CI using K series terms and the normal critical value. However, as I already emphasized, there is no formal data-dependent method to choose such large enough K^*: It also depends on the sample sizes and unknown smoothness of the underlying function. If $\bar{K}$ is smaller than the K^*, then $CI_{\mathtt{maxK}}$ may undercover because of bias problems. If $\bar{K}$ is larger than K^*, then $CI_{\mathtt{maxK}}$ may be too wide because of large variance, or the normal distribution may be a poor approximation with $\bar{K}$ in small sample size. In contrast, $CI_{\mathtt{inf}}^{\mathtt{Robust}}$ and $CI_{\mathtt{inf}}$ is least affected with those small K with large bias, and performs quite well even in small sample size.

In sum, $CI_{\mathtt{inf}}^{\mathtt{Robust}}$ seems to work well in various simulation experiments. It is the only method close to nominal coverage and it is least affected by biases. In terms of coverage, $CI_{\mathtt{inf}}$ also performs well but it can be quite conservative. In some simulation results, coverage of $CI_{\mathtt{pms}}^{\mathtt{Robust}}$ is close to the nominal level, thus it is also advisable to report.

In addition to length comparisons, I also provide power of the different test statistics. In Figure 7, I report power functions of the three different test statistics to test $H_0 : \theta = \theta_0$ against fixed alternatives $H_1 : \theta = \theta_0 + \delta$ where $\theta_0 = g_2(x)$ evaluated at some point x. Of course, the power depends on different point of interest x. I consider two cases where bias of series estimator for $g_2(x)$ is small ($x = 0.5$) and relatively large ($x = 0.4$). I plot following rejection probability based on $\mathrm{Inf}\, T_n(\theta), \mathrm{Sup}\, T_n(\theta)$, and $|T_n(\widehat{K}, \theta)|$ with appropriate critical values as a functions of δ: (1) $P(|T_n(\widehat{K}_{\mathtt{cv}}, \theta_0 + \delta)| > z_{1-\alpha/2})$ with $\widehat{K}_{\mathtt{cv}}$; (2) $P(\mathrm{Inf}\, T_n(\theta_0 + \delta) > \widehat{c}_{1-\alpha}^{\mathrm{inf}})$; (3) $P(\mathrm{Inf}\, T_n(\theta_0 + \delta) > z_{1-\alpha/2})$; (4) $P(\mathrm{Sup}\, T_n(\theta_0 + \delta) > \widehat{c}_{1-\alpha}^{\mathrm{sup}})$; (5) $P(|T_n(\widehat{K}_{\mathtt{cv}}, \theta_0 + \delta)| > \widehat{c}_{1-\alpha}^{\mathrm{sup}})$. As expected, Figure 7-(a) and (b) show that the tests based on $\mathrm{Inf}\, T_n(\theta)$ are the only method to control size or bound the size distortions when bias exists for some Ks.

9 Illustrative empirical application : Nonparametric estimation of labor supply function and wage elasticity with nonlinear budget set

In this section, I illustrate robust inference procedures by revisiting a paper by Blomquist and Newey (2002). Understanding how tax and policy affect individual labor supply has been central issues in labor economics (see Hausman (1985) and Blundell and MaCurdy

(1999), among many others). Focusing on the conditional mean of hours of work given the individual budget set, Blomquist and Newey (2002) estimate labor supply function using nonparametric series estimation. They also estimate other functionals such as wage elasticity of the expected labor supply and find some evidence of possible misspecification of the usual parametric model (e.g. maximum likelihood estimation (MLE)).

Specifically, they consider following models by exploiting additive structure follows from the utility maximization with piecewise linear budget sets.

$$h_i = g(x_i) + \varepsilon_i, \quad E(\varepsilon_i | x_i) = 0, \tag{9.1}$$

$$g(x_i) = g_1(y_J, w_J) + \sum_{j=1}^{J-1} [g_2(y_j, w_j, \ell_j) - g_2(y_{j+1}, w_{j+1}, \ell_j)], \tag{9.2}$$

where h_i is the hours of the ith individual and $x_i = (y_1, \cdots, y_J, w_1, \cdots, w_J, \ell_1, \cdots, \ell_J)$ is the budget set that can be represented by 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, I use the similar notations with theirs. Equation (9.2) for the conditional mean function follows from Theorem 2.1 of Blomquist and Newey (2002), and this additive structure greatly reduce dimensionality. They consider following power series for $g(x)$

$$p_k(x) = (y_J^{p_1(k)} w_J^{q_1(k)}, \sum_{j=1}^{J-1} \ell_j^{m(k)} (y_j^{p_2(k)} w_j^{q_2(k)} - y_{j+1}^{p_2(k)} w_{j+1}^{q_2(k)})). \tag{9.3}$$

Using the data from the Swedish "Level of Living" survey in 1973, 1980 and 1990, they pool the data from three waves and use the data from 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

$$E_w = \bar{w}/\bar{h}[\frac{\partial g(w, \cdots, w, \bar{y}, \cdots, \bar{y})}{\partial w}]|_{w=\bar{w}}, \tag{9.4}$$

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

Table 1 is exactly the same table used 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 each row. For example, estimates in the second raw use the term in the first row $(1, y_J, w_J)$ with additional terms $(\Delta y, \Delta w)$. Here, $\ell^m \Delta y^p w^q$ denotes approximating term $\sum_i \ell_j^m (y_j^p w_j^q - y_{j+1}^p w_{j+1}^q)$. They also report cross-validation criteria,

CV, for each model specification. In their formula, series terms are chosen to maximize CV, which minimizes asymptotic MSE. In addition to their original table, I also report CI for each specification. As we can see from the table, it is ambiguous which large K should be used for the inference. We do not have compelling reason to select one of the large K for the confidence interval to be reported.

I report proposed robust confidence interval, $CI_{\texttt{inf}}^{\texttt{Robust}}$ as well as $CI_{\texttt{inf}}$, $CI_{\texttt{pms}}^{\texttt{Robust}}$ defined in Sections 5 and 6. For this, I exploit the covariance structure in the joint asymptotic distribution of the t-statistics under homoskedastic error; the variance-covariance matrix is only a function of the variance of series estimators. Therefore, construction of the critical value using the Monte Carlo method defined in (4.10) only requires estimated variance for different specifications that are reported in the table of Blomquist and Newey (2002). It is quite straightforward to construct the proposed CI without any replication of the data sets in this case and this is one of the computational advantages of our procedure. If we have the dataset, then we could also implement critical value based on general variance forms under heteroskedasticity or bootstrap critical value. Using Monte-Carlo method, estimated critical values are $\widehat{c}_{1-\alpha}^{\inf} = 0.9668, \widehat{c}_{1-\alpha}^{\sup} = 2.4764$, respectively.

Robust CI based on the infimum of the t-statistics, $CI_{\texttt{inf}}^{\texttt{Robust}}$ is $[0.0271, 0.1111]$ and this is quite comparable to the CI with some large K, for example, $CI = [0.0273, 0.1045]$ using all the additional terms up to the 6th row. Morover, $CI_{\texttt{inf}}^{\texttt{Robust}}$ is substantially tighter than $CI_{\texttt{maxK}} = [0.0148, 0.1280]$ using the largest number of series terms $\bar{K}$ as well as those based on the second largest series terms, $[0.0214, 0.1336]$.

$CI_{\texttt{inf}}$ using normal critical value is $[0.0148, 0.1384]$, and this turns out to be the union of CI with the largest and the third largest number of series terms. Naive post-selection CI with $\widehat{K}_{\texttt{cv}}$ is $CI_{\texttt{pms}}^{\texttt{Naive}} = [0.0247, 0.0839]$, and this seems somewhat narrow in this case. $CI_{\texttt{pms}}^{\texttt{Robust}}$ widens naive confidence interval to $[0.0169, 0.0916]$.

10 Conclusion

This paper considers the construction of inference methods with data-dependent number of series terms in nonparametric series regression model. New inference methods proposed in this paper are based on two innovations. First, I provide an empirical process theory for the t-statistic sequences indexed by the number of series terms over a set. Second, I introduce tests based on the infimum of the t-statistics over different series terms and show that the tests control the asymptotic size with undersmoothing condition or bound the size distortions without undersmoothing condition. Pointwise confidence interval for the true regression function is obtained by test statistic inversion. To construct the critical value and a valid

CI, I suggest using a simple Monte Carlo simulation based method. In various simulation experiments, CI based on the infimum t-statistics performs well; coverage is close to the nominal level and least affected by finite sample bias. I illustrate proposed CI by revisiting empirical example of Blomquist and Newey (2002). I also provide methods of constructing a valid CI after selecting the number of series terms by adjusting the conventional normal critical value to the critical value based on the supremum of the t-statistics. Furthermore, I provide an extension of the proposed methods in the partially linear model setup.

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.

Armstrong, T. B. (2015): "Adaptive Testing on a Regression Function at a Point," *The Annals of Statistics*, 43, 2086-2101.

Armstrong, T. B. and M. Kolesár (2015): "A Simple Adjustment for Bandwidth Snooping," Working Paper.

Athey, S. and G.W. Imbens (2015): "A Measure of Robustness to Misspecification," *American Economic Review: Papers and Proceedings*, 105, 476-480.

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.

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.

Gallant, A.R. and G. Souza (1991): "On the Asymptotic Normality of Fourier Flexible Form Estimates," *Journal of Econometrics*, 50, 329-353.

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.

Hausman, J. A. and W. K. Newey (1995): "Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss", *Econometrica*, 63, 1445-1476.

Heckman, J. J., L. J. Lochner, and P. E. Todd (2006): "Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond," *Handbook of the Economics of Education*, In: E. A. Hanushek, and F. Welch (Eds.), Vol. 1, Elsevier, Chapter 7.

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.

Horowitz, J. L. and V. G. Spokoiny (2001): "An Adaptive, Rate-Optimal Test of a Parametric Mean-Regression Model Against a Nonparametric Alternative" *Econometrica*, 69, 599-631.

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.

Ichimura H. and P. E. Todd (2007): "Implementing Nonparametric and Semiparametric Esstimators," *Handbook of Econometrics*, In: J.J. Heckman, E. Leamer (Eds.), vol. 6B., Elsevier, Chapter 74.

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.

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. (1990): "Efficient Instrumental Variables Estimation of Nonlinear Models," *Econometrica*, 58(4), 809-837.

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.

NEWEY, W. K., F. HSIEH, J. ROBINS (2003): "Undersmoothing and Bias Corrected Functional Estimation," *Working Paper.*

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.

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.

Varian, H. R. (2014): "Big Data: New Tricks for Econometrics," *Journal of Economic Perspectives*, 28, 3-28.

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.

A Proofs

In the Appendix, we define additional notations for the empirical process theory used in the proof of Theorem 3.1. Given measurable space $(S, \mathcal{S})$, let $\mathcal{F}$ as a class of measurable functions $f : \mathcal{S} \to \mathbb{R}$. We define $N(\epsilon, \mathcal{F}, L_2(Q))$ as covering numbers relative to the $L_2(Q)$ norms, which is the minimal number of the $L_2(Q)$ balls of radius ϵ to cover $\mathcal{F}$ with $L_2(Q)$ norms $||f||_{Q,2} = (\int |f|^2 dQ)^{1/2}$ and measure Q. Uniform entropy numbers relative to L_2 are defined as $\sup_Q \log N(\epsilon ||F||_{Q,2}, \mathcal{F}, L_2(Q))$ where 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 to discuss 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 universal constant that does not depend on n.

A.1 Proof of Theorem 3.1

For any sequence $\{K(n) = \lfloor \pi \bar{K}(n) \rfloor : n \geq 1\} \in \prod_{n=1}^{\infty} \mathcal{K}_n$ under Assumptions 3.1 and 3.2, we first define orthonormalized vector of basis functions

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

We observe that

$$
\begin{aligned}
&\widehat{g}_K(x) = P_K(x)'(P^{K\prime} P^K)^{-1} P^{K\prime} y = \tilde{P}_K(x)'(\tilde{P}^{K\prime} \tilde{P}^K)^{-1} \tilde{P}^{K\prime} y,\\
&V_K(x) = P_K(x)' Q_K^{-1} \Omega_K Q_K^{-1} P_K(x) = \tilde{P}_K(x)' \tilde{\Omega}_K \tilde{P}_K(x),\\
&\tilde{\Omega}_K = E(\tilde{P}_{Ki} \tilde{P}_{Ki}' \varepsilon_i^2).
\end{aligned}
$$

Without loss of generality, we may impose normalization of $Q_{\bar{K}} = I$ or $Q_K = E(P_{Ki} P_{Ki}') = I_K$ uniformly over $K \in \mathcal{K}_n$, since $\widehat{g}_K(x)$ 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 in (2.2), with abuse of notation, using orthornormalized series terms $\tilde{P}_{Ki}$. That is, $y_i = \tilde{P}_{Ki}' \beta_K + \varepsilon_{Ki}, E[\tilde{P}_{Ki} \varepsilon_{Ki}] = 0$ where $\varepsilon_{Ki} = r_{Ki} + \varepsilon_i$, $r_K(x) = g_0(x) - \tilde{P}_K(x)' \beta_K, r_{Ki} = r_K(x_i)$, and $r_K \equiv (r_{K1}, \cdots r_{Kn})'$. We also define $\widehat{Q}_K \equiv \frac{1}{n} \tilde{P}^{K\prime} \tilde{P}^K$, $\underline{\sigma}^2 \equiv \inf_x E[\varepsilon_i^2 | x_i = x], \bar{\sigma}^2 \equiv \sup_x E[\varepsilon_i^2 | x_i = x]$. We first provide useful lemmas which will be used in the proof of Theorem 3.1. Versions of proof of Lemma 1 are available in the literature, such as Newey (1997), 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 can be used in Assumption 3.2, but lead to different bounds in (A.1)-(A.3) in the following Lemma 1.

Lemma 1. *Under Assumptions 3.1 and 3.2, for any $K \in \mathcal{K}_n$, following holds*

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

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

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

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 $\sqrt{\frac{1}{nV_K}}\tilde{P}_K(x)'\left(\widehat{Q}_K^{-1} - I_K\right)\tilde{P}^{K\prime}\varepsilon$. Conditional on the sample $X = [x_1, \cdots, x_n]$, this term has mean zero and variance,

$$
\begin{aligned}
&\frac{1}{nV_K}\tilde{P}_K(x)'\left(\widehat{Q}_K^{-1} - I_K\right)\tilde{P}^{K\prime}E(\varepsilon\varepsilon'|X)\tilde{P}^K\left(\widehat{Q}_K^{-1} - I_K\right)\tilde{P}_K(x)\\
&\le \frac{\bar{\sigma}^2}{V_K}\tilde{P}_K(x)'\left(\widehat{Q}_K^{-1} - I_K\right)\widehat{Q}_K\left(\widehat{Q}_K^{-1} - I_K\right)\tilde{P}_K(x)\\
&= \frac{\bar{\sigma}^2}{V_K}\tilde{P}_K(x)'\left(\widehat{Q}_K - I_K\right)\widehat{Q}_K^{-1}\left(\widehat{Q}_K - I_K\right)\tilde{P}_K(x)\\
&\le \frac{\bar{\sigma}^2\tilde{P}_K(x)'\tilde{P}_K(x)}{V_K}\lambda_{max}(\widehat{Q}_K^{-1})||\left(\widehat{Q}_K - I_K\right)||^2\\
&= O_P(\lambda_K^2\zeta_K^2\log(K)/n)
\end{aligned}
$$

where the first and the last inequality uses $V_K \le \bar{\sigma}^2\tilde{P}_K(x)'\tilde{P}_K(x)$, $V_K \ge \underline{\sigma}^2\tilde{P}_K(x)'\tilde{P}_K(x)$ by Assumption 3.2(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 $|\lambda_{min}(\widehat{Q}_K) - 1| \le ||\widehat{Q}_K - I_K|| = o_p(1)$ by (A.1) and Assumption 3.2(iv)-(v). By Chebyshev's inequality, we have that

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

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

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

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

$$
\begin{aligned}
|\sqrt{\frac{1}{nV_K}}\tilde{P}_K(x)'\left(\widehat{Q}_K^{-1} - I_K\right)\tilde{P}^{K\prime}r_K| &\le C||\widehat{Q}_K^{-1}||\cdot||\left(\widehat{Q}_K - I_K\right)||||\frac{1}{\sqrt{n}}\sum_{i=1}^n \tilde{P}_{Ki}r_{Ki}||\\
&= O_p(\sqrt{\frac{\lambda_K^2\zeta_K^2\log(K)}{n}}\ell_K c_K\sqrt{K})
\end{aligned}
$$

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

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

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

since $E[(\frac{\tilde{P}_K(x)'\tilde{P}_{Ki}}{V_K^{1/2}})^2] \asymp 1$ by Assumption 3.2(ii) and $E(r_{Ki})^2 \leq (\ell_K c_K)^2$ by Assumption 3.2(iii). Therefore, we have (A.3) by Chebyshev's inequality and using $E[\tilde{P}_{Ki}r_{Ki}] = 0$ from projection model. This completes the proof. *Q.E.D.*

Proof of Theorem 3.1. For any $\pi \in \Pi = [\underline{\pi}, 1]$, we first show the decomposition of the t-statistic in equation (3.2).

$$\begin{aligned}
T_n^*(\pi,\theta_0) &= T_n(\lfloor \pi\bar{K}\rfloor,\theta)\\
&= \sqrt{\frac{n}{V_\pi}}\tilde{P}_\pi(x)'(\widehat{\beta}_{\lfloor \pi\bar{K}\rfloor} - \beta_{\lfloor \pi\bar{K}\rfloor}) - \sqrt{\frac{n}{V_\pi}}r_\pi\\
&= \sqrt{\frac{1}{nV_\pi}}\tilde{P}_\pi(x)'\tilde{P}^{\lfloor \pi\bar{K}\rfloor\prime}(\varepsilon + r_{\lfloor \pi\bar{K}\rfloor\prime})\\
&\quad + \sqrt{\frac{1}{nV_\pi}}\tilde{P}_\pi(x)'\left(\widehat{Q}^{-1}_{\lfloor \pi\bar{K}\rfloor} - I_{\lfloor \pi\bar{K}\rfloor}\right)\tilde{P}^{\lfloor \pi\bar{K}\rfloor\prime}(\varepsilon + r_{\lfloor \pi\bar{K}\rfloor\prime}) - \sqrt{\frac{n}{V_\pi}}r_\pi\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\tilde{P}_\pi(x)'\tilde{P}_{\pi i}\varepsilon_i}{V_\pi^{1/2}} + R_1(\lfloor \pi\bar{K}\rfloor) + R_2(\lfloor \pi\bar{K}\rfloor) - \sqrt{n}V_\pi^{-1/2}r_\pi
\end{aligned}$$

where $R_1(K), R_2(K)$ are defined in (A.2), (A.3).

By Lemma 1, we have $R_1(K) = O_p(\sqrt{\frac{\zeta_K^2\log K}{n}}(1+\ell_K c_K\sqrt{K})) = o_p(1)$, $R_2(K) = O_p(\ell_K c_K) = o_p(1)$ for any $K = \lfloor \pi\bar{K}\rfloor \in \mathcal{K}_n$ under Assumptions 3.1 and 3.2. Therefore we have following decomposition for any $\pi \in \Pi$

$$T_n^*(\pi,\theta_0) = t_n^*(\pi) - \sqrt{n}V_\pi^{-1/2}r_\pi + o_p(1), \tag{A.5}$$

where

$$t_n^*(\pi) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\tilde{P}_\pi(x)'\tilde{P}_{\pi i}\varepsilon_i}{V_\pi^{1/2}}. \tag{A.6}$$

Now we show weak convergence of the empicial process $\{t_n^*(\cdot) : n \geq 1\}$ to the mean zero Gaussian process $\mathbb{T}(\cdot)$ defined in the Theorem 3.1. Let $\mathcal{F}_n = \{f_{n,\pi} : \pi \in \Pi\}$ be a sequence of classes of measurable functions $f_{n,\pi}:(\mathcal{E}\times\mathcal{X})$ to $\mathbb{R}$ indexed by π,

$$f_{n,\pi}(\varepsilon,t)=\frac{\tilde{P}_{\pi}(x)'\tilde{P}_{\pi}(t)\varepsilon}{V_{\pi}^{1/2}(x)}=\frac{\tilde{P}_{\lfloor\bar{K}\pi\rfloor}(x)'\tilde{P}_{\lfloor\bar{K}\pi\rfloor}(t)\varepsilon}{V_{\lfloor\bar{K}\pi\rfloor}^{1/2}(x)},(\varepsilon,t)\in\mathcal{E}\times\mathcal{X}. \tag{A.7}$$

Consider empirical process $\{t_n^*(\pi):\pi\in\Pi\}=\{n^{-1/2}\sum_{i=1}^n f_{n,\pi}(\varepsilon_i,x_i):\pi\in\Pi\}$ indexed by classes of functions $\mathcal{F}_n=\{f_{n,\pi}:\pi\in\Pi\}$. We want to show weak convergence of the stochastic process 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 follows from marginal convergence to a Gaussian process 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

$$Ef_{n,\pi_1}f_{n,\pi_2}-Ef_{n,\pi_1}Ef_{n,\pi_2}=\frac{\tilde{P}_{\pi_1}(x)'E(\tilde{P}_{\pi_1}(x_i)\tilde{P}_{\pi_2}(x_i)'\varepsilon_i^2)\tilde{P}_{\pi_2}(x)}{V_{\pi_1}^{1/2}V_{\pi_2}^{1/2}}. \tag{A.8}$$

This term converges to the claimed covariance function $\Sigma(\pi_1,\pi_2)$. This covariance kernel can be bounded below and above some constant $0<c,C<\infty$ for all n,

$$c\le\underline{\sigma}^2\frac{V_{\pi_1}^{1/2}}{V_{\pi_2}^{1/2}}\le\frac{\tilde{P}_{\pi_1}(x)'E(\tilde{P}_{\pi_1}(x_i)\tilde{P}_{\pi_2}(x_i)'\varepsilon_i^2)\tilde{P}_{\pi_2}(x)}{V_{\pi_1}^{1/2}V_{\pi_2}^{1/2}}\le\bar{\sigma}^2\frac{V_{\pi_1}^{1/2}}{V_{\pi_2}^{1/2}}\le C \tag{A.9}$$

by using $\underline{\sigma}^2\tilde{P}_{\pi}(x)'\tilde{P}_{\pi}(x)\le V_{\pi}\le\bar{\sigma}^2\tilde{P}_{\pi}(x)'\tilde{P}_{\pi}(x)$ from Assumption 3.2(ii). We also use the fact that $V_{\pi_1}^{1/2}\asymp V_{\pi_2}^{1/2}\asymp||\tilde{P}_{\bar{K}}||$ for any π_1,π_2 under Assumption 3.1 and 3.2.

To show the finite dimensional convergence, by the Cramér-Wold device, it suffices to show that for any $\pi_1<\cdots<\pi_M$,

$$\delta' t_n^*\xrightarrow{d}N(0,\delta'\Sigma\delta)\quad\forall\delta\in\mathbb{R}^M \tag{A.10}$$

where $t_n^*=(t_n^*(\pi_1),...,t_n^*(\pi_M))',\Sigma_{jl}=\lim_{n\to\infty}\Sigma_{jl,n}$, $\Sigma_{jl,n}\equiv\frac{\tilde{P}_{\pi_j}(x)'E(\tilde{P}_{\pi_j i}\tilde{P}'_{\pi_l i}\varepsilon_i^2)\tilde{P}_{\pi_l}(x)}{V_{\pi_j}^{1/2}V_{\pi_l}^{1/2}}$. To show (A.10) 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 $\omega_{ni}=(\delta'\Sigma_n\delta)^{-1/2}\sum_{j=1}^M\delta_j\frac{\tilde{P}_{\pi_j}(x)'\tilde{P}_{\pi_j i}\varepsilon_i}{V_{\pi_j}^{1/2}}$. 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 3.2, we have $||\sum_{j=1}^M\delta_j\frac{\tilde{P}_{\pi_j}(x)'\tilde{P}_{\pi_j i}}{V_{\pi_j}^{1/2}}||_\infty\lesssim\zeta_{\bar{K}}\lambda_{\bar{K}}$. Moreover, $(\delta'\Sigma_n\delta)^{-1}\lesssim 1$.

Therefore, for any $a > 0$,

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} E(|\omega_{ni}|^2 1\{|\omega_{ni}| > a\sqrt{n}\}) \\
&\lesssim M\sum_{j=1}^{M} E[|\frac{\tilde{P}_{\pi_j}(x)'\tilde{P}_{\pi_j i}\varepsilon_i}{V_{\pi_j}^{1/2}}|^2 1\{|\sum_{j=1}^{M}\delta_j\frac{\tilde{P}_{\pi_j}(x)'\tilde{P}_{\pi_j i}}{V_{\pi_j}^{1/2}}\varepsilon_i| > a\sqrt{n}\}] \\
&\leq M\sum_{j=1}^{M} E(|\frac{\tilde{P}_{\pi_j}(x)'\tilde{P}_{\pi_j i}}{V_{\pi_j}^{1/2}}|^2)\sup_x E[\varepsilon_i^2 1\{|\varepsilon_i| > a(\sqrt{n}/(\zeta_{\bar{K}}\lambda_{\bar{K}})\}|x_i = x],
\end{aligned}
$$

where the last term goes to 0 under $n \to \infty$ by Assumption 3.2(ii), since $E[(\frac{\tilde{P}_\pi(x)'\tilde{P}_{\pi i}}{V_\pi^{1/2}})^2] \asymp 1$ for any π and $(\zeta_{\bar{K}}\lambda_{\bar{K}})/\sqrt{n} = o(1)$ by Assumption 3.2(iv). Thus, Lindberg condition is verified and therefore (A.10) holds by Lindberg-Feller CLT and Slutzky's Theorem. We show that the finite dimensional convergence to a Gaussian distribution with covariance kernel in the Theorem 3.1.

Now, we only need to show stochastic equicontinuity. Define $\alpha(x,\pi) \equiv \tilde{P}_\pi(x)/V_\pi^{1/2}(x) = \tilde{P}_\pi(x)/||\Omega_\pi^{1/2}\tilde{P}_\pi(x)||$. Note that $|f_{n,\pi}(\varepsilon,t)| = |\alpha(x,\pi)'P_\pi(t)\varepsilon| \leq C|f_{n,1}(\varepsilon,t)| \leq C|\varepsilon|\zeta_{\bar{K}}\lambda_{\bar{K}}$. We define envelope function $F_n(\varepsilon,t) \equiv |f_{n,1}(\varepsilon,t)| \vee 1$. 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 $a > 0$,

$$
\begin{aligned}
E(F_n^2 1\{F_n > a\sqrt{n}\}) &= E[(\frac{\tilde{P}_1(x)'\tilde{P}_1(x_i)}{V_{\underline{\pi}}^{1/2}}\varepsilon_i)^2 1\{|\varepsilon_i| > a(\sqrt{n}/(\zeta_{\bar{K}}\lambda_{\bar{K}})\}] && \text{(A.11)} \\
&\leq \sup_x E[\varepsilon_i^2 1\{|\varepsilon_i| > a(\sqrt{n}/(\zeta_{\bar{K}}\lambda_{\bar{K}})\}|X_i = x] = o(1) && \text{(A.12)}
\end{aligned}
$$

since $(\zeta_{\bar{K}}\lambda_{\bar{K}})/\sqrt{n} = o(1)$ by Assumption 3.2(ii). Moreover, for every $\delta_n \to 0$,

$$
\sup_{\rho(\pi_1,\pi_2)<\delta_n} E(f_{n,\pi_1} - f_{n,\pi_2})^2 \to 0 \tag{A.13}
$$

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

Define $\kappa_{1,n} \equiv \sup_{x\in\mathcal{X}}\sup_{\pi\neq\pi'}\frac{||\tilde{P}_{\pi'-\pi}(x)||}{||\pi'-\pi||}$ 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, $\kappa_{1,n} \lesssim \sup_{x\in\mathcal{X}}||\tilde{P}_{\pi'-\pi}(x)|| \lesssim \zeta_{\bar{K}}\lambda_{\bar{K}}$ under Assumption 3.1 and 3.2. Also define $\kappa_{2,n} \equiv \sup_{\pi\neq\pi'}\frac{|V_{\pi'}(x)-V_\pi(x)|}{||\pi'-\pi||}$.

Then, for any $\pi, \pi' \in \Pi = [\underline{\pi}, 1]$ such that $\pi < \pi'$, following holds for sufficiently large n,

$$
\begin{aligned}
&|\alpha(x,\pi')'P_{\pi'}(t) - \alpha(x,\pi)'P_{\pi}(t)| = |\frac{\tilde{P}_{\pi'}(x)'\tilde{P}_{\pi'}(t)}{V_{\pi'}^{1/2}(x)} - \frac{\tilde{P}_{\pi}(x)'\tilde{P}_{\pi}(t)}{V_{\pi}^{1/2}(x)}| && \text{(A.14)}\\
&\leq |\frac{\tilde{P}_{\pi'}(x)'\tilde{P}_{\pi'}(t) - \tilde{P}_{\pi}(x)'\tilde{P}_{\pi}(t)}{V_{\pi'}^{1/2}(x)}| + |\tilde{P}_{\pi}(x)'\tilde{P}_{\pi}(t)(\frac{1}{V_{\pi'}^{1/2}(x)} - \frac{1}{V_{\pi}^{1/2}(x)})| && \text{(A.15)}\\
&\leq (\sup_{\pi} \frac{1}{|V_{\pi}^{1/2}(x)|})|\tilde{P}_{\pi'-\pi}(x)'\tilde{P}_{\pi'-\pi}(t)| + |\frac{\tilde{P}_{\pi}(x)'\tilde{P}_{\pi}(t)}{V_{\pi}^{1/2}(x)}(\frac{V_{\pi'}(x) - V_{\pi}(x)}{V_{\pi'}^{1/2}(x)(V_{\pi}^{1/2}(x) + V_{\pi'}^{1/2}(x))})| && \text{(A.16)}\\
&\leq C_1\kappa_{1,n}||\pi' - \pi|| + C_2\zeta_{\bar{K}}\lambda_{\bar{K}}\frac{1}{\inf_{\pi}|V_{\pi}(x)|}\kappa_{2,n}||\pi' - \pi|| && \text{(A.17)}\\
&\leq C_3\zeta_{\bar{K}}\lambda_{\bar{K}}||\pi' - \pi|| + C_4\zeta_{\bar{K}}\lambda_{\bar{K}}||\pi' - \pi|| = A\zeta_{\bar{K}}\lambda_{\bar{K}}||\pi' - \pi|| && \text{(A.18)}
\end{aligned}
$$

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 $\kappa_{1,n}, \kappa_{2,n}$, $||\sup_{\pi}(1/|V_{\pi}^{1/2}(x)|)\tilde{P}_{\pi'-\pi}(x)|| \lesssim 1$, and $|\frac{\tilde{P}_{\pi}(x)'\tilde{P}_{\pi}(t)}{V_{\pi}^{1/2}(x)}| \lesssim \zeta_{\bar{K}}\lambda_{\bar{K}}$ under Assumption 3.1 and 3.2. The last inequality uses $\kappa_{1,n} \lesssim \zeta_{\bar{K}}\lambda_{\bar{K}}$, $\kappa_{2,n} \lesssim \sup_{\pi} V_{\pi}(x)$, and $V_{\pi}(x) \asymp V_{\pi'}(x)$ for any $\pi, \pi' \in \Pi$.

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 $\mathcal{F}_n = \{f_{n,\pi} : \pi \in \Pi\}$ satisfy Lipschitz conditions, thus it is VC classes, 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 $\nu(\pi) = \lim_{n\to\infty} -\sqrt{n}V_{\pi}^{-1/2}r_{\pi}$ and the equation (A.5), we have $T_n^*(\pi, \theta_0) \Rightarrow \mathbb{T}(\pi) + \nu(\pi)$ for $\pi \in \Pi$. In addition, if Assumption 3.3 holds, then $|\sqrt{n}V_{\pi}^{-1/2}r_{\pi}| = O(\sqrt{n}V_{\pi}^{-1/2}\ell_{\lfloor\pi\bar{K}\rfloor}c_{\lfloor\pi\bar{K}\rfloor}) = o(1)$ for any $\pi \in \Pi$. Therefore, $T_n^*(\pi, \theta_0) \Rightarrow \mathbb{T}(\pi)$. This completes the proof.

Q.E.D.

A.2 Proof of Theorem 3.2

Proof. We prove the finite dimensional convergence using similar arguments to those used in the proof of Theorem 3.1. We repeat this here, as Assumption 3.4 impose different rates of K compare with the Assumption 3.1. If some elements of $|\nu_m| = +\infty$ 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 profer bounded random vector. Thus, continuous mapping theorem cannot be directly applied to obtain asymptotic distribution results. To circumvent this issue, remaining proof use the same type of argument as in Theorem 1 of Andrews and Guggenberger (2009) in the moment inequality literature.

By Lemma 1 and similar arguments as in Theorem 3.1, 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 $t_n(m) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\tilde{P}_{K_m}(x)'\tilde{P}_{K_m i}\varepsilon_i}{V_{K_m}^{1/2}}$ and $\nu_n(m) = -\sqrt{n} V_{K_m}^{-1/2} r_{K_m}(x)$ is defined in Assumption 3.2. To obtain joint asymptotic distribution of $t_n(m)$, we need to show

$$\delta' t_n \xrightarrow{d} N(0, \delta'\Sigma\delta) \quad \forall \delta \in \mathbb{R}^M \tag{A.22}$$

where $t_n = (t_n(1), ..., t_n(M))'$, $\Sigma_{jl} = \lim_{n\to\infty} \Sigma_{jl,n}$, $\Sigma_{jl,n} \equiv \frac{\tilde{P}_{K_j}(x)'E(\tilde{P}_{K_j i}\tilde{P}'_{K_l i}\varepsilon_i^2)\tilde{P}_{K_l}(x)}{V_{K_j}^{1/2} V_{K_l}^{1/2}}$. Similarly to the proof of Theorem 3.1, we define $\omega_{ni} = (\delta'\Sigma_n\delta)^{-1/2} \sum_{j=1}^M \delta_j \frac{\tilde{P}_{K_j}(x)'\tilde{P}_{K_j i}\varepsilon_i}{V_{K_j}^{1/2}}$. Observe that $E\omega_{ni} = 0$, and $\frac{1}{n}\sum_{i=1}^n E[\omega_{ni}^2] = 1$, and

$$\|\sum_{j=1}^M \frac{\tilde{P}_{K_j}(x)'\tilde{P}_{K_j}}{\sqrt{V_{K_j}}}\|_\infty \lesssim \sum_{j=1}^M \zeta_{K_j}\lambda_{K_j} \lesssim \zeta_{K_M}\lambda_{K_M}$$

by Assumptions 3.2 and 3.4. Lindberg's condition can be verified similarly as in the proof of Theorem 3.1. Therefore, finite dimensional convergence holds by Lindberg-Feller CLT and Slutzky's Theorem.

Next, we let $G(\cdot)$ be a strictly increasing continuous df 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

$$G_{n,m} \xrightarrow{d} G(Z_m + \nu(m)) \tag{A.23}$$

by finite dimensional CLT under Assumptions 3.2, 3.4 and the continuous mapping theorem. If $\nu(m) = +\infty$,

$$G_{n,m} \xrightarrow{p} 1 \tag{A.24}$$

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

$$G_{n,m} \xrightarrow{p} 0 \tag{A.25}$$

as $G(x) \to 0$ as $x \to -\infty$. Since (A.23), (A.24), and (A.25) holds jointly, following holds for any strictly increasing continuous distribution function on $\mathbb{R}$, $G(\cdot)$,

$$G_n \equiv (G_{n,1}, \cdots, G_{n,M})' \xrightarrow{d} G_\infty \equiv (G(Z_1 + \nu(1)), \cdots, G(Z_M + \nu(M)))' \tag{A.26}$$

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 $t = (t_1, \cdots, t_M)' \in \mathbb{R}^{M-1}_{[\pm\infty]} \times \mathbb{R}$, define $G_{(M)}(t) \equiv (G(x_1), \cdots, G(x_M))' \in [0,1]^{M-1} \times (0,1)$. For $y = (y_1, \cdots, y_M)' \in (0,1]^{M-1} \times (0,1)$, define $G^{-1}_{(M)}(y) \equiv (G^{-1}(y_1), \cdots, G^{-1}(y_M))' \in \mathbb{R}^{M-1}_{[\pm\infty]} \times \mathbb{R}$. Define also $S^*(y)$ for $y \in (0,1]^{M-1} \times (0,1)$,

$$S^*(y) \equiv S(G^{-1}_{(M)}(y)). \tag{A.27}$$

Note that $S^*(y)$ is continuous at all $y \in (0,1]^{M-1} \times (0,1)$ since $S(t)$ is continuous at all $t \in \mathbb{R}^{M-1}_{[\pm\infty]} \times \mathbb{R}$. Then, we have

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

where the first equality holds by the definition of $G^{-1}_{(M)}(\cdot)$, the second equality uses the definition of S^*. Convergence in the third line holds by (A.26), and the fourth and fifth equality uses the definition of S^*.

Q.E.D.

A.3 Proof of Corollary 4.1

Proof. Under Assumptions 3.1, 3.2 and $\sup_\pi |\nu(\pi)| < \infty$, we have $T_n^*(\pi, \theta_0) \Rightarrow \mathbb{T}(\pi) + \nu(\pi)$ by Theorem 3.1. Then, $\operatorname{Inf} T_n(\theta_0) = \inf_{K \in \mathcal{K}_n} |T_n(K, \theta_0)| = \inf_{\pi \in \Pi} |T_n^*(\pi, \theta_0)| \xrightarrow{d} \inf_\pi |\mathbb{T}(\pi) + \nu(\pi)|$ holds by continuous mapping theorem. In addition, if Assumption 3.3 holds, $\operatorname{Inf} T_n(\theta_0) \xrightarrow{d} \inf_\pi |\mathbb{T}(\pi)|$ by Theorem 3.1.

For the second part of Corollary, we first define $S(t) \equiv \inf_m |t_m|$ for $t = (t_1, \cdots, t_M) \in \mathbb{R}^{M-1}_{[\pm\infty]} \times \mathbb{R}$. Note that $S(t)$ is continuous at all $t \in \mathbb{R}^{M-1}_{[\pm\infty]} \times \mathbb{R}$ under Assumption 3.4 (especially, assumption of at least one $|\nu_m| = O(1)$) by restricting the domain of functions appropriately. Then, we have

$$\operatorname{Inf} T_n(\theta_0) = S(T_n(\theta_0)) \xrightarrow{d} S(Z + \nu) = \inf_m |Z_m + \nu_m|. \tag{A.28}$$

by Theorem 3.2. If $|\nu_m| = +\infty$, corresponding elements of $|Z_m + \nu_m| = +\infty$ by construction. This completes the proof of Corollary 4.1.

Q.E.D.

A.4 Proof of Corollary 4.2

Proof. We first provide (4.3) in Corollary 4.2.1. Under Assumptions 3.1-3.3, we have shown that $\operatorname{Inf} T_n(\theta_0) \xrightarrow{d} \xi_{\inf} = \inf_{\pi \in [\underline{\pi}, 1]} |\mathbb{T}(\pi)|$ in Corollary 4.1.1. Therefore,

$$\limsup_{n \to \infty} P(\operatorname{Inf} T_n(\theta_0) > c^{\inf}_{1-\alpha}) = \lim_{n \to \infty} P(\operatorname{Inf} T_n(\theta_0) > c^{\inf}_{1-\alpha}) = P(\xi_{\inf} > c^{\inf}_{1-\alpha}) = \alpha$$

where the first equality holds under subsequence $\{u_n\}$ of $\{n\}$ by the definition of $\limsup$, the second equality uses the Corollary 4.1.1 and the definition of $c^{\inf}_{1-\alpha}$ in (4.2). Moreover,

$$\limsup_{n \to \infty} P(\operatorname{Inf} T_n(\theta_0) > z_{1-\alpha/2}) = P(\xi_{\inf} > z_{1-\alpha/2}) \leq P(|\mathbb{T}(\pi)| > z_{1-\alpha/2}) = \alpha$$

where the inequality uses $\xi_{\inf} = \inf_{\pi \in [\underline{\pi}, 1]} |\mathbb{T}(\pi)| \leq |\mathbb{T}(\pi)|$ and $\mathbb{T}(\pi) \overset{d}{=} N(0, 1)$ for any single π.

Next, we prove Corollary 4.2.2. Under Assumptions 3.1, 3.2 and $\sup_\pi |\nu(\pi)| < \infty$, we have $\operatorname{Inf} T_n(\theta_0) \xrightarrow{d} \inf_{\pi\in[\underline{\pi},1]} |\mathbb{T}(\pi)+\nu(\pi)|$ with asymptotic bias $\nu(\pi)$. We have

$$\begin{aligned}
\limsup_{n\to\infty} P(\operatorname{Inf} T_n(\theta_0) > c_{1-\alpha}^{\inf}) &= P(\inf_{\pi\in[\underline{\pi},1]} |\mathbb{T}(\pi)+\nu(\pi)| > c_{1-\alpha}^{\inf}) \\
&\le \inf_{\pi} P(|\mathbb{T}(\pi)+\nu(\pi)| > c_{1-\alpha}^{\inf}) \\
&= \inf_{\pi}[1-(P(Z\le c_{1-\alpha}^{\inf}-|\nu(\pi)|)-P(Z\le -c_{1-\alpha}^{\inf}-|\nu(\pi)|))] \\
&= \inf_{\pi} F(c_{1-\alpha}^{\inf}, |\nu(\pi)|) = F(c_{1-\alpha}^{\inf}, \inf_{\pi}|\nu(\pi)|)
\end{aligned}$$

where the first inequality uses $\inf_{\pi\in[\underline{\pi},1]} |\mathbb{T}(\pi)+\nu(\pi)| \le |\mathbb{T}(\pi)+\nu(\pi)|$ for all π, the second equality uses $\mathbb{T}(\pi) \overset{d}{=} Z \sim N(0,1)$ and the definition of $F(\cdot)$. Finally, the last equality holds since $F(c,|\nu|)$ is monotone increasing function of $|\nu|$. Similarly,

$$\limsup_{n\to\infty} P(\operatorname{Inf} T_n(\theta_0) > z_{1-\alpha/2}) = P(\inf_{\pi\in[\underline{\pi},1]} |\mathbb{T}(\pi)+\nu(\pi)| > z_{1-\alpha/2}) \le F(z_{1-\alpha/2}, \inf_{\pi}|\nu(\pi)|).$$

Corollary 4.2.3 can be similarly derived with $\inf_m |\nu(m)| = 0$ under Assumption 3.4 and using the fact that $F(z_{1-\alpha/2},0)=\alpha$. This completes the proof. (If we further assume $\Sigma = I_M$ in Theorem 3.2, then $\limsup_{n\to\infty} P(\operatorname{Inf} T_n(\theta_0) > c) = \prod_{m=M-M_1+1}^{M} F(c,|\nu(m)|)$ holds for any $0<c<\infty$, by the independence of Z_m, $m=1,...,M$ and when $|\nu(m)|=\infty$ for $m=1,\cdots,M-M_1$ since $F(c,|\nu(m)|)=1$ for $|\nu(m)|=\infty$.) *Q.E.D.*

A.5 Proof of Corollary 4.3

Proof. Under Assumptions 3.2, 3.3, 4.1, and 4.2, following finite dimensional convergence holds by Theorem 3.1,

$$T_n(\theta) = (T_n(K_1,\theta_0),\cdots,T_n(K_M,\theta_0))' \xrightarrow{d} Z = (Z_1,\cdots,Z_M)', \quad Z\sim N(0,\Sigma) \tag{A.29}$$

Under Assumptions 3.2-3.4, above also holds by Theorem 3.2. Note that $T_{n,\widehat{V}}(K,\theta) = \frac{\sqrt{n}(\widehat{\theta}_K-\theta_0)}{\widehat{V}_K^{1/2}} = \frac{V_K^{1/2}}{\widehat{V}_K^{1/2}} T_n(K,\theta)$. Then following holds

$$(T_{n,\widehat{V}}(K_1,\theta_0),\cdots,T_{n,\widehat{V}}(K_M,\theta_0))' = AT_n(\theta) \xrightarrow{d} Z \tag{A.30}$$

by Assumption 4.3 and Slutzky Theorem, where $A \equiv diag\{\frac{V_{K_1}^{1/2}}{\widehat{V}_{K_1}^{1/2}},\cdots,\frac{V_{K_M}^{1/2}}{\widehat{V}_{K_M}^{1/2}}\}$, and $A \xrightarrow{p} I_M$

Next consider $\widehat{c}_{1-\alpha}^{\inf}$ which is $(1-\alpha)$ quantile of $\inf_{m=1,\cdots,M} |Z_{m,\widehat{\Sigma}}|$ defined in (4.10),

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

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

$$\widehat{\Sigma}_{jl} = \frac{\widehat{V}_{K_j}^{1/2}}{\widehat{V}_{K_l}^{1/2}} = \frac{\widehat{V}_{K_j}^{1/2}}{V_{K_j}^{1/2}} \frac{V_{K_j}^{1/2}}{V_{K_l}^{1/2}} \frac{V_{K_l}^{1/2}}{\widehat{V}_{K_l}^{1/2}} \xrightarrow{p} \Sigma_{jl} \tag{A.31}$$

by Assumption 4.3. Therefore, $\widehat{\Sigma} \xrightarrow{p} \Sigma$, $Z_{\widehat{\Sigma}} \xrightarrow{d} Z_{\Sigma}$, and $\inf_{m=1,\cdots,M} |Z_{m,\widehat{\Sigma}}| \xrightarrow{d} \inf_{m=1,\cdots,M} |Z_{m,\Sigma}|$ hold. Thus, $\widehat{c}_{1-\alpha}^{\inf} \xrightarrow{p} c_{1-\alpha}^{\inf}$. *Q.E.D.*

A.6 Proof of Corollary 5.1

Proof. We first show Corollary 5.1.1. Note that $\operatorname{Inf} T_n(\theta_0) = \inf_{K \in \mathcal{K}_n} |T_{n,\widehat{V}}(K, \theta)| \xrightarrow{d} \inf_m |Z_m|$ by Corollary 4.3. We have

$$\begin{aligned}
\liminf_{n\to\infty} P(\theta_0 \in CI_{\texttt{inf}}^{\texttt{Robust}}) &= \liminf_{n\to\infty} P(\operatorname{Inf} T_n(\theta_0) \leq c_{1-\alpha}^{inf} + o_p(1)) \\
&= P(\inf_m |Z_m| \leq c_{1-\alpha}^{inf}) = 1-\alpha
\end{aligned}$$

where the first and the second equality holds by Corollary 4.3 and Corollary 4.1.1 under Assumptions 3.2, 3.3, 4.1, 4.2, and 4.3. Similarly,

$$\liminf_{n\to\infty} P(\theta_0 \in CI_{\texttt{inf}}) = P(\inf_m |Z_m| \leq z_{1-\alpha/2}) \geq P(|Z_m| \leq z_{1-\alpha/2}) = 1-\alpha. \tag{A.32}$$

Corollary 5.1.2 and 5.1.3 can be similarly derived from Corollary 4.2.2 and 4.2.3, respectively. *Q.E.D.*

A.7 Proof of Corollary 6.1

Proof. Similar to the proof of Corollary 4.3, we can also verify $\sup_{m=1,\cdots,M} |Z_{m,\widehat{\Sigma}}| \xrightarrow{d} \sup_{m=1,\cdots,M} |Z_{m,\Sigma}|$, $\widehat{c}_{1-\alpha}^{\sup} \xrightarrow{p} c_{1-\alpha}^{\sup}$, and $\operatorname{Sup} T_n(\theta_0) = \sup_m |T_{n,\widehat{V}}(K_m, \theta_0)| \xrightarrow{d} \sup_m |Z_{m,\Sigma}|$ either under Assumptions

3.2, 3.3, 4.1, 4.2, and 4.3 or under Assumptions 3.2, 3.3, 3.4, and 4.3. Therefore, we have

$$
\begin{aligned}
\liminf_{n\to\infty} P(\theta_0 \in CI^{\texttt{Robust}}_{\texttt{pms}}) &= \liminf_{n\to\infty} P(|T_{n,\widehat{V}}(\widehat{K},\theta_0)| \le \widehat{c}^{sup}_{1-\alpha}) && \text{(A.33)}\\
&\ge \liminf_{n\to\infty} P(\operatorname{Sup} T_n(\theta_0) \le \widehat{c}^{sup}_{1-\alpha}) && \text{(A.34)}\\
&= P(\sup_m |Z_{m,\Sigma}| \le c^{sup}_{1-\alpha}) = 1-\alpha && \text{(A.35)}
\end{aligned}
$$

where the first inequality uses $|T_{n,\widehat{V}}(\widehat{K},\theta_0)| \le \operatorname{Sup} T_n(\theta_0)$ for any $\widehat{K} \in \mathcal{K}_n$. *Q.E.D.*

A.8 Proof of Theorem 7.1

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

$$
\begin{aligned}
&\sqrt{n}(\widehat{\theta}_K - \theta_0) = \widehat{\Gamma}_K^{-1} S_K\\
&\widehat{\Gamma}_K = \frac{1}{n}(W' M_K W), \quad S_K = \frac{1}{\sqrt{n}} W' M_K (g + \varepsilon)
\end{aligned}
$$

where $g = [g_1, \cdots, g_n]', g_i = g_0(x_i), g_w = [g_{w1}, \cdots, g_{wn}]', g_{wi} = g_{w0}(x_i) = E[w_i|x_i], v = [v_1, \cdots, v_n]$.

Under Assumption 7.1 and conditional homoskedastic error terms, $E[v_i^2|x_i] = E[v_i^2]$,

$$
\widehat{\Gamma}_K = \Gamma_K + o_p(1), \quad \Gamma_K = (1 - K/n)E[v_i^2] \tag{A.36}
$$

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

$$
\begin{aligned}
S_K &= \frac{1}{\sqrt{n}} v' M_K \varepsilon + \frac{1}{\sqrt{n}} g_w' M_K g + \frac{1}{\sqrt{n}} (v' M_K g + g_w' M_K \varepsilon) && \text{(A.37)}\\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n M_{K,ii} v_i \varepsilon_i - \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{j=1,j<i}^n P_{K,ij}(v_i \varepsilon_j + v_j \varepsilon_i) + o_p(1) && \text{(A.38)}
\end{aligned}
$$

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 7.1. Under conditional homoskedastic error $E[\varepsilon_i^2|w_i, x_i] = \sigma_\varepsilon^2$ following holds

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

by Theorem 1 of Cattaneo, Jansson and Newey (2015a) which follows from Lemma A2 in

Chao, Swanson, Hausman, Newey and Woutersen (2012).

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

$$\delta_1 T_n(K_1, \theta_0) + \delta_2 T_n(K_2, \theta_0) \xrightarrow{d} N(0, (\delta_1^2 + \delta_2^2 + 2\delta_1\delta_2 v_{12})) \quad \forall(\delta_1, \delta_2) \in \mathbb{R}^2 \tag{A.39}$$

where $v_{12} = \lim_{n\to\infty} V_{K_1}^{1/2}/V_{K_2}^{1/2}$. We closely follows the proof of Lemma A2 in Chao, Swanson, Hausman, Newey and Woutersen (2012). Define $Y_n, Y_{1,n}$ and $Y_{2,n}$ as follows

$$Y_n = \delta_1 Y_{1,n} + \delta_2 Y_{2,n}, \tag{A.40}$$

$$Y_{1,n} = \omega_{1,1n} + \sum_{i=2}^{n} y_{1,in}, \quad y_{1,in} = \omega_{1,in} + \bar{y}_{1,in}, \tag{A.41}$$

$$Y_{2,n} = \omega_{2,1n} + \sum_{i=2}^{n} y_{2,in}, \quad y_{2,in} = \omega_{2,in} + \bar{y}_{2,in}, \tag{A.42}$$

where $\omega_{1,in} = V_{K_1}^{-1/2}\Gamma_{K_1}^{-1} M_{K_1,ii}/\sqrt{n}$, $\bar{y}_{1,in} = \sum_{j<i}(u_{1,j}P_{K_1,ij}\varepsilon_i + u_{1,i}P_{K_1,ij}\varepsilon_j)/\sqrt{n}$, $u_{1,i} = V_{K_1}^{-1/2}\Gamma_{K_1}^{-1} v_i$ and $\omega_{2,in}, \bar{y}_{2,in}$ are similarly defined with appropriate terms $P_{K_2}, V_{K_2}, \Gamma_{K_2}$ with K_2. Similar to the proof of Lemma A2 in Chao, Swanson, Hausman, Newey and Woutersen (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

$$\sum_{i=2}^{n}(\delta_1 y_{1,in} + \delta_2 y_{2,in}) \xrightarrow{d} N(0, \delta_1^2 + \delta_2^2 + 2\delta_1\delta_2 v_{12}). \tag{A.43}$$

It remains to provide Lindeberg-Feller condition.

$$\begin{aligned} E[(\sum_{i=2}^{n} \delta_1 y_{1,in} + \delta_2 y_{2,in})^2|X] = \delta_1^2 E[(\sum_{i=2}^{n} y_{1,in})^2|X] + \delta_2^2 E[(\sum_{i=2}^{n} y_{2,in})^2|X] \\ +2\delta_1\delta_2 E[\sum_{i=2}^{n}\sum_{j=2}^{n} y_{1,in} y_{2,in}|X], \end{aligned} \tag{A.44}$$

where the first and second terms in (A.44) goes to δ_1^2, δ_2^2 a.s., respectively, as in the proof of Lemma A.2 in Chao, Swanson, Hausman, Newey and Woutersen (2012). Note that $E[\omega_{1,in}\bar{y}_{2,jn}|X] = 0, E[\omega_{2,in}\bar{y}_{1,jn}|X] = 0$ for all i, j, 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:

$$E[Y_{1,n}Y_{2,n}|X] = \frac{1}{n}V_{K_1}^{-1/2}\Gamma_{K_1}^{-1}E[v'M_{K_1}\varepsilon v'M_{K_2}\varepsilon|X]\Gamma_{K_2}^{-1}V_{K_2}^{-1/2} \tag{A.45}$$
$$= \frac{1}{n}V_{K_1}^{-1/2}\Gamma_{K_1}^{-1}\sigma_\varepsilon^2 E[v'M_{K_2}v|X]\Gamma_{K_2}^{-1}V_{K_2}^{-1/2} \tag{A.46}$$
$$= V_{K_1}^{-1/2}\Gamma_{K_1}^{-1}\sigma_\varepsilon^2\Gamma_{K_2}\Gamma_{K_2}^{-1}V_{K_2}^{-1/2} \tag{A.47}$$
$$= V_{K_1}^{1/2}/V_{K_2}^{1/2} \tag{A.48}$$

where the second equality uses conditional homoskedasticity $E[\varepsilon^2|X,Z] = \sigma_\varepsilon^2$ 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 $V_{K_1} = \sigma_\varepsilon^2\Gamma_{K_1}^{-1}$. Therefore, we calculate components of last terms in (A.44) as follows

$$E[\sum_{i=2}^n\sum_{j=2}^n y_{1,in}y_{2,in}|X] = E[Y_{1,n}Y_{2,n}|X] - \sum_{i=2}^n E[\omega_{1,1n}y_{2,in}|X]$$
$$- \sum_{i=2}^n E[\omega_{2,1n}y_{1,in}|X] - E[\omega_{1,1n}\omega_{2,1n}|X] \tag{A.49}$$
$$= V_{K_1}^{1/2}/V_{K_2}^{1/2} - E[\omega_{1,1n}\omega_{2,1n}|X] \to v_{12} \quad a.s. \tag{A.50}$$

Also as in the proof of Lemma A.2 of Chao, Swanson, Hausman, Newey and Woutersen (2012), we have

$$\sum_{i=2}^n E[(\delta_1 y_{1,in} + \delta_2 y_{2,in})^4|X] \lesssim \sum_{i=2}^n E[(y_{1,in})^4|X] + \sum_{i=2}^n E[(y_{2,in})^4|X] \to 0 \quad a.s. \tag{A.51}$$

Thus, by similar arguments following the proof of Lemma A.2 in Chao, Swanson, Hausman, Newey and Woutersen (2012), we can apply the martingale central limit theorem. Then, by Slutzky theorem, joint convergence holds with the claimed covariance. By Theorem 2 in Cattaneo, Jansson and Newey (2015a), Assumption 4.3 holds with the following variance estimator for V_K

$$\widehat{V}_K = s^2\widehat{\Gamma}_K^{-1}, \quad s^2 = \frac{1}{n-1-K}\sum_{i=1}^n \widehat{\varepsilon}_i^2, \quad \widehat{\varepsilon}_i^2 = \sum_{j=1}^n M_{K,ij}(y_j - \widehat{\theta}_K w_j). \tag{A.52}$$

Then, we can show the coverage results using similar arguments to those used in the proof of Corollary 5.1. This completes the proof. *Q.E.D.*

B Figures and Tables

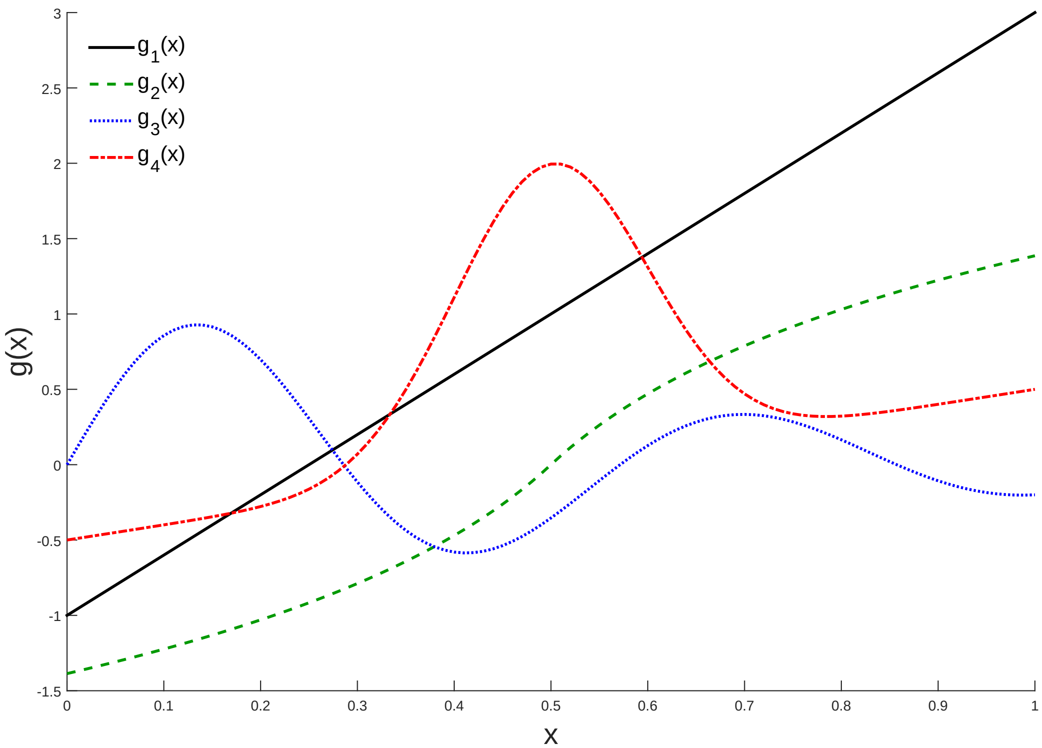


Figure 1: Different functions of $g(x)$.
Solid lines (Black) are $g_1(x) = 4x - 1$; Dashed lines (Green) are $g_2(x) = \ln(|6x - 3| + 1)sgn(x - 1/2)$; Dotted lines (Blue) are $g_3(x) = \sin(7\pi x/2)/[1 + 2x^2(sgn(x) + 1)]$; and Dash-dot lines (Red) are $g_4(x) = x - 1/2 + 5\phi(10(x - 1/2))$, where $\phi(\cdot)$ is standard normal pdf.

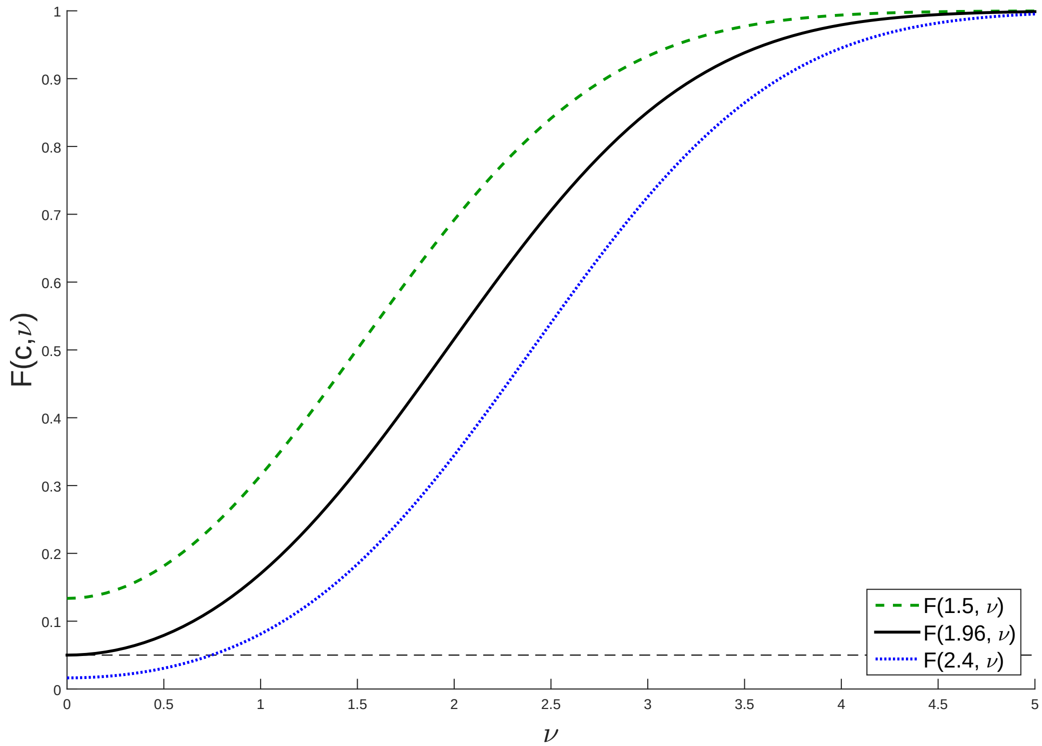


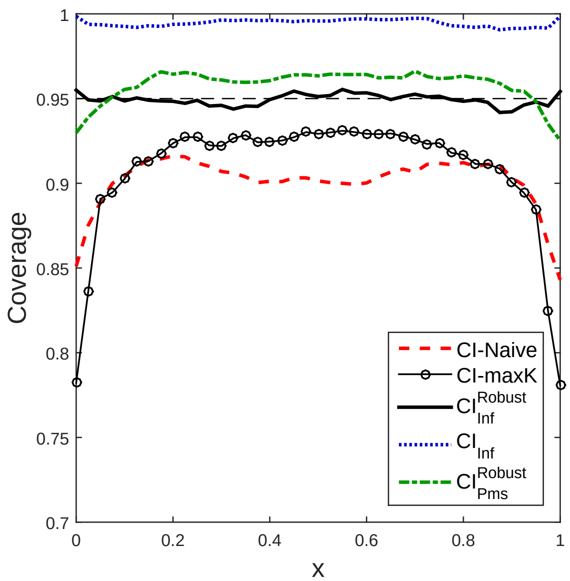
Figure 2: Plots of $F(c, \nu)$ as a function of ν for c = 1.5, 1.96, 2.4.

Figure 3: Coverage - Polynomials

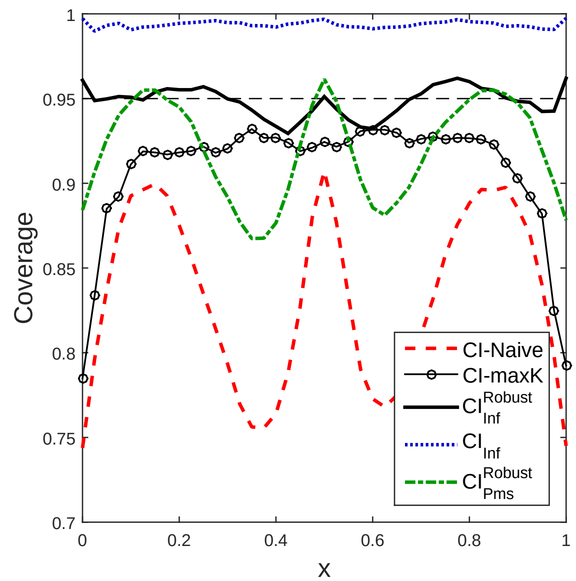
Nominal 95% Coverage of Various CIs for $g(x)$:

(1) $CI^{\texttt{Naive}}_{\texttt{pms}}$ with $\widehat{K}_{\texttt{cv}}$ (2) $CI_{\texttt{maxK}}$ with $\bar{K}$ (3) $CI^{\texttt{Robust}}_{\texttt{inf}}$ (4) $CI_{\texttt{inf}}$ (5) $CI^{\texttt{Robust}}_{\texttt{pms}}$ with $\widehat{K}_{\texttt{cv}}$

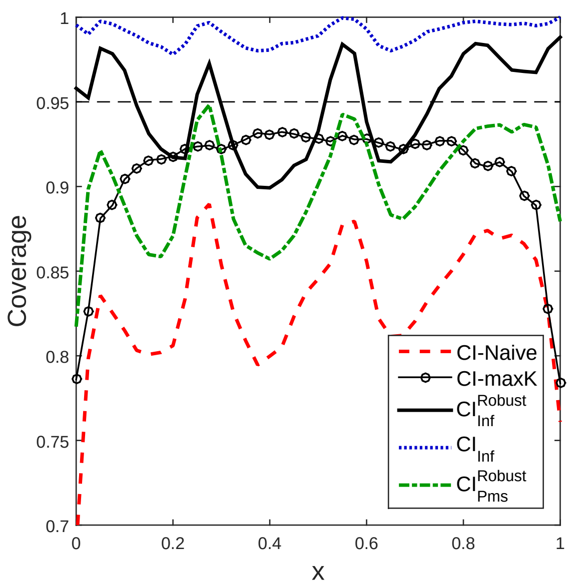
(a) $g_1(x) = 4x - 1$



(b) $g_2(x) = \ln(|6x-3|+1)sgn(x-1/2)$



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



(d) $g_4(x) = x - 1/2 + 5\phi(10(x-1/2))$

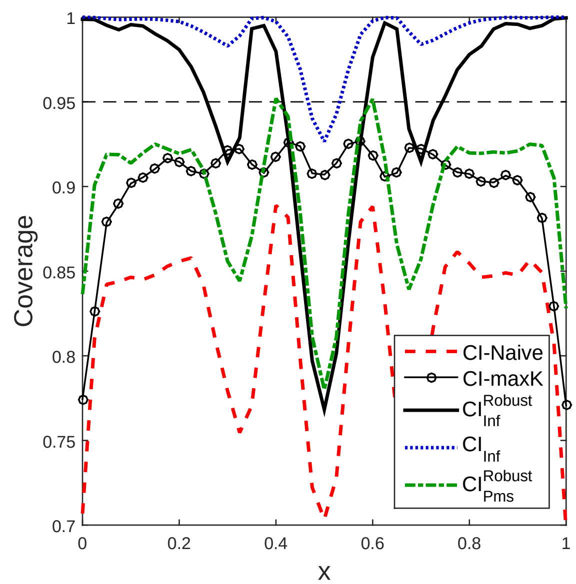
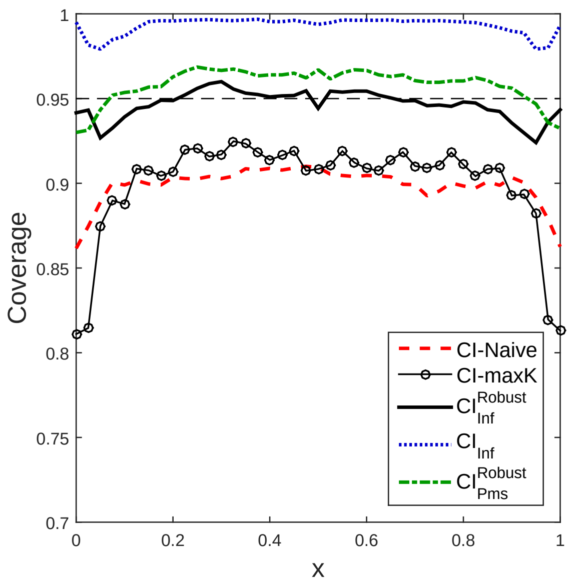


Figure 4: Coverage - Splines

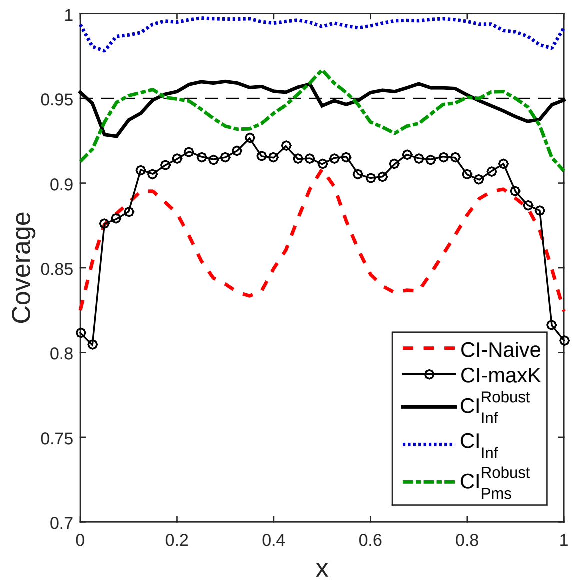
Nominal 95% Coverage of Various CIs for $g(x)$:

(1) $CI^{\texttt{Naive}}_{\texttt{pms}}$ with $\widehat{K}_{\texttt{cv}}$ (2) $CI_{\texttt{maxK}}$ with $\bar{K}$ (3) $CI^{\texttt{Robust}}_{\texttt{inf}}$ (4) $CI_{\texttt{inf}}$ (5) $CI^{\texttt{Robust}}_{\texttt{pms}}$ with $\widehat{K}_{\texttt{cv}}$

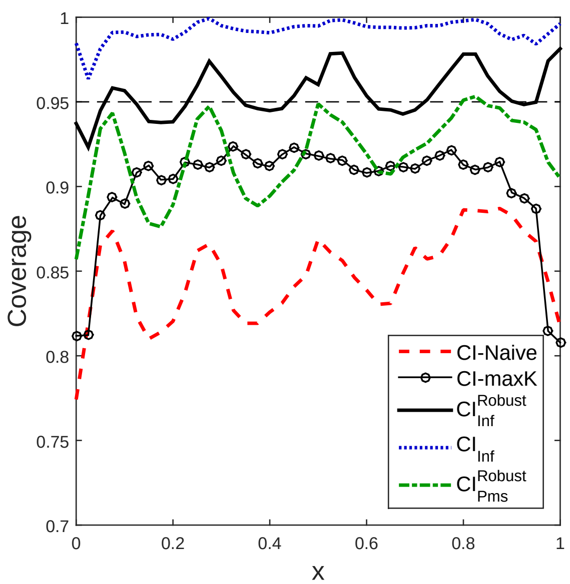
(a) $g_1(x) = 4x - 1$



(b) $g_2(x) = \ln(|6x-3|+1)sgn(x-1/2)$



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



(d) $g_4(x) = x - 1/2 + 5\phi(10(x-1/2))$

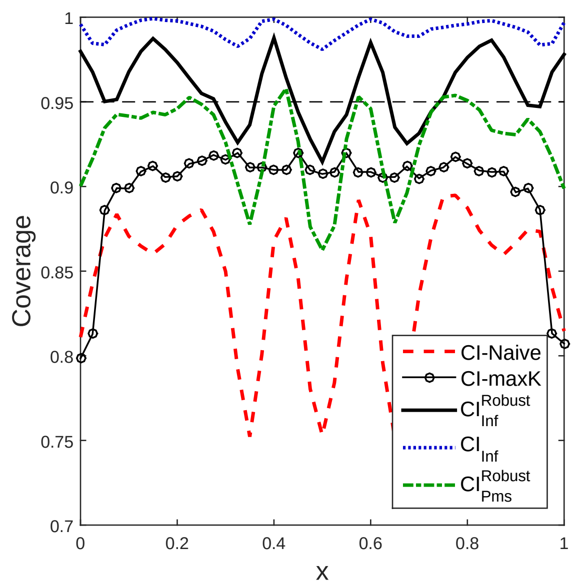
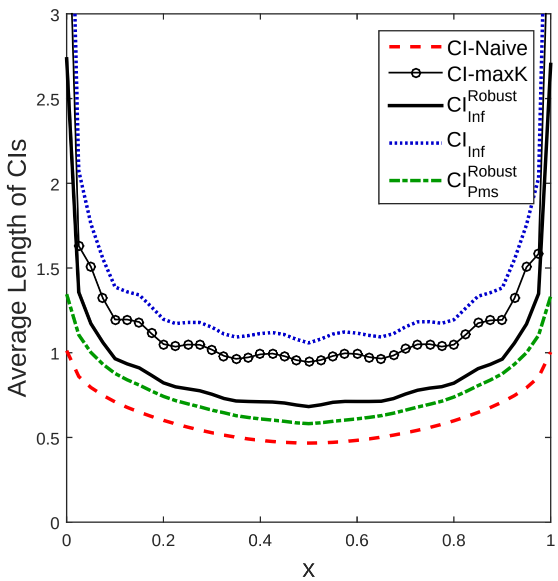
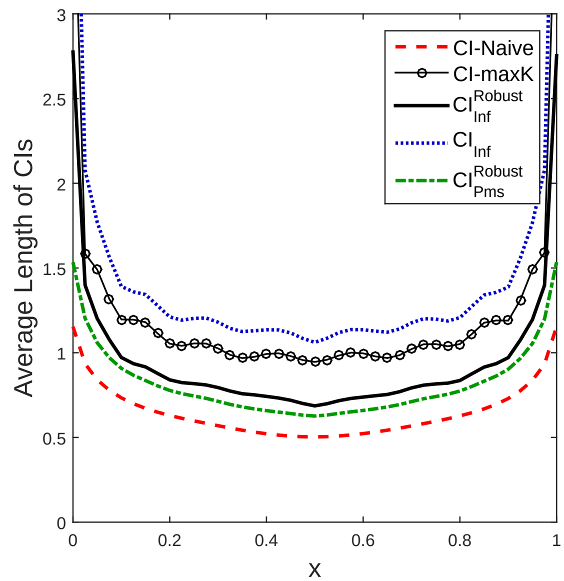


Figure 5: Length of CIs - Polynomials
Average lengths of nominal 95% CIs for $g(x)$:
(1) $CI^{\texttt{Naive}}_{\texttt{pms}}$ with $\widehat{K}_{\texttt{cv}}$ (2) $CI_{\texttt{maxK}}$ with $\bar{K}$ (3) $CI^{\texttt{Robust}}_{\texttt{inf}}$ (4) $CI_{\texttt{inf}}$ (5) $CI^{\texttt{Robust}}_{\texttt{pms}}$ with $\widehat{K}_{\texttt{cv}}$

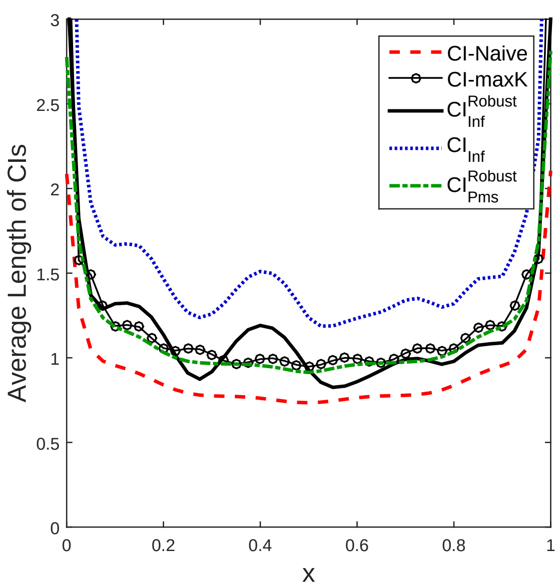
(a) $g_1(x) = 4x - 1$



(b) $g_2(x) = \ln(|6x-3|+1)sgn(x-1/2)$



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



(d) $g_4(x) = x - 1/2 + 5\phi(10(x-1/2))$

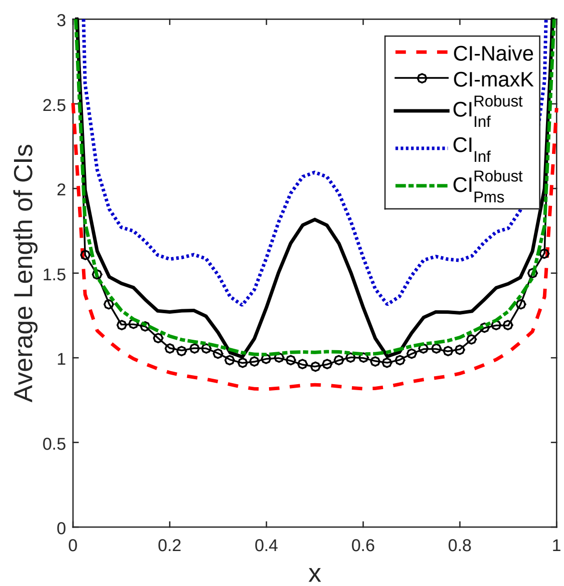
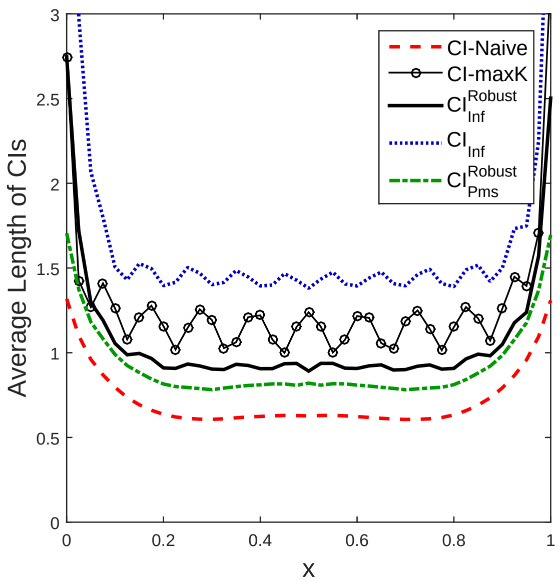


Figure 6: Length of CIs - Splines

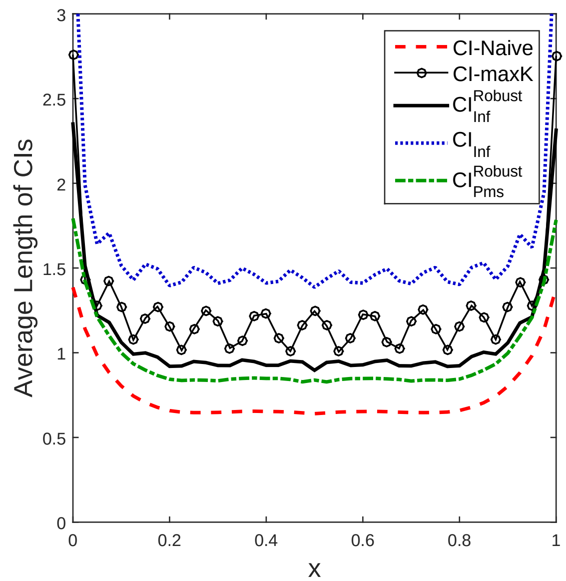
Average lengths of nominal 95% CIs for $g(x)$:

(1) $CI^{\texttt{Naive}}_{\texttt{pms}}$ with $\widehat{K}_{\texttt{cv}}$ (2) $CI_{\texttt{maxK}}$ with $\bar{K}$ (3) $CI^{\texttt{Robust}}_{\texttt{inf}}$ (4) $CI_{\texttt{inf}}$ (5) $CI^{\texttt{Robust}}_{\texttt{pms}}$ with $\widehat{K}_{\texttt{cv}}$

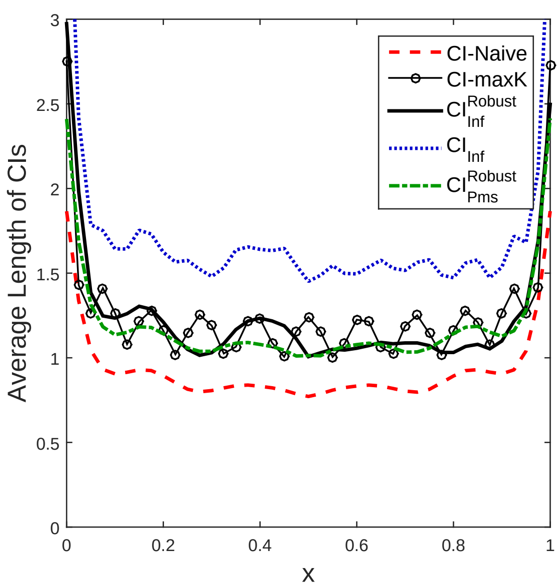
(a) $g_1(x) = 4x - 1$



(b) $g_2(x) = \ln(|6x - 3| + 1)sgn(x - 1/2)$



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



(d) $g_4(x) = x - 1/2 + 5\phi(10(x - 1/2))$

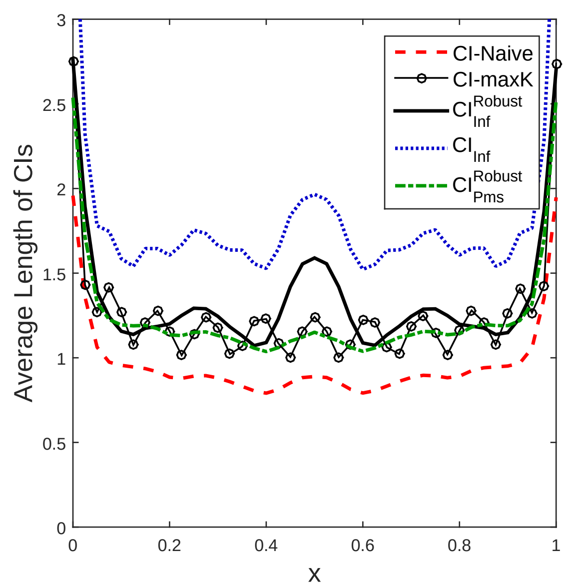
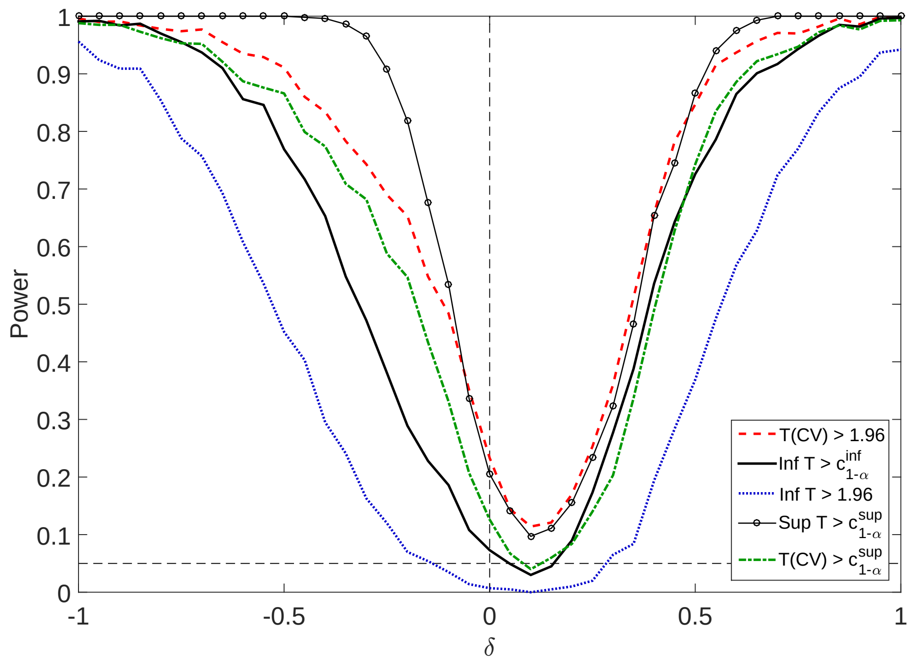


Figure 7: Power function against fixed alternatives. Design 2 : $g_2(x) = \ln(|6x - 3| + 1)sgn(x - 1/2)$. $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_0 + \delta$, where $\theta_0 = g_2(x)$ at $x = 0.4$ for figure (a) and $x = 0.5$ for figure (b). Using Polynomials.

(a) $\theta_0 = g_2(x), x = 0.4$



(b) $\theta_0 = g_2(x), x = 0.5$

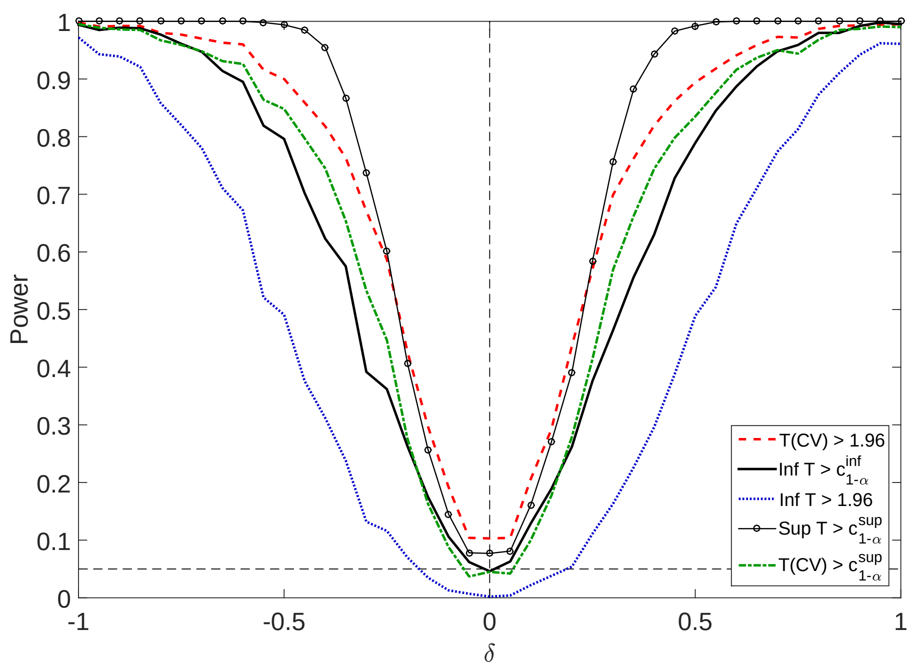


Table 1: Nonparametric Wage Elasticity of Hours of Work Estimates in Blomquist and Newey (Table 1, 2002). Wage elasticity evaluated at the mean wage and income

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]
$\Delta y w$	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]
$\ell \Delta y^2, \ell \Delta w^2, \ell \Delta y w$	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: $\widehat{c}^{\,\mathrm{inf}}_{1-\alpha} = 0.9668$, $\widehat{c}^{\,\mathrm{sup}}_{1-\alpha} = 2.4764$				
Test $H_0 : E_w = 0$, $\quad \mathrm{Inf}\, T_n(\theta_0) = 2.4706 > \widehat{c}^{\,\mathrm{inf}}_{1-\alpha}$				
$CI^{\texttt{Robust}}_{\texttt{inf}} = [0.0271, 0.1111]$ $CI_{\texttt{inf}} = [0.0148, 0.1384]$, $\quad CI^{\texttt{Robust}}_{\texttt{pms}} = [0.0169, 0.0916]$				

[1] y : non-labor income, w : marginal wage rates, ℓ: the end point of the segment in a piecewise linear budget set.

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

C Supplementary material

The supremum of the t-statistics and confidence intervals uniform in the number of series terms

In this supplementary material, we consider the supremum of the t-statistics over all series terms and discuss more about inference methods based on this test statistic. In another direction, this paper also derives the robust inference method after searching over different specifications for nonparametric series estimation.

Suppose a researcher reports only 'favorable' subset of positive results and hiding large different specifications which shows overall mixed results or pretending not to search. These practices may lead to distorted inference and the misleading conclusion if we take variability of the first step specification search into account. For example, if a researcher computes many t-statistics and chooses the largest one, then usual standard normal critical value must be adjusted to control size. The importance of specification search (or data mining/ data snooping) has long been alerted in various other contexts (see Leamer (1983), White (2000), Romano and Wolf (2005), Hansen (2005), and recent papers by Varian (2014), Athey and Imbens (2015), and Armstrong and Kolesár (2015)). Considering the supremum statistic is quite natural to control size of the joint test in multiple testing literature.

Specification search is widely used in estimating the parametric model in a less clear way. Although, nonparametric series estimation gives systematic way of doing specification search by restricting domain of search as $K \in [\underline{K}, \bar{K}]$, little justification has been done, especially for the inference problems. Here, we introduce the tests based on the supremum of the t-statistics over all series terms using the critical values from its asymptotic distribution. We show that this also controls size with undersmoothing conditions. This tests can be used to construct CIs which are uniform in K that have a correct coverage. That is, all confidence intervals using the critical value from supremum t-statistics jointly cover the true parameter at the nominal level, asymptotically. This robust inference method is one way to improve the credibility of inference by admitting search over large sets of different models in nonparametric regression and doing some corrections as usual in multiple testing literature.

We consider a following 'supremum' t-statistic

$$\operatorname{Sup} T_n(\theta) = \sup_{K \in \mathcal{K}_n} |T_n(K, \theta)|. \tag{C.1}$$

The supremum of the t-statistics is appropriate in the context of multiple testing, and is known to control the size of the family wise error rate (FWE). We may consider the specification search over large sets of $\mathcal{K}_n$ as simultaneously testing a single hypothesis H_0

based on different test statistics $T_n(K,\theta)$ over $K \in \mathcal{K}_n$. Multiple testing setup is more natural when we focus on the pseudo-true parameter θ_K, i.e., the best linear approximation for $g_0(x)$. One can consider simultaneous testing of individual hypothesis $H_{K,0} : \theta_K = \theta_0$ vs $H_{K,1} : \theta_K \neq \theta_0$ for different $K \in \mathcal{K}_n$. Here, controlling FWE corresponds to control following probability asymptotically, $FWE = P(\text{reject at least one hypothesis } H_{K,0}, K \in \mathcal{K}_n) \leq \alpha$.

To derive asymptotic size of the test and coverage of CI based on the $\operatorname{Sup} T_n(\theta)$, we first provide asymptotic null limiting distribution of the supremum statistics analogous to the Corollary 1 for the infimum test statistic, $\operatorname{Inf} T_n(\theta)$.

Corollary C.1. *1. Under Assumptions 3.1-3.2 and* $\sup_\pi |\nu(\pi)| < \infty$, $\operatorname{Sup} T_n(\theta_0) \xrightarrow{d} \sup_{\pi \in [\underline{\pi},1]} |\mathbb{T}(\pi) + \nu(\pi)|$, *where* $\mathbb{T}(\pi)$ *is the mean zero Gaussian process defined in Theorem 3.1. In addition, if Assumption 3.3 holds, then* $\operatorname{Sup} T_n(\theta_0) \xrightarrow{d} \xi_{\sup} = \sup_{\pi \in [\underline{\pi},1]} |\mathbb{T}(\pi)|$.

2. *Suppose Assumptions 3.2 and 3.4 hold. In addition, if* $\sup_m |\nu(m)| < \infty$ *are satisifed, then* $\operatorname{Sup} T_n(\theta_0) \xrightarrow{d} \sup_{m=1,\cdots,M} |Z_m + \nu(m)|$ *where* Z_m *is an element of* $M \times 1$ *normal vector* $Z \sim N(0, I_M)$ *and* $\nu = (\nu(1), \cdots, \nu(M))'$ *defined in Theorem 3.2. If* $\sup_m |\nu(m)| = \infty$, *then* $\operatorname{Sup} T_n(\theta_0) \xrightarrow{p} \infty$.

Corollary C.1.2 shows that $\operatorname{Sup} T_n(\theta_0)$ converges in probability to infinity under alternative set Assumption 3.4. This implies that the supremum of the t-statistics can be sensitive to those oversmoothing sequences (small K) with high bias. Next Corollary provides the asymptotic size of the test based on $\operatorname{Sup} T_n(\theta)$ similar to Corollary 4.2.

Corollary C.2. *1. Under Assumptions 3.1-3.3, following holds*

$$\limsup_{n \to \infty} P(\operatorname{Sup} T_n(\theta_0) > c_{1-\alpha}^{\sup}) = \alpha. \tag{C.2}$$

2. *Under Assumptions 3.1-3.2, and* $\sup_\pi |\nu_\pi| < \infty$, *following holds*

$$\limsup_{n \to \infty} P(\operatorname{Sup} T_n(\theta_0) > c_{1-\alpha}^{\sup}) \geq F(c_{1-\alpha}^{\sup}, \sup_\pi |\nu(\pi)|) \tag{C.3}$$

where $F(c, |\nu|) = 1 - \Phi(c - |\nu|) + \Phi(-c - |\nu|)$ *with standard normal cumulative distribution function* $\Phi(\cdot)$.

3. *Under Assumptions 3.2, 3.4, and* $\sup_m |\nu(m)| = \infty$, $\limsup_{n \to \infty} P(\operatorname{Sup} T_n(\theta_0) > c) = 1$ *for any* $0 < c < \infty$.

Contrary to the $\operatorname{Inf} T_n(\theta)$ test statistic,(C.3) in Corollary C.2.2 shows that the test based on $\operatorname{Sup} T_n(\theta)$ can be sensitive to the large asymptotic bias, and this leads to the over-rejection of the test. Suppose $F(c^{\text{sup}}_{1-\alpha}, q) = \alpha$ for some $q > 0$. If $\sup_\pi |\nu(\pi)| > q$, then the asymptotic size is strictly greater than α. This also can be seen from the results in C.2.3. If $|\nu(m)| = \infty$ for any m, then the asymptotic size of the test is equal to 1.

Next, we define CI_{sup} based on $\operatorname{Sup} T_n(\theta)$ and the critical value $\widehat{c}^{\text{sup}}_{1-\alpha}$ in Section 6.

$$\begin{aligned} CI_{\mathsf{sup}} &\equiv \{\theta : \sup_{K\in\mathcal{K}_n} |T_{n,\widehat{V}}(K,\theta)| \le \widehat{c}^{\text{sup}}_{1-\alpha}\} \\ &= \bigcap_{K\in\mathcal{K}_n} \{\theta : |T_{n,\widehat{V}}(K,\theta)| \le \widehat{c}^{\text{sup}}_{1-\alpha}\} = [\sup_K(\widehat{\theta}_K - \widehat{c}^{\text{sup}}_{1-\alpha} s(\widehat{\theta}_K)), \inf_K(\widehat{\theta}_K + \widehat{c}^{\text{sup}}_{1-\alpha} s(\widehat{\theta}_K))]. \end{aligned} \tag{C.4}$$

Note that CI_{sup} is an intersection of all CIs in $\mathcal{K}_n$ using critical value $\widehat{c}^{\text{sup}}_{1-\alpha}$.

Corollary C.3. *1. Under Assumptions 3.2, 4.1, 4.2, and 4.3,*

$$\liminf_{n\to\infty} P(\theta_K \in [\widehat{\theta}_K \pm \widehat{c}^{\text{sup}}_{1-\alpha} s(\widehat{\theta}_K)] \quad \forall K \in \mathcal{K}_n) = 1 - \alpha. \tag{C.5}$$

In addition, if Assumption 3.3 (undersmoothing) holds,

$$\liminf_{n\to\infty} P(\theta_0 \in CI_{sup}) = \liminf_{n\to\infty} P(\theta_0 \in CI_K = [\widehat{\theta}_K \pm \widehat{c}^{\text{sup}}_{1-\alpha} s(\widehat{\theta}_K)] \quad \forall K \in \mathcal{K}_n) = 1 - \alpha. \tag{C.6}$$

2. Under Assumptions 3.2, 4.1, 4.2, 4.3, and $\sup_m |\nu(m)| < \infty$,

$$\liminf_{n\to\infty} P(\theta_0 \in CI_{sup}) \le 1 - F(c^{\text{sup}}_{1-\alpha}, \sup_m |\nu(m)|). \tag{C.7}$$

3. Under Assumptions 3.2, 3.4, 4.3, and $\sup_m |\nu(m)| = \infty$, $\liminf_{n\to\infty} P(\theta_0 \in CI_{sup}) = 0$.

By using an appropriate critical value from the distribution of $\operatorname{Sup} T_n(\theta)$, (C.5) gives asymptotic coverage of the uniform confidence intervals over $K \in \mathcal{K}_n$ for the pseudo-true value θ_K. (C.6) gives asymptotic coverage probability of CI_{sup} for the true value θ_0 with undersmoothing assumption, which is same as joint coverage of uniform confidence intervals over $K \in \mathcal{K}_n$.

Corollary C.3.2 and C.3.3 show that the coverage can be sensitive to the asymptotic bias. Especially, uniform coverage results based on $\operatorname{Sup} T_n(\theta)$ in (C.6) can be highly sensitive to some small $K \in \mathcal{K}$ has large asymptotic bias, so that the coverage probability can be far below than the nominal level. Recall that CI_{sup} is constructed by intersection of all confidence intervals in $\mathcal{K}_n$ using larger critical value $\widehat{c}_{1-\alpha}^{\sup}$ than the normal critical value. Intersection can give tighter CI, however, if one of the estimator has a large bias, resulting CI can be too narrow to cover the true parameter. In worst scenario, intersection can be empty sets so that the coverage of uniform CIs can be 0. This was formally stated in C.3.3. Under Assumption 3.4, if $|\nu(m)| = \infty$ for some m then asymptotic coverage probability of CI_{sup} is exactly 0.

C.1 Proof of the results in Section C

C.1.1 Proof of Corollary C.1

Proof. The first part follows from Theorem 3.1 and continuous mapping theorem similar to the proof of Corollary 4.1. For the second part of Corollary C.1, consider $S(t) = \sup_m |t_m|$ for $t = (t_1, \cdots, t_M)$ similarly as in the proof of Corollary 4.1. We have

$$\operatorname{Sup} T_n(\theta_0) = \sup_m |T_n(K_m, \theta_0)| = S(T_n(\theta_0)). \tag{C.8}$$

If $\sup_m |\nu(m)| < \infty$, $S(t)$ is continuous at all $t \in \mathbb{R}^M$. Therefore, following holds

$$\operatorname{Sup} T_n(\theta_0) \xrightarrow{d} S(Z + \nu) = \sup_m |Z_m + \nu(m)| \tag{C.9}$$

by Theorem 3.2. If $|\nu_m| = +\infty$ for some m, then then $|T_n(K_m, \theta_0)| \xrightarrow{p} +\infty$, therefore $\operatorname{Sup} T_n(\theta_0) \xrightarrow{p} +\infty$. *Q.E.D.*

C.1.2 Proof of Corollary C.2

Proof. First, we observe that $|T_n(\widehat{K}, \theta_0| \leq \operatorname{Sup} T_n(\theta_0)$ for any $\widehat{K} \in \mathcal{K}_n$. Then we have

$$\limsup_{n\to\infty} P(|T_n(\widehat{K}, \theta)| > c_{1-\alpha}^{\sup}) \leq \limsup_{n\to\infty} P(\operatorname{Sup} T_n(\theta_0) > c_{1-\alpha}^{\sup}) = P(\xi_{\sup} > c_{1-\alpha}^{\sup}) = \alpha$$

by Corollary C.1.1. Next, without assuming Assumption 3.3, we have

$$\begin{aligned}
\limsup_{n\to\infty} P(\operatorname{Sup} T_n(\theta_0) > c_{1-\alpha}^{\sup}) &= P(\sup_{\pi\in[\underline{\pi},1]} |\mathbb{T}(\pi) + \nu(\pi)| > c_{1-\alpha}^{\sup}) \\
&= 1 - P(\sup_{\pi\in[\underline{\pi},1]} |\mathbb{T}(\pi) + \nu(\pi)| \leq c_{1-\alpha}^{\sup}) \\
&\geq \sup_{\pi} [1 - P(|\mathbb{T}(\pi) + \nu(\pi)| \leq c_{1-\alpha}^{\sup})] \\
&= \sup_{\pi} F(c_{1-\alpha}^{\sup}, |\nu(\pi)|) = F(c_{1-\alpha}^{\sup}, \sup_{\pi} |\nu(\pi)|)
\end{aligned}$$

where the first inequality uses $P(\sup_{\pi\in[\underline{\pi},1]}|\mathbb{T}(\pi)+\nu(\pi)| \leq c^{\text{sup}}_{1-\alpha}) \leq P(|\mathbb{T}(\pi)+\nu(\pi)| \leq c^{\text{sup}}_{1-\alpha})$ for all π. The third and last equality use the definition of F and monotone increasing property of $F(c,|\nu|)$ with respect to $|\nu|$.

Next, we consider Corollary C.2.3 under alternative set assumption. If $\sup_m|\nu(m)| = \infty$, then $\text{Sup}\,T_n(\theta_0) \xrightarrow{p} +\infty$ by Corollary C.1.2. Thus, $\limsup_{n\to\infty} P(\text{Sup}\,T_n(\theta_0) > c) = 1$ since $F(c,\infty) = 1$ for any $0 < c < \infty$. *Q.E.D.*

C.1.3 Proof of Corollary C.3

Proof. This follows from Corollary 4.3 and Corollary C.2 similar to the proof of Corollary 6.1. Recall that the t-statistic can be written as

$$T_{n,\widehat{V}}(K,\theta_0) = \frac{\sqrt{n}(\widehat{\theta}_K - \theta_0)}{\widehat{V}_K^{1/2}} = \frac{\sqrt{n}(\widehat{\theta}_K - \theta_K)}{\widehat{V}_K^{1/2}} + \frac{\sqrt{n}r_K}{\widehat{V}_K^{1/2}} \tag{C.10}$$

First, consider (C.5),

$$\liminf_{n\to\infty} P(\theta_K \in [\widehat{\theta}_K \pm \widehat{c}^{\text{sup}}_{1-\alpha}s(\widehat{\theta}_K)] \quad \forall K \in \mathcal{K}_n) \tag{C.11}$$

$$= \liminf_{n\to\infty} P(|\frac{\sqrt{n}(\widehat{\theta}_K - \theta_K)}{\widehat{V}_K^{1/2}}| \leq \widehat{c}^{sup}_{1-\alpha} \quad \forall K \in \mathcal{K}_n) = \liminf_{n\to\infty} P(\sup_K|\frac{\sqrt{n}(\widehat{\theta}_K - \theta_K)}{\widehat{V}_K^{1/2}}| \leq \widehat{c}^{sup}_{1-\alpha}) \tag{C.12}$$

$$= P(\sup_m|Z_m| \leq c^{sup}_{1-\alpha}) = 1-\alpha \tag{C.13}$$

where the last equality follows from Theorem 3.1 and Corollary 4.3 under Assumptions 3.2, 4.1, 4.2, and 4.3. In addition, if Assumption 3.3 holds, we have that

$$\liminf_{n\to\infty} P(\theta_0 \in CI_{\texttt{sup}}) = \liminf_{n\to\infty} P(\text{Sup}\,T_n(\theta_0) \leq \widehat{c}^{sup}_{1-\alpha}) \tag{C.14}$$

$$(= \liminf_{n\to\infty} P(|T_{n,\widehat{V}}(K,\theta_0)| \leq \widehat{c}^{sup}_{1-\alpha} \quad \forall K \in \mathcal{K}_n)) \tag{C.15}$$

$$= P(\sup_m|Z_m| \leq c^{sup}_{1-\alpha}) = 1-\alpha. \tag{C.16}$$

This completes the first part of Corollary C.3. The second part can be shown similarly to the proof of Corollary C.2.2. For the last part, if $\sup_m|\nu(m)| = \infty$, then $\liminf_{n\to\infty} P(\theta_0 \in CI_{\texttt{sup}}) = 0$ by Corollary C.2.3 and Corollary 4.3.

Q.E.D.