SIEVE INFERENCE ON SEMI-NONPARAMETRIC TIME SERIES MODELS

By

Xiaohong Chen, Zhipeng Liao and Yixiao Sun

February 2012

COWLES FOUNDATION DISCUSSION PAPER NO. 1849

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Sieve Inference on Semi-nonparametric Time Series Models*

Xiaohong Chen†, Zhipeng Liao‡, and Yixiao Sun§

First Draft: October 2009; This draft: February 2012

Abstract

The method of sieves has been widely used in estimating semiparametric and nonparametric models. In this paper, we first provide a general theory on the asymptotic normality of plug-in sieve M estimators of possibly irregular functionals of semi/nonparametric time series models. Next, we establish a surprising result that the asymptotic variances of plug-in sieve M estimators of irregular (i.e., slower than root-T estimable) functionals do not depend on temporal dependence. Nevertheless, ignoring the temporal dependence in small samples may not lead to accurate inference. We then propose an easy-to-compute and more accurate inference procedure based on a "pre-asymptotic" sieve variance estimator that captures temporal dependence. We construct a "pre-asymptotic" Wald statistic using an orthonormal series long run variance (OS-LRV) estimator. For sieve M estimators of both regular (i.e., root-T estimable) and irregular functionals, a scaled "pre-asymptotic" Wald statistic is asymptotically F distributed when the series number of terms in the OS-LRV estimator is held fixed. Simulations indicate that our scaled "pre-asymptotic" Wald test with F critical values has more accurate size in finite samples than the usual Wald test with chi-square critical values.

Keywords: Weak Dependence; Sieve M Estimation; Sieve Riesz Representor; Irregular Functional; Misspecification; Pre-asymptotic Variance; Orthogonal Series Long Run Variance Estimation; F Distribution

---

*We acknowledge useful comments from T. Christensen, J. Hahn, L. Hansen, J. Hidalgo, M. Jansson, D. Kaplan, O. Linton, P. Phillips, D. Pouzo, J. Powell, H. White, and other participants at 2011 Econometrics Society Australasian Meeting in Adelaide, the Pre-conference Workshop of 2011 Asian Meeting of the Econometric Society in Seoul, econometrics workshops at Yale, UC Berkeley, UCLA, UCSD and Stanford. Chen and Sun acknowledge financial support from the National Science Foundation under the respective grant numbers: SES-0838161 and SES-0752443. Any errors are the responsibility of the authors.

†Department of Economics, Yale University, 30 Hillhouse, Box 208281, New Haven, CT 06520. Email: xiaohong.chen@yale.edu

‡Department of Economics, UC Los Angeles, 8379 Bunche Hall, Mail Stop: 147703, Los Angeles, CA 90095. Email: zhipeng.liao@econ.ucla.edu

§Department of Economics, UC San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508. Email: yisun@ucsd.edu

1 Introduction

Many economic and financial time series (and panel time series) are nonlinear and non-Gaussian; see, e.g., Granger (2003). For policy and welfare analysis, it is important to uncover complicated nonlinear economic relations in dynamic structural models. Unfortunately, it is difficult to correctly parameterize nonlinear dynamic functional relations. Even if the nonlinear functional relation among the observed variables is correctly specified by economic theory or by chance, misspecifying distributions of nonseparable latent variables could lead to inconsistent estimates of structural parameters of interest. These reasons, coupled with the availability of larger data sets, motivate the growing popularity of semiparametric and nonparametric models and methods in economics and finance.

The method of sieves (Grenander, 1981) is a general procedure for estimating semiparametric and nonparametric models, and has been widely used in economics, finance, statistics and other disciplines. In particular, the method of sieve extremum estimation optimizes a random criterion function over a sequence of approximating parameter spaces, *sieves*, that becomes dense in the original infinite dimensional parameter space as the complexity of the sieves grows to infinity with the sample size T. See, e.g., Chen (2007, 2011) for detailed reviews of some well-known empirical applications of the method and existing theoretical properties of sieve extremum estimators.

In this paper, we consider inference on possibly misspecified semi-nonparametric time series models via the method of sieves. We focus on sieve M estimation, which optimizes a sample average of a criterion function over a sequence of finite dimensional sieves whose complexity grows to infinity with the sample size T. Prime examples include sieve quasi maximum likelihood, sieve (nonlinear) least squares, sieve generalized least squares, and sieve quantile regression. For general sieve M estimators with weakly dependent data, White and Wooldridge (1991) establish the consistency, and Chen and Shen (1998) establish the convergence rate and the $\sqrt{T}$ asymptotic normality of plug-in sieve M estimators of regular (i.e., $\sqrt{T}$ estimable) functionals. To the best of our knowledge, there is no published work on the limiting distributions of plug-in sieve M estimators of irregular (i.e., slower than $\sqrt{T}$ estimable) functionals. There is also no published inferential result for general sieve M estimators of regular or irregular functionals for possibly misspecified semi-nonparametric time series models.

We first provide a general theory on the asymptotic normality of plug-in sieve M estimators of possibly irregular functionals in semi/nonparametric time series models. This result extends that of Chen and Shen (1998) for sieve M estimators of regular functionals to sieve M estimators of irregular functionals. It also extends that of Chen and Liao (2008) for sieve M estimators of irregular functionals with iid data to time series settings. The asymptotic normality result is rate-adaptive in the sense that researchers do not need to know *a priori* whether the functional of interest is $\sqrt{T}$ estimable or not.

For weakly dependent data and for regular functionals, it is known that the asymptotic variance expression depends on the temporal dependence and is usually equal to the long run variance (LRV) of a scaled moment (or score) process. It is often believed that this result would also hold for sieve estimators of irregular functionals such as the evaluation functionals and weighted integration functionals. Contrary to this common belief, we show that under some general conditions the asymptotic variance of the plug-in sieve estimator for weakly dependent data is the same as that for iid data. This is a very surprising result, as sieve estimators are often regarded as global estimators, and hence autocorrelation is not expected to vanish in the limit (as $T \to \infty$).

Our asymptotic theory suggests that, for weakly dependent time series data with a large sample size, temporal dependence could be ignored in making inference on irregular functionals via the method of sieves. This resembles the earlier well-known asymptotic results for time series density and regression functions estimated via kernel and local polynomial regression methods. See, e.g., Robinson (1983), Fan and Yao (2003), Li and Racine (2007), Gao (2007) and the references therein. However, simulation studies indicate that inference procedures based on asymptotic variance estimates ignoring autocorrelation may not perform well when the sample size is small (relatively to the degree of temporal dependence). See, e.g., Conley, Hansen and Liu (1997) and Pritsker (1998) for earlier discussion of this problem with kernel density estimation for interest rate data sets.

In this paper, for both regular and irregular functionals of semi-nonparametric time series models, we propose computationally simple, accurate and robust inference procedures based on estimates of "pre-asymptotic" sieve variances capturing temporal dependence. That is, we treat the underlying triangular array sieve score process as a generic time series and ignore the fact that it becomes less temporally dependent when the sieve number of terms in approximating unknown functions grows to infinity as T goes to infinity. This "pre-asymptotic" approach enables us to conduct easy-to-compute and accurate inference on semi-nonparametric time series models by adopting any existing autocorrelation robust inference procedures for (misspecified) parametric time series models.

For semi-nonparametric time series models, we could compute various "pre-asymptotic" Wald statistics using various existing LRV estimators for regular functionals of (misspecified) parametric time series models, such as the kernel LRV estimators considered by Newey and West (1987), Andrews (1991), Jansson (2004), Kiefer and Vogelsang (2005), Sun (2011b) and others. Nevertheless, to be consistent with our focus on the method of sieves and to derive a simple and accurate asymptotic approximation, we compute a "pre-asymptotic" Wald statistic using an orthonormal series LRV (OS-LRV) estimator. The OS-LRV estimator has already been used in constructing autocorrelation robust inference on regular functionals of parametric time series models; see, e.g., Phillips (2005), Müller (2007), Sun (2011a), and the references therein. We extend these results to robust inference on both regular and irregular functionals of semi-nonparametric time series models.[1]

For both regular and irregular functionals, we show that the "pre-asymptotic" t statistic and a scaled

[1]We thank Peter Phillips for suggesting that we consider autocorrelation robust inference for semi-nonparametric time series models.

Wald statistic converge to the standard t distribution and F distribution respectively when the series number of terms in the OS-LRV estimator is held fixed; and that the t distribution and F distribution approach the standard normal and chi-square distributions respectively when the series number of terms in the OS-LRV estimator goes to infinity. Our "pre-asymptotic" t and F approximations achieve triple robustness in the following sense: they are asymptotically valid regardless of (1) whether the functional is regular or not; (2) whether there is temporal dependence or not; and (3) whether the series number of terms in the OS-LRV estimator is held fixed or not.

To facilitate the practical use of our inference procedure, we show that, in finite samples and for linear sieve M estimators, our "pre-asymptotic" sieve test statistics (i.e. t statistic and Wald statistic) for semi-nonparametric time series models are numerically equivalent to the corresponding test statistics one would obtain if the models are treated as if they were parametric.[2] These results are of much use to applied researchers, and demonstrate the advantage of the sieve method for inference on semi-nonparametric time series models.

To investigate the finite sample performance of our proposed "pre-asymptotic" robust inference procedures on semi-nonparametric time series models, we conduct a detailed simulation study using a partially linear regression model. For both regular and irregular functionals, we find that our test using the "pre-asymptotic" scaled Wald statistic and F critical values has more accurate size than the "pre-asymptotic" Wald test using chi-square critical values. For irregular functionals, we find that they both perform better than the Wald test using a consistent estimate of the asymptotic variance ignoring autocorrelation. These are especially true when the time series (with moderate sample size) has strong temporal dependence and the number of joint hypotheses being tested is large. Based on our simulation studies, we recommend the use of the "pre-asymptotic" scaled Wald statistic using an OS-LRV estimator and F approximation in empirical applications.

The rest of the paper is organized as follows. Section 2 presents the plug-in sieve M estimator of functionals of interest and gives two illustrative examples. Section 3 establishes the asymptotic normality of the plug-in sieve M estimators of possibly irregular functionals. Section 4 shows the surprising result that the asymptotic variances of plug-in sieve M estimators of irregular functionals for weakly dependent time series data are the same as if they were for i.i.d. data. Section 5 presents the "pre-asymptotic" OS-LRV estimator and F approximation. Section 6 proves the numerical equivalence result. Section 7 reports the simulation evidence, and the last section briefly concludes. Appendix A contains all the proofs, and Appendix B discusses the properties of the hidden delta functions associated with sieve M estimation of evaluation functionals.

Notation. In this paper, we denote $f_A(a)$ ($F_A(a)$) as the marginal probability density (cdf) of

[2] Here we slightly abuse terminology and define a parametric model to be a model with a fixed finite number of unknown parameters of interest, although the model may contain infinite dimensional nuisance parameters that are not needed to be estimated, such as Hansen (1982)'s GMM models.

a random variable A evaluated at a and $f_{AB}(a,b)$ ($F_{AB}(a,b)$) the joint density (cdf) of the random variables A and B. We use $\equiv$ to introduce definitions. For any vector-valued A, we let A' denote its transpose and $||A||_E \equiv \sqrt{A'A}$, although sometimes we also use $|A| = \sqrt{A'A}$ without too much confusion. Denote $L^p(\Omega, d\mu)$, $1 \leq p < \infty$, as a space of measurable functions with $||g||_{L^p(\Omega,d\mu)} \equiv \{\int_\Omega |g(t)|^p d\mu(t)\}^{1/p} < \infty$, where Ω is the support of the sigma-finite positive measure $d\mu$ (sometimes $L^p(\Omega)$ and $||g||_{L^p(\Omega)}$ are used when $d\mu$ is the Lebesgue measure). For any (possibly random) positive sequences $\{a_T\}_{T=1}^\infty$ and $\{b_T\}_{T=1}^\infty$, $a_T = O_p(b_T)$ means that $\lim_{c\to\infty}\limsup_T \Pr(a_T/b_T > c) = 0$; $a_T = o_p(b_T)$ means that for all $\varepsilon > 0$, $\lim_{T\to\infty}\Pr(a_T/b_T > \varepsilon) = 0$; and $a_T \asymp b_T$ means that there exist two constants $0 < c_1 \leq c_2 < \infty$ such that $c_1 a_T \leq b_T \leq c_2 a_T$. We use $\mathcal{A}_T \equiv \mathcal{A}_{k_T}$, $\mathcal{H}_T \equiv \mathcal{H}_{k_T}$ and $\mathcal{V}_T \equiv \mathcal{V}_{k_T}$ to denote various sieve spaces. To simplify the presentation, we assume that $\dim(\mathcal{V}_T) = \dim(\mathcal{A}_T) \asymp \dim(\mathcal{H}_T) \asymp k_T$, all of which grow to infinity with the sample size T.

2 Sieve M Estimation and Examples

2.1 Basic Setting

We assume that the data $\{Z_t = (Y_t', X_t')'\}_{t=1}^T$ is from a strictly stationary and weakly dependent process defined on an underlying complete probability space. Let the support of Z_t be $\mathcal{Z} \subseteq \mathbb{R}^{d_z}, 1 \leq d_z < \infty$, and let $\mathcal{Y}$ and $\mathcal{X}$ be the supports of Y and X respectively. Let $(\mathcal{A}, d)$ denote an infinite dimensional metric space. Let $\ell : \mathcal{Z} \times \mathcal{A} \to \mathbb{R}$ be a measurable function and $E[\ell(Z,\alpha)]$ be a population criterion. For simplicity we assume that there is a unique $\alpha_0 \in (\mathcal{A}, d)$ such that $E[\ell(Z,\alpha_0)] > E[\ell(Z,\alpha)]$ for all $\alpha \in (\mathcal{A}, d)$ with $d(\alpha,\alpha_0) > 0$. Different models in economics correspond to different choices of the criterion functions $E[\ell(Z,\alpha)]$ and the parameter spaces $(\mathcal{A}, d)$. A model does not need to be correctly specified and α_0 could be a pseudo-true parameter. Let $f : (\mathcal{A}, d) \to \mathbb{R}$ be a known measurable mapping. In this paper we are interested in estimation of and inference on $f(\alpha_0)$ via the method of sieves.

Let $\mathcal{A}_T$ be a sieve space for the whole parameter space $(\mathcal{A}, d)$. Then there is an element $\Pi_T\alpha_0 \in \mathcal{A}_T$ such that $d(\Pi_T\alpha_0, \alpha_0) \to 0$ as $\dim(\mathcal{A}_T) \to \infty$ (with T). An *approximate sieve M estimator* $\widehat{\alpha}_T \in \mathcal{A}_T$ of α_0 solves

$$\frac{1}{T}\sum_{t=1}^T \ell(Z_t, \widehat{\alpha}_T) \geq \sup_{\alpha\in\mathcal{A}_T} \frac{1}{T}\sum_{t=1}^T \ell(Z_t, \alpha) - O_p(\varepsilon_T^2), \tag{2.1}$$

where the term $O_p(\varepsilon_T^2) = o_p(T^{-1})$ denotes the maximization error when $\widehat{\alpha}_T$ fails to be the exact maximizer over the sieve space. We call $f(\widehat{\alpha}_T)$ the *plug-in sieve M estimator* of $f(\alpha_0)$.

Under very mild conditions (see, e.g., Chen (2007, Theorem 3.1) and White and Wooldridge (1991)), the sieve M estimator $\widehat{\alpha}_T$ is consistent for α_0: $d(\widehat{\alpha}_T, \alpha_0) = O_p\{\max[d(\widehat{\alpha}_T, \Pi_T\alpha_0), d(\Pi_T\alpha_0, \alpha_0)]\} = o_p(1)$.

2.2 Examples

The method of sieve M estimation includes many special cases. Different choices of criterion functions $\ell(Z_t,\alpha)$ and different choices of sieves $\mathcal{A}_T$ lead to different examples of sieve M estimation. As an illustration, we provide two examples below. See, e.g., Chen (2007, 2011) for additional applications.

Example 2.1 ***(Partially linear ARX regression)*** *Suppose that the time series data $\{Y_t\}_{t=1}^T$ is generated by*

$$Y_t = X_{1,t}'\theta_0 + h_0\left(X_{2,t}\right) + u_t,\ \ E\left[u_t|X_{1,t},X_{2,t}\right]=0, \tag{2.2}$$

where $X_{1,t}$ and $X_{2,t}$ are d_1 and d_2 dimensional random vectors respectively, and $X_{1,t}$ could include finitely many lagged Y_t's. Let $\theta_0 \in \Theta \subset \mathbb{R}^{d_1}$ and $h_0 \in \mathcal{H}$ a function space. Let $\alpha_0 = (\theta_0', h_0)' \in \mathcal{A} = \Theta\times\mathcal{H}$. Examples of functionals of interest could be $f(\alpha_0)=\lambda'\theta_0$ or $h_0(\overline{x}_2)$ where $\lambda \in \mathbb{R}^{d_1}$ and $\overline{x}_2$ is some point in the support of $X_{2,t}$.

Let $\mathcal{X}_j$ be the support of X_j for $j=1,2$. For simplicity we assume that $\mathcal{X}_2$ is a convex and bounded subset of $\mathbb{R}^{d_2}$. For the sake of concreteness we let $\mathcal{H}=\Lambda^s(\mathcal{X}_2)$ (a Hölder space):

$$\Lambda^s(\mathcal{X}_2) = \left\{ h \in C^{[s]}\left(\mathcal{X}_2\right) : \sup_{j\le[s]}\sup_{x\in\mathcal{X}_2}\left|h^{(j)}(x)\right| < \infty,\ \sup_{x,x'\in\mathcal{X}_2}\frac{\left|h^{([s])}(x)-h^{([s])}\left(x'\right)\right|}{\left|x-x'\right|^{s-[s]}} < \infty \right\},$$

where $[s]$ is the largest integer that is strictly smaller than s. The Hölder space $\Lambda^s(\mathcal{X}_2)$ (with $s>0.5d_2$) is a smooth function space that is widely assumed in the semi-nonparametric literature. We can then approximate $\mathcal{H}=\Lambda^s(\mathcal{X}_2)$ by various linear sieve spaces:

$$\mathcal{H}_T = \left\{ h(\cdot) : h(\cdot) = \sum_{j=1}^{k_T}\beta_j p_j(\cdot) = \beta' P_{k_T}(\cdot),\ \beta\in\mathbb{R}^{k_T} \right\}, \tag{2.3}$$

where the known sieve basis $P_{k_T}(\cdot)$ could be tensor-products of splines, wavelets, Fourier series and others; see, e.g., Newey (1997) and Chen (2007).

Let $\ell(Z_t,\alpha) = -\left[Y_t - X_{1,t}'\theta - h\left(X_{2,t}\right)\right]^2/4$ with $Z_t=(Y_t, X_{1,t}', X_{2,t}')'$ and $\alpha=(\theta',h)'\in\mathcal{A}=\Theta\times\mathcal{H}$. Let $\mathcal{A}_T=\Theta\times\mathcal{H}_T$ be a sieve for $\mathcal{A}$. We can estimate $\alpha_0=(\theta_0',h_0)'\in\mathcal{A}$ by the sieve least squares (LS) (a special case of sieve M estimation):

$$\widehat{\alpha}_T \equiv (\widehat{\theta}_T', \widehat{h}_T)' = \arg\max_{(\theta,h)\in\Theta\times\mathcal{H}_T}\frac{1}{T}\sum_{t=1}^T \ell(Z_t,\theta,h). \tag{2.4}$$

A functional of interest $f(\alpha_0)$ (such as $\lambda'\theta_0$ or $h_0(\overline{x}_2)$) is then estimated by the plug-in sieve LS estimator $f(\widehat{\alpha}_T)$ (such as $\lambda'\widehat{\theta}_T$ or $\widehat{h}_T(\overline{x}_2)$).

This example is very similar to example 2 in Chen and Shen (1998) and example 4.2.3 in Chen (2007). One can slightly modify their proofs to get the convergence rate of $\widehat{\alpha}_T$ and the $\sqrt{T}$-asymptotic normality of $\lambda'\widehat{\theta}_T$. But neither paper provides a variance estimator for $\lambda'\widehat{\theta}_T$. The results in our paper immediately lead to the asymptotic normality of $f(\widehat{\alpha}_T)$ for possibly irregular functionals $f(\alpha_0)$ and provide simple, accurate inference on $f(\alpha_0)$.

Example 2.2 ***(Possibly misspecified copula-based time series model)*** *Suppose that $\{Y_t\}_{t=1}^T$ is a sample of strictly stationary first order Markov process generated from $(F_Y, C_0(\cdot,\cdot))$, where F_Y is the true unknown continuous marginal distribution, and $C_0(\cdot,\cdot)$ is the true unknown copula for (Y_{t-1}, Y_t) that captures all the temporal and tail dependence of $\{Y_t\}$. The τ-th conditional quantile of Y_t given $Y^{t-1} = (Y_{t-1}, ..., Y_1)$ is:*

$$Q_\tau^Y(y) = F_Y^{-1}\left(C_{2|1}^{-1}\left[\tau|F_Y(y)\right]\right),$$

where $C_{2|1}[\cdot|u] \equiv \frac{\partial}{\partial u}C_0(u,\cdot)$ is the conditional distribution of $U_t \equiv F_Y(Y_t)$ given $U_{t-1} = u$, and $C_{2|1}^{-1}[\tau|u]$ is its τ-th conditional quantile. The conditional density function of Y_t given Y^{t-1} is

$$p^0(\cdot|Y^{t-1}) = f_Y(\cdot)c_0\left(F_Y(Y_{t-1}), F_Y(\cdot)\right),$$

where $f_Y(\cdot)$ and $c_0(\cdot,\cdot)$ are the density functions of $F_Y(\cdot)$ and $C_0(\cdot,\cdot)$ respectively. A researcher specifies a parametric form $\{c(\cdot,\cdot;\theta) : \theta \in \Theta\}$ for the copula density function, but it could be misspecified in the sense $c_0(\cdot,\cdot) \notin \{c(\cdot,\cdot;\theta) : \theta \in \Theta\}$. Let θ_0 be the pseudo true copula dependence parameter:

$$\theta_0 = \arg\max_{\theta\in\Theta} \int_0^1 \int_0^1 c(u,v;\theta)c_0(u,v)dudv.$$

Let $(\theta_0', f_Y)'$ be the parameters of interest. Examples of functionals of interest could be $\lambda'\theta_0$, $f_Y(\overline{y})$, $F_Y(\overline{y})$ or $Q_{0.01}^Y(\overline{y}) = F_Y^{-1}\left(C_{2|1}^{-1}\left[\tau|F_Y(y);\theta_0\right]\right)$ for any $\lambda \in \mathbb{R}^{d_\theta}$ and some $\overline{y} \in supp(Y_t)$.

We could estimate $(\theta_0', f_Y)'$ by the method of sieve quasi ML using different parameterizations and different sieves for f_Y. For example, let $h_0 = \sqrt{f_Y}$ and $\alpha_0 = (\theta_0', h_0)'$ be the (pseudo) true unknown parameters. Then $f_Y(\cdot) = h_0^2(\cdot) / \int_{-\infty}^{\infty} h_0^2(y)\,dy$, and $h_0 \in L^2(\mathbb{R})$. For the identification of h_0, we can assume that $h_0 \in \mathcal{H}$:

$$\mathcal{H} = \left\{ h(\cdot) = p_0(\cdot) + \sum_{j=1}^{\infty} \beta_j p_j(\cdot) : \sum_{j=1}^{\infty} \beta_j^2 < \infty \right\}, \tag{2.5}$$

where $\{p_j\}_{j=0}^{\infty}$ is a complete orthonormal basis functions in $L^2(\mathbb{R})$, such as Hermite polynomials, wavelets and other orthonormal basis functions. Here we normalize the coefficient of the first basis function $p_0(\cdot)$ to be 1 in order to achieve the identification of $h_0(\cdot)$. Other normalization could also be used. It is now obvious that $h_0 \in \mathcal{H}$ could be approximated by functions in the following sieve space:

$$\mathcal{H}_T = \left\{ h(\cdot) = p_0(\cdot) + \sum_{j=1}^{k_T} \beta_j p_j(\cdot) = p_0(\cdot) + \beta' P_{k_T}(\cdot) : \beta \in \mathbb{R}^{k_T} \right\}. \tag{2.6}$$

Let $Z_t' = (Y_{t-1}, Y_t)$, $\alpha = (\theta', h)' \in \mathcal{A} = \Theta \times \mathcal{H}$ and

$$\ell(Z_t, \alpha) = \log \left\{ \frac{h^2(Y_t)}{\int_{-\infty}^{\infty} h^2(y)\,dy} \right\} + \log \left\{ c\left(\int_{-\infty}^{Y_{t-1}} \frac{h^2(y)}{\int_{-\infty}^{\infty} h^2(x)\,dx} dy, \int_{-\infty}^{Y_t} \frac{h^2(y)}{\int_{-\infty}^{\infty} h^2(x)\,dx} dy; \theta \right) \right\}. \quad (2.7)$$

Then $\alpha_0 = (\theta_0', h_0)' \in \mathcal{A} = \Theta \times \mathcal{H}$ could be estimated by the sieve quasi MLE $\widehat{\alpha}_T = (\widehat{\theta}_T', \widehat{h}_T)' \in \mathcal{A}_T = \Theta \times \mathcal{H}_T$ that solves:

$$\sup_{\alpha \in \Theta \times \mathcal{H}_T} \frac{1}{T} \left\{ \sum_{t=2}^{T} \ell(Z_t, \alpha) + \log \left\{ \frac{h^2(Y_1)}{\int_{-\infty}^{\infty} h^2(y)\,dy} \right\} \right\} - O_p(\varepsilon_T^2). \quad (2.8)$$

A functional of interest $f(\alpha_0)$ (such as $\lambda'\theta_0$, $f_Y(\overline{y}) = h_0^2(\overline{y}) / \int_{-\infty}^{\infty} h_0^2(y)\,dy$, $F_Y(\overline{y})$ or $Q_{0.01}^Y(\overline{y})$) is then estimated by the plug-in sieve quasi MLE $f(\widehat{\alpha}_T)$ (such as $\lambda'\widehat{\theta}$, $\widehat{f}_Y(\overline{y}) = \widehat{h}_T^2(\overline{y}) / \int_{-\infty}^{\infty} \widehat{h}_T^2(y)\,dy$, $\widehat{F}_Y(\overline{y}) = \int_{-\infty}^{\overline{y}} \widehat{f}_Y(y)dy$ or $\widehat{Q}_{0.01}^Y(\overline{y}) = \widehat{F}_Y^{-1}(C_{2|1}^{-1}[\tau|\widehat{F}_Y(y); \widehat{\theta}])$).

Under correct specification, Chen, Wu and Yi (2009) establish the rate of convergence of the sieve MLE $\widehat{\alpha}_T$ and provide a sieve likelihood-ratio inference for regular functionals including $f(\alpha_0) = \lambda'\theta_0$ or $F_Y(\overline{y})$ or $Q_{0.01}^Y(\overline{y})$. Under misspecified copulas, by applying Chen and Shen (1998), we can still derive the convergence rate of the sieve quasi MLE $\widehat{\alpha}_T$ and the $\sqrt{T}$ asymptotic normality of $f(\widehat{\alpha}_T)$ for regular functionals. However, the sieve likelihood ratio inference given in Chen, Wu and Yi (2009) is no longer valid under misspecification. The results in this paper immediately lead to the asymptotic normality of $f(\widehat{\alpha}_T)$ (such as $\widehat{f}_Y(\overline{y}) = \widehat{h}_T^2(\overline{y}) / \int_{-\infty}^{\infty} \widehat{h}_T^2(y)\,dy$) for any possibly irregular functional $f(\alpha_0)$ (such as $f_Y(\overline{y})$) as well as valid inferences under potential misspecification.

3 Asymptotic Normality of Sieve M-Estimators

In this section, we establish the asymptotic normality of plug-in sieve M estimators of possibly irregular functionals of semi-nonparametric time series models. We also give a closed-form expression for the sieve Riesz representor that appears in our asymptotic normality result.

3.1 Local Geometry

Given the existing consistency result ($d(\widehat{\alpha}_T, \alpha_0) = o_p(1)$), we can restrict our attention to a shrinking d-neighborhood of α_0. We equip $\mathcal{A}$ with an inner product induced norm $\|\alpha - \alpha_0\|$ that is weaker than $d(\alpha, \alpha_0)$ (i.e., $\|\alpha - \alpha_0\| \leq cd(\alpha, \alpha_0)$ for a constant c), and is locally equivalent to $\sqrt{E[\ell(Z_t, \alpha_0) - \ell(Z_t, \alpha)]}$ in a shrinking d-neighborhood of α_0. For strictly stationary weakly dependent data, Chen and Shen (1998) establish the convergence rate $\|\widehat{\alpha}_T - \alpha_0\| = O_p(\xi_T) = o\left(T^{-1/4}\right)$. The convergence rate result implies that $\widehat{\alpha}_T \in \mathcal{B}_T \subset \mathcal{B}_0$ with probability approaching one, where

$$\mathcal{B}_0 \equiv \{\alpha \in \mathcal{A} : \|\alpha - \alpha_0\| \leq C\xi_T \log(\log(T))\}; \; \mathcal{B}_T \equiv \mathcal{B}_0 \cap \mathcal{A}_T. \quad (3.1)$$

Hence, we can now regard $\mathcal{B}_0$ as the effective parameter space and $\mathcal{B}_T$ as its sieve space.

Define

$$\alpha_{0,T} \equiv \arg\min_{\alpha \in \mathcal{B}_T} ||\alpha - \alpha_0||. \tag{3.2}$$

Let $\mathcal{V}_T \equiv clsp(\mathcal{B}_T) - \{\alpha_{0,T}\}$, where $clsp(\mathcal{B}_T)$ denotes the closed linear span of $\mathcal{B}_T$ under $\|\cdot\|$. Then $\mathcal{V}_T$ is a finite dimensional Hilbert space under $\|\cdot\|$. Similarly the space $\mathcal{V} \equiv clsp(\mathcal{B}_0) - \{\alpha_0\}$ is a Hilbert space under $\|\cdot\|$. Moreover, $\mathcal{V}_T$ is dense in $\mathcal{V}$ under $\|\cdot\|$. To simplify the presentation, we assume that $\dim(\mathcal{V}_T) = \dim(\mathcal{A}_T) \asymp k_T$, all of which grow to infinity with T. By definition we have $\langle \alpha_{0,T} - \alpha_0, v_T \rangle = 0$ for all $v_T \in \mathcal{V}_T$.

As demonstrated in Chen and Shen (1998) and Chen (2007), there is lots of freedom to choose such a norm $\|\alpha - \alpha_0\|$ that is weaker than $d(\alpha, \alpha_0)$ and is locally equivalent to $\sqrt{E[\ell(Z, \alpha_0) - \ell(Z, \alpha)]}$. In some parts of this paper, for the sake of concreteness, we present results for a specific choice of the norm $\|\cdot\|$. We suppose that for all α in a shrinking d-neighborhood of α_0, $\ell(Z, \alpha) - \ell(Z, \alpha_0)$ can be approximated by $\Delta(Z, \alpha_0)[\alpha - \alpha_0]$ such that $\Delta(Z, \alpha_0)[\alpha - \alpha_0]$ is linear in $\alpha - \alpha_0$. Denote the remainder of the approximation as:

$$r(Z, \alpha_0)[\alpha - \alpha_0, \alpha - \alpha_0] \equiv 2\left\{\ell(Z, \alpha) - \ell(Z, \alpha_0) - \Delta(Z, \alpha_0)[\alpha - \alpha_0]\right\}. \tag{3.3}$$

When $\lim_{\tau \to 0}[(\ell(Z, \alpha_0 + \tau[\alpha - \alpha_0]) - \ell(Z, \alpha_0))/\tau]$ is well defined, we could let $\Delta(Z, \alpha_0)[\alpha - \alpha_0] = \lim_{\tau \to 0}[(\ell(Z, \alpha_0 + \tau[\alpha - \alpha_0]) - \ell(Z, \alpha_0))/\tau]$, which is called the directional derivative of $\ell(Z, \alpha)$ at α_0 in the direction $[\alpha - \alpha_0]$. Define

$$\|\alpha - \alpha_0\| = \sqrt{E\left(-r(Z, \alpha_0)[\alpha - \alpha_0, \alpha - \alpha_0]\right)} \tag{3.4}$$

with the corresponding inner product $\langle \cdot, \cdot \rangle$

$$\langle \alpha_1 - \alpha_0, \alpha_2 - \alpha_0 \rangle = E\left\{-r(Z, \alpha_0)[\alpha_1 - \alpha_0, \alpha_2 - \alpha_0]\right\} \tag{3.5}$$

for any α_1, α_2 in the shrinking d-neighborhood of α_0. In general this norm defined in (3.4) is weaker than $d(\cdot, \cdot)$. Since α_0 is the unique maximizer of $E[\ell(Z, \alpha)]$ on $\mathcal{A}$, under mild conditions $\|\alpha - \alpha_0\|$ defined in (3.4) is locally equivalent to $\sqrt{E[\ell(Z, \alpha_0) - \ell(Z, \alpha)]}$.

For any $v \in \mathcal{V}$, we define $\frac{\partial f(\alpha_0)}{\partial \alpha}[v]$ to be the pathwise (directional) derivative of the functional $f(\cdot)$ at α_0 and in the direction of $v = \alpha - \alpha_0 \in \mathcal{V}$:

$$\frac{\partial f(\alpha_0)}{\partial \alpha}[v] = \left.\frac{\partial f(\alpha_0 + \tau v)}{\partial \tau}\right|_{\tau = 0} \text{ for any } v \in \mathcal{V}. \tag{3.6}$$

For any $v_T = \alpha_T - \alpha_{0,T} \in \mathcal{V}_T$, we let

$$\frac{\partial f(\alpha_0)}{\partial \alpha}[v_T] = \frac{\partial f(\alpha_0)}{\partial \alpha}[\alpha_T - \alpha_0] - \frac{\partial f(\alpha_0)}{\partial \alpha}[\alpha_{0,T} - \alpha_0]. \tag{3.7}$$

So $\frac{\partial f(\alpha_0)}{\partial \alpha}[\cdot]$ is also a linear functional on $\mathcal{V}_T$.

Note that $\mathcal{V}_T$ is a finite dimensional Hilbert space. As any linear functional on a finite dimensional Hilbert space is bounded, we can invoke the Riesz representation theorem to deduce that there is a $v_T^* \in \mathcal{V}_T$ such that

$$\frac{\partial f(\alpha_0)}{\partial \alpha}[v] = \langle v_T^*, v \rangle \quad \text{for all } v \in \mathcal{V}_T \tag{3.8}$$

and that

$$\frac{\partial f(\alpha_0)}{\partial \alpha}[v_T^*] = \|v_T^*\|^2 = \sup_{v \in \mathcal{V}_T, v \neq 0} \frac{|\frac{\partial f(\alpha_0)}{\partial \alpha}[v]|^2}{\|v\|^2} \tag{3.9}$$

We call v_T^* the *sieve Riesz representor* of the functional $\frac{\partial f(\alpha_0)}{\partial \alpha}[\cdot]$ on $\mathcal{V}_T$.

We emphasize that the sieve Riesz representation (3.8)–(3.9) of the linear functional $\frac{\partial f(\alpha_0)}{\partial \alpha}[\cdot]$ on $\mathcal{V}_T$ always exists regardless of whether $\frac{\partial f(\alpha_0)}{\partial \alpha}[\cdot]$ is bounded on the infinite dimensional space $\mathcal{V}$ or not.

- If $\frac{\partial f(\alpha_0)}{\partial \alpha}[\cdot]$ is bounded on the infinite dimensional Hilbert space $\mathcal{V}$, i.e.

$$\|v^*\| \equiv \sup_{v \in \mathcal{V}, v \neq 0} \left\{ |\frac{\partial f(\alpha_0)}{\partial \alpha}[v]| / \|v\| \right\} < \infty, \tag{3.10}$$

  then $\|v_T^*\| = O(1)$ (in fact $\|v_T^*\| \nearrow \|v^*\| < \infty$ and $\|v^* - v_T^*\| \to 0$ as $T \to \infty$); we say that $f(\cdot)$ is *regular* (at $\alpha = \alpha_0$). In this case, we have $\frac{\partial f(\alpha_0)}{\partial \alpha}[v] = \langle v^*, v \rangle$ for all $v \in \mathcal{V}$, and v^* is the Riesz representor of the functional $\frac{\partial f(\alpha_0)}{\partial \alpha}[\cdot]$ on $\mathcal{V}$.

- If $\frac{\partial f(\alpha_0)}{\partial \alpha}[\cdot]$ is unbounded on the infinite dimensional Hilbert space $\mathcal{V}$, i.e.

$$\sup_{v \in \mathcal{V}, v \neq 0} \left\{ |\frac{\partial f(\alpha_0)}{\partial \alpha}[v]| / \|v\| \right\} = \infty, \tag{3.11}$$

  then $\|v_T^*\| \nearrow \infty$ as $T \to \infty$; and we say that $f(\cdot)$ is *irregular* (at $\alpha = \alpha_0$).

As it will become clear later, the convergence rate of $f(\widehat{\alpha}_T) - f(\alpha_0)$ depends on the order of $\|v_T^*\|$.

3.2 Asymptotic Normality

To establish the asymptotic normality of $f(\widehat{\alpha}_T)$ for possibly irregular nonlinear functionals, we assume:

Assumption 3.1 (local behavior of functional)

(i) $\sup_{\alpha \in \mathcal{B}_T} \left| f(\alpha) - f(\alpha_0) - \frac{\partial f(\alpha_0)}{\partial \alpha}[\alpha - \alpha_0] \right| = o\left(T^{-\frac{1}{2}} \|v_T^*\|\right)$;

(ii) $\left| \frac{\partial f(\alpha_0)}{\partial \alpha}[\alpha_{0,T} - \alpha_0] \right| = o\left(T^{-\frac{1}{2}} \|v_T^*\|\right)$.

Assumption 3.1.(i) controls the linear approximation error of possibly nonlinear functional $f(\cdot)$. It is automatically satisfied when $f(\cdot)$ is a linear functional, but it may rule out some highly nonlinear functionals. Assumption 3.1.(ii) controls the bias part due to the finite dimensional sieve approximation of $\alpha_{0,T}$ to α_0. It is a condition imposed on the growth rate of the sieve dimension $\dim(\mathcal{A}_T)$, and requires that the sieve approximation error rate is of smaller order than $T^{-\frac{1}{2}}\left\|v_T^*\right\|$. When $f(\cdot)$ is a regular functional, we have $\|v_T^*\| \nearrow \|v^*\| < \infty$, and since $\langle \alpha_{0,T} - \alpha_0, v_T^* \rangle = 0$ (by definition of $\alpha_{0,T}$), we have:

$$\left|\frac{\partial f(\alpha_0)}{\partial \alpha}[\alpha_{0,T} - \alpha_0]\right| = |\langle v^*, \alpha_{0,T} - \alpha_0 \rangle| = |\langle v^* - v_T^*, \alpha_{0,T} - \alpha_0 \rangle| \leq \|v^* - v_T^*\| \times \|\alpha_{0,T} - \alpha_0\|,$$

thus Assumption 3.1.(ii) is satisfied if

$$\|v^* - v_T^*\| \times \|\alpha_{0,T} - \alpha_0\| = o(T^{-1/2}) \quad \text{when } f(\cdot) \text{ is regular}, \tag{3.12}$$

which is similar to condition 4.1(ii)(iii) imposed in Chen (2007, p. 5612) for regular functionals.

Next, we make an assumption on the relationship between $\|v_T^*\|$ and the asymptotic standard deviation of $f(\widehat{\alpha}_T) - f(\alpha_{0,T})$. It will be shown that the asymptotic standard deviation is the limit of the "standard deviation" (sd) norm $\|v_T^*\|_{sd}$ of v_T^*, defined as

$$\|v_T^*\|_{sd}^2 \equiv Var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \Delta(Z_t, \alpha_0)[v_T^*]\right). \tag{3.13}$$

Note that $\|v_T^*\|_{sd}^2$ is the finite dimensional sieve version of the long run variance of the score process $\Delta(Z_t, \alpha_0)[v_T^*]$. Since $v_T^* \in \mathcal{V}_T$, the sd norm $\|v_T^*\|_{sd}$ depends on the sieve dimension $\dim(\mathcal{A}_T)$ that grows with the sample size T.

Assumption 3.2 (sieve variance) $\|v_T^*\| / \|v_T^*\|_{sd} = O(1)$.

By definition of $\|v_T^*\|$ given in (3.9), $0 < \|v_T^*\|$ is non-decreasing in $\dim(\mathcal{V}_T)$, and hence is non-decreasing in T. Assumption 3.2 then implies that $\liminf_{T\to\infty} \|v_T^*\|_{sd} > 0$. Define

$$u_T^* \equiv \frac{v_T^*}{\left\|v_T^*\right\|_{sd}} \tag{3.14}$$

to be the normalized version of v_T^*. Then Assumption 3.2 implies that $\|u_T^*\| = O(1)$.

Let $\mu_T\{g(Z)\} \equiv T^{-1}\sum_{t=1}^{T}[g(Z_t) - Eg(Z_t)]$ denote the centered empirical process indexed by the function g. Let $\varepsilon_T = o(T^{-1/2})$. For notational economy, we use the same ε_T as that in (2.1).

Assumption 3.3 (local behavior of criterion) *(i)* $\mu_T\{\Delta(Z, \alpha_0)[v]\}$ *is linear in* $v \in \mathcal{V}$*;*

$$\text{(ii)} \quad \sup_{\alpha \in \mathcal{B}_T} \mu_T\{\ell(Z, \alpha \pm \varepsilon_T u_T^*) - \ell(Z, \alpha) - \Delta(Z, \alpha_0)[\pm\varepsilon_T u_T^*]\} = O_p(\varepsilon_T^2);$$

$$\text{(iii)} \quad \sup_{\alpha \in \mathcal{B}_T} \left| E[\ell(Z_t, \alpha) - \ell(Z_t, \alpha \pm \varepsilon_T u_T^*)] - \frac{||\alpha \pm \varepsilon_T u_T^* - \alpha_0||^2 - ||\alpha - \alpha_0||^2}{2} \right| = O(\varepsilon_T^2).$$

Assumptions 3.3.(ii) and (iii) are essentially the same as conditions 4.2 and 4.3 of Chen (2007, p. 5612) respectively. In particular, the stochastic equicontinuity assumption 3.3.(ii) can be easily verified by applying Lemma 4.2 of Chen (2007).

Assumption 3.4 (CLT) $\sqrt{T}\mu_T\left\{\Delta(Z,\alpha_0)\left[u_T^*\right]\right\} \to_d N(0,1)$, *where* $N(0,1)$ *is a standard normal distribution.*

Assumption 3.4 is a very mild one, which effectively combines conditions 4.4 and 4.5 of Chen (2007, p. 5612). This can be easily verified by applying any existing triangular array CLT for weakly dependent data (see, e.g., White (2004) for references).

We are now ready to state the asymptotic normality theorem for the plug-in sieve M estimator.

Theorem 3.1 *Let Assumptions 3.1.(i), 3.2 and 3.3 hold. Then*

$$\sqrt{T}\frac{f(\widehat{\alpha}_T)-f(\alpha_{0,T})}{\left\|v_T^*\right\|_{sd}} = \sqrt{T}\mu_T\left\{\Delta(Z,\alpha_0)\left[u_T^*\right]\right\} + o_p\left(1\right); \tag{3.15}$$

If further Assumptions 3.1.(ii) and 3.4 hold, then

$$\sqrt{T}\frac{f(\widehat{\alpha}_T)-f(\alpha_0)}{\left\|v_T^*\right\|_{sd}} = \sqrt{T}\mu_T\left\{\Delta(Z,\alpha_0)\left[u_T^*\right]\right\} + o_p\left(1\right) \to_d N(0,1). \tag{3.16}$$

In light of Theorem 3.1, we call $\|v_T^*\|_{sd}^2$ defined in (3.13) the "*pre-asymptotic*" sieve variance of the estimator $f(\widehat{\alpha}_T)$. When the functional $f(\alpha_0)$ is regular (i.e., $\|v_T^*\| = O(1)$), we have $\|v_T^*\|_{sd} \asymp \|v_T^*\| = O(1)$ typically; so $f(\widehat{\alpha}_T)$ converges to $f(\alpha_0)$ at the parametric rate of $1/\sqrt{T}$. When the functional $f(\alpha_0)$ is irregular (i.e., $\|v_T^*\| \to \infty$), we have $\|v_T^*\|_{sd} \to \infty$ (under Assumption 3.2); so the convergence rate of $f(\widehat{\alpha}_T)$ becomes slower than $1/\sqrt{T}$. Regardless of whether the "pre-asymptotic" sieve variance $\|v_T^*\|_{sd}^2$ stays bounded asymptotically (i.e., as $T \to \infty$) or not, it always captures whatever true temporal dependence exists in finite samples.

Note that $\|v_T^*\|_{sd}^2 = Var\left(\Delta(Z,\alpha_0)[v_T^*]\right)$ if either the score process $\{\Delta(Z_t,\alpha_0)[v_T^*]\}_{t\leq T}$ is a martingale difference array or if data $\{Z_t\}_{t=1}^T$ is iid. Therefore, Theorem 3.1 recovers the asymptotic normality result in Chen and Liao (2008) for sieve M estimators of possibly irregular functionals with iid data.

For regular functionals of semi-nonparametric time series models, Chen and Shen (1998) and Chen (2007, Theorem 4.3) establish that $\sqrt{T}\left(f(\widehat{\alpha}_T)-f(\alpha_0)\right) \to_d N(0,\sigma_{v^*}^2)$ with

$$\sigma_{v^*}^2 = \lim_{T\to\infty} Var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Delta(Z_t,\alpha_0)[v^*]\right) = \lim_{T\to\infty}\|v_T^*\|_{sd}^2 \in (0,\infty). \tag{3.17}$$

Our Theorem 3.1 is a natural extension of their results to allow for irregular functionals as well.

3.3 Sieve Riesz Representor

To apply the asymptotic normality Theorem 3.1 one needs to verify Assumptions 3.1–3.4. Once we compute the sieve Riesz representor $v_T^* \in \mathcal{V}_T$, Assumptions 3.1 and 3.2 can be easily checked, while Assumptions 3.3 and 3.4 are standard ones and can be verified in the same ways as those in Chen and Shen (1998) and Chen (2007) for regular functionals of semi-nonparametric models. Although it may be difficult to compute the Riesz representor $v^* \in \mathcal{V}$ in a closed form for a regular functional on the infinite dimensional space $\mathcal{V}$ (see e.g., Ai and Chen (2003) for discussions), we can always compute the sieve Riesz representor $v_T^* \in \mathcal{V}_T$ defined in (3.8) and (3.9) explicitly. Therefore, Theorem 3.1 is easily applicable to a large class of semi-nonparametric time series models, regardless of whether the functionals of interest are $\sqrt{T}$ estimable or not.

3.3.1 Sieve Riesz representors for general functionals

For the sake of concreteness, in this subsection we focus on a large class of semi-nonparametric models where the population criterion $E[\ell(Z_t, \theta, h(\cdot))]$ is maximized at $\alpha_0 = (\theta_0', h_0(\cdot))' \in \mathcal{A} = \Theta \times \mathcal{H}$, Θ is a compact subset in $\mathbb{R}^{d_\theta}$, $\mathcal{H}$ is a class of real valued continuous functions (of a subset of Z_t) belonging to a Hölder, Sobolev or Besov space, and $\mathcal{A}_T = \Theta \times \mathcal{H}_T$ is a finite dimensional sieve space. The general cases with multiple unknown functions require only more complicated notation.

Let $\|\cdot\|$ be the norm defined in (3.4) and $\mathcal{V}_T = \mathbb{R}^{d_\theta} \times \{v_h(\cdot) = P_{k_T}(\cdot)'\beta : \beta \in \mathbb{R}^{k_T}\}$ be dense in the infinite dimensional Hilbert space $(\mathcal{V}, \|\cdot\|)$. By definition, the sieve Riesz representor $v_T^* = (v_{\theta,T}^{*\prime}, v_{h,T}^*(\cdot))' = (v_{\theta,T}^{*\prime}, P_{k_T}(\cdot)'\beta_T^*)' \in \mathcal{V}_T$ of $\frac{\partial f(\alpha_0)}{\partial \alpha}[\cdot]$ solves the following optimization problem:

$$\frac{\partial f(\alpha_0)}{\partial \alpha}[v_T^*] = \|v_T^*\|^2 = \sup_{v=(v_\theta', v_h)' \in \mathcal{V}_T, v\neq 0} \frac{\left|\frac{\partial f(\alpha_0)}{\partial \theta'} v_\theta + \frac{\partial f(\alpha_0)}{\partial h}[v_h(\cdot)]\right|^2}{E\left(-r\left(Z_t, \theta_0, h_0(\cdot)\right)[v, v]\right)}$$

$$= \sup_{\gamma=(v_\theta', \beta')' \in \mathbb{R}^{d_\theta + k_T}, \gamma\neq 0} \frac{\gamma' F_{k_T} F_{k_T}' \gamma}{\gamma' R_{k_T} \gamma}, \tag{3.18}$$

where

$$F_{k_T} \equiv \left(\frac{\partial f(\alpha_0)}{\partial \theta'}, \frac{\partial f(\alpha_0)}{\partial h}[P_{k_T}(\cdot)']\right)' \tag{3.19}$$

is a $(d_\theta + k_T) \times 1$ vector,[3] and

$$\gamma' R_{k_T} \gamma \equiv E\left(-r\left(Z_t, \theta_0, h_0(\cdot)\right)[v, v]\right) \quad \text{for all } v = \left(v_\theta', P_{k_T}(\cdot)'\beta\right)' \in \mathcal{V}_T, \tag{3.20}$$

with

$$R_{k_T} = \begin{pmatrix} I_{11} & I_{T,12} \\ I_{T,21} & I_{T,22} \end{pmatrix} \quad \text{and} \quad R_{k_T}^{-1} := \begin{pmatrix} I_T^{11} & I_T^{12} \\ I_T^{21} & I_T^{22} \end{pmatrix} \tag{3.21}$$

[3] When $\frac{\partial f(\alpha_0)}{\partial h}[\cdot]$ applies to a vector (matrix), it stands for element-wise (column-wise) operations. We follow the same convention for other operators such as $\Delta(Z_t, \alpha_0)[\cdot]$ and $-r(Z_t, \alpha_0)[\cdot, \cdot]$ throughout the paper.

being $(d_\theta+k_T)\times(d_\theta+k_T)$ positive definite matrices. For example if the criterion function $\ell(z,\theta,h(\cdot))$ is twice continuously pathwise differentiable with respect to $(\theta,h(\cdot))$, then we have $I_{11}=E\left[-\frac{\partial^2\ell(Z_t,\theta_0,h_0(\cdot))}{\partial\theta\partial\theta'}\right]$, $I_{T,22}=E\left[-\frac{\partial^2\ell(Z_t,\theta_0,h_0(\cdot))}{\partial h\partial h}[P_{k_T}(\cdot),P_{k_T}(\cdot)']\right]$, $I_{T,12}=E\left[\frac{\partial^2\ell(Z_t,\theta_0,h_0(\cdot))}{\partial\theta\partial h}[P_{k_T}(\cdot)]\right]$ and $I_{T,21}\equiv I'_{T,12}$.

The sieve Riesz representation (3.8) becomes: for all $v=(v'_\theta,P_{k_T}(\cdot)'\beta)'\in\mathcal{V}_T$,

$$\frac{\partial f(\alpha_0)}{\partial\alpha}[v]=F'_{k_T}\gamma=\langle v^*_T,v\rangle=\gamma^{*\prime}_T R_{k_T}\gamma \quad \text{for all } \gamma=(v'_\theta,\beta')'\in\mathbb{R}^{d_\theta+k_T}. \tag{3.22}$$

It is obvious that the optimal solution of γ in (3.18) or in (3.22) has a closed-form expression:

$$\gamma^*_T=\left(v^{*\prime}_{\theta,T},\beta^{*\prime}_T\right)'=R^{-1}_{k_T}F_{k_T}. \tag{3.23}$$

The sieve Riesz representor is then given by

$$v^*_T=\left(v^{*\prime}_{\theta,T},v^*_{h,T}(\cdot)\right)'=\left(v^{*\prime}_{\theta,T},P_{k_T}(\cdot)'\beta^*_T\right)'\in\mathcal{V}_T.$$

Consequently,

$$\|v^*_T\|^2=\gamma^{*\prime}_T R_{k_T}\gamma^*_T=F'_{k_T}R^{-1}_{k_T}F_{k_T}, \tag{3.24}$$

which is finite for each sample size T but may grow with T.

Finally the score process can be expressed as

$$\Delta(Z_t,\alpha_0)[v^*_T]=\left(\Delta_\theta(Z_t,\theta_0,h_0(\cdot))',\Delta_h(Z_t,\theta_0,h_0(\cdot))[P_{k_T}(\cdot)']\right)\gamma^*_T\equiv S_{k_T}(Z_t)'\gamma^*_T.$$

Thus

$$Var\left(\Delta(Z_t,\alpha_0)[v^*_T]\right)=\gamma^{*\prime}_T E\left[S_{k_T}(Z_t)S_{k_T}(Z_t)'\right]\gamma^*_T \tag{3.25}$$

and $\|v^*_T\|^2_{sd}=\gamma^{*\prime}_T Var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T S_{k_T}(Z_t)\right)\gamma^*_T$.

To verify Assumptions 3.1 and 3.2 for irregular functionals, it is handy to know the exact speed of divergence of $\|v^*_T\|^2$. We assume

Assumption 3.5 *The smallest and largest eigenvalues of R_{k_T} defined in (3.20) are bounded and bounded away from zero uniformly for all k_T.*

Assumption 3.5 imposes some regularity conditions on the sieve basis functions, which is a typical assumption in the linear sieve (or series) literature. For example, Newey (1997) makes similar assumptions in his paper on series LS regression.

Remark 3.2 *Assumption 3.5 implies that*

$$||v^*_T||^2\asymp||\gamma^*_T||^2_E\asymp||F_{k_T}||^2_E=||\frac{\partial f(\alpha_0)}{\partial\theta}||^2_E+||\frac{\partial f(\alpha_0)}{\partial h}[P_{k_T}(\cdot)]||^2_E.$$

Then: $f(\cdot)$ is regular at $\alpha=\alpha_0$ if $\lim_{k_T}||\frac{\partial f(\alpha_0)}{\partial h}[P_{k_T}(\cdot)]||^2_E<\infty$; $f(\cdot)$ is irregular at $\alpha=\alpha_0$ if $\lim_{k_T}||\frac{\partial f(\alpha_0)}{\partial h}[P_{k_T}(\cdot)]||^2_E=\infty$.

3.3.2 Examples

We first consider three typical linear functionals of semi-nonparametric models.

For the *Euclidean parameter functional* $f(\alpha) = \lambda'\theta$, we have $F_{k_T} = (\lambda', \mathbf{0}'_{k_T})'$ with $\mathbf{0}'_{k_T} = [0, ..., 0]_{1\times k_T}$, and hence $v^*_T = (v^{*\prime}_{\theta,T}, P_{k_T}(\cdot)'\beta^*_T)' \in \mathcal{V}_T$ with $v^*_{\theta,T} = I^{11}_T\lambda$, $\beta^*_T = I^{21}_T\lambda$, and

$$\|v^*_T\|^2 = F'_{k_T}R^{-1}_{k_T}F_{k_T} = \lambda' I^{11}_T\lambda.$$

If the largest eigenvalue of I^{11}_T, $\lambda_{\max}(I^{11}_T)$, is bounded above by a finite constant uniformly in k_T, then $\|v^*_T\|^2 \leq \lambda_{\max}(I^{11}_T) \times \lambda'\lambda < \infty$ uniformly in T, and the functional $f(\alpha) = \lambda'\theta$ is regular.

For the *evaluation functional* $f(\alpha) = h(\overline{x})$ for $\overline{x} \in \mathcal{X}$, we have $F_{k_T} = (\mathbf{0}'_{d_\theta}, P_{k_T}(\overline{x})')'$, and hence $v^*_T = (v^{*\prime}_{\theta,T}, P_{k_T}(\cdot)'\beta^*_T)' \in \mathcal{V}_T$ with $v^*_{\theta,T} = I^{12}_T P_{k_T}(\overline{x})$, $\beta^*_T = I^{22}_T P_{k_T}(\overline{x})$, and

$$\|v^*_T\|^2 = F'_{k_T}R^{-1}_{k_T}F_{k_T} = P'_{k_T}(\overline{x}) I^{22}_T P_{k_T}(\overline{x}).$$

So if the smallest eigenvalue of I^{22}_T, $\lambda_{\min}(I^{22}_T)$, is bounded away from zero uniformly in k_T, then $\|v^*_T\|^2 \geq \lambda_{\min}(I^{22}_T)||P_{k_T}(\overline{x})||^2_E \to \infty$, and the functional $f(\alpha) = h(\overline{x})$ is irregular.

For the *weighted integration functional* $f(\alpha) = \int_{\mathcal{X}} w(x)h(x)dx$ for a weighting function $w(x)$, we have $F_{k_T} = (\mathbf{0}'_{d_\theta}, \int_{\mathcal{X}} w(x)P_{k_T}(x)'dx)'$, and hence $v^*_T = (v^{*\prime}_{\theta,T}, P_{k_T}(\cdot)'\beta^*_T)'$ with $v^*_{\theta,T} = I^{12}_T \int_{\mathcal{X}} w(x)P_{k_T}(x)dx$, $\beta^*_T = I^{22}_T \int_{\mathcal{X}} w(x)P_{k_T}(x)dx$, and

$$\|v^*_T\|^2 = F'_{k_T}R^{-1}_{k_T}F_{k_T} = \left\{\int_{\mathcal{X}} w(x)P_{k_T}(x)dx\right\}' I^{22}_T \int_{\mathcal{X}} w(x)P_{k_T}(x)dx.$$

Suppose that the smallest and largest eigenvalues of I^{22}_T are bounded and bounded away from zero uniformly for all k_T. Then $||v^*_T||^2 \asymp ||\int_{\mathcal{X}} w(x)P_{k_T}(x)dx||^2_E$. Thus $f(\alpha) = \int_{\mathcal{X}} w(x)h(x)dx$ is regular if $\lim_{k_T} ||\int_{\mathcal{X}} w(x)P_{k_T}(x)dx||^2_E < \infty$; is irregular if $\lim_{k_T} ||\int_{\mathcal{X}} w(x)P_{k_T}(x)dx||^2_E = \infty$.

We finally consider an example of nonlinear functionals that arises in Example 2.2 when the parameter of interest is $\alpha_0 = (\theta'_0, h_0)'$ with $h^2_0 = f_Y$ being the true marginal density of Y_t. Consider the functional $f(\alpha) = h^2(\overline{y}) / \int_{-\infty}^{\infty} h^2(y)\, dy$. Note that $f(\alpha_0) = f_Y(\overline{y}) = h^2_0(\overline{y})$ and $h_0(\cdot)$ is approximated by the linear sieve $\mathcal{H}_T$ given in (2.6). Then $F_{k_T} = \left(\mathbf{0}'_{d_\theta}, \frac{\partial f(\alpha_0)}{\partial h}[P_{k_T}(\cdot)']\right)'$ with

$$\frac{\partial f(\alpha_0)}{\partial h}[P_{k_T}(\cdot)] = 2h_0(\overline{y})\left(P_{k_T}(\overline{y}) - h_0(\overline{y})\int_{-\infty}^{\infty} h_0(y)\, P_{k_T}(y)dy\right),$$

and hence $v^*_T = (v^{*\prime}_{\theta,T}, P_{k_T}(\cdot)'\beta^*_T)' \in \mathcal{V}_T$ with $v^*_{\theta,T} = I^{12}_T \frac{\partial f(\alpha_0)}{\partial h}[P_{k_T}(\cdot)]$, $\beta^*_T = I^{22}_T \frac{\partial f(\alpha_0)}{\partial h}[P_{k_T}(\cdot)]$, and

$$\|v^*_T\|^2 = F'_{k_T}R^{-1}_{k_T}F_{k_T} = \frac{\partial f(\alpha_0)}{\partial h}[P_{k_T}(\cdot)'] I^{22}_T \frac{\partial f(\alpha_0)}{\partial h}[P_{k_T}(\cdot)].$$

So if the smallest eigenvalue of I^{22}_T is bounded away from zero uniformly in k_T, then $\|v^*_T\|^2 \geq const. \times ||\frac{\partial f(\alpha_0)}{\partial h}[P_{k_T}(\cdot)]||^2_E \to \infty$, and the functional $f(\alpha) = h^2(\overline{y}) / \int_{-\infty}^{\infty} h^2(y)\, dy$ is irregular at $\alpha = \alpha_0$.

4 Asymptotic Variance of Sieve Estimators of Irregular Functionals

In this section, we derive the asymptotic expression of the "pre-asymptotic" sieve variance $\|v_T^*\|_{sd}^2$ for irregular functionals. We provide general sufficient conditions under which the asymptotic variance does not depend on the temporal dependence. We also show that evaluation functionals and some weighted integrals satisfy these conditions.

4.1 Exact Form of the Asymptotic Variance

By definition of the "pre-asymptotic" sieve variance $||v_T^*||_{sd}^2$ and the strict stationarity of the data $\{Z_t\}_{t=1}^T$, we have:

$$
\begin{aligned}
||v_T^*||_{sd}^2 &= Var\left(\Delta(Z,\alpha_0)[v_T^*]\right) + 2\sum_{t=1}^{T-1}\left(1-\frac{t}{T}\right)E\left(\Delta(Z_1,\alpha_0)[v_T^*]\Delta(Z_{t+1},\alpha_0)[v_T^*]\right) \qquad (4.1)\\
&= Var\left(\Delta(Z,\alpha_0)[v_T^*]\right) \times \left[1 + 2\sum_{t=1}^{T-1}\left(1-\frac{t}{T}\right)\rho_T^*(t)\right],
\end{aligned}
$$

where $\{\rho_T^*(t)\}$ is the autocorrelation coefficient of the triangular array $\{\Delta(Z_t,\alpha_0)[v_T^*]\}_{t\leq T}$:

$$
\rho_T^*(t) \equiv \frac{E\left(\Delta(Z_1,\alpha_0)[v_T^*]\Delta(Z_{t+1},\alpha_0)[v_T^*]\right)}{Var\left(\Delta(Z,\alpha_0)[v_T^*]\right)}. \qquad (4.2)
$$

Loosely speaking, one could say that the triangular array $\{\Delta(Z_t,\alpha_0)[v_T^*]\}_{t\leq T}$ is weakly dependent if

$$
\sum_{t=1}^{T-1}\left(1-\frac{t}{T}\right)\rho_T^*(t) = O(1). \qquad (4.3)
$$

Then we have $||v_T^*||_{sd}^2 = O\left\{Var\left(\Delta(Z,\alpha_0)[v_T^*]\right)\right\}$.

When $f(\cdot)$ is irregular, we have $\|v_T^*\| \to \infty$ as $\dim(\mathcal{V}_T) \to \infty$ (as $T \to \infty$). This and Assumption 3.2 imply that $\|v_T^*\|_{sd} \to \infty$, and so $Var\left(\Delta(Z,\alpha_0)[v_T^*]\right) \to \infty$ under (4.3) as $T \to \infty$ for irregular functionals. In this section we provide some sufficient conditions to ensure that, as $T \to \infty$, although the variance term blows up (i.e., $Var\left(\Delta(Z,\alpha_0)[v_T^*]\right) \to \infty$), the individual autocovariance term stays bounded or diverges at a slower rate, and hence the sum of autocorrelation coefficients becomes asymptotically negligible (i.e., $\sum_{t=1}^{T-1}\rho_T^*(t) = o(1)$). In the following we denote

$$
C_T \equiv \sup_{t\in[1,T)}\left|E\left\{\Delta(Z_1,\alpha_0)[v_T^*]\Delta(Z_{t+1},\alpha_0)[v_T^*]\right\}\right|.
$$

Assumption 4.1 *(i)* $\|v_T^*\| \to \infty$ *as* $T \to \infty$, *and* $\|v_T^*\|^2 / Var\left(\Delta(Z,\alpha_0)[v_T^*]\right) = O(1)$; *(ii) There is an increasing integer sequence* $\{d_T \in [2,T)\}$ *such that*

$$
\text{(a)}\ \frac{d_T C_T}{Var\left(\Delta(Z,\alpha_0)[v_T^*]\right)} = o(1) \quad \text{and} \quad \text{(b)}\ \left|\sum_{t=d_T}^{T-1}\left(1-\frac{t}{T}\right)\rho_T^*(t)\right| = o(1).
$$

More primitive sufficient conditions for Assumption 4.1 are given in the next subsection.

Theorem 4.1 *Let Assumption 4.1 hold. Then:* $\left| \frac{\|v_T^*\|_{sd}^2}{Var(\Delta(Z,\alpha_0)[v_T^*])} - 1 \right| = o(1)$*; If further Assumptions 3.1, 3.3 and 3.4 hold, then*

$$\frac{\sqrt{T}\left[f(\widehat{\alpha}_T) - f(\alpha_0)\right]}{\sqrt{Var\left(\Delta(Z,\alpha_0)[v_T^*]\right)}} \to_d N(0,1). \tag{4.4}$$

Theorem 4.1 shows that when the functional $f(\cdot)$ is irregular (i.e., $\|v_T^*\| \to \infty$), time series dependence does not affect the asymptotic variance of a general sieve M estimator $f(\widehat{\alpha}_T)$. Similar results have been proved for nonparametric kernel and local polynomial estimators of evaluation functionals of conditional mean and density functions. See for example, Robinson (1983) and Masry and Fan (1997). However, whether this is the case for general sieve M estimators of unknown functionals has been a long standing question. Theorem 4.1 gives a positive answer. This may seem surprising at first sight as sieve estimators are often regarded as global estimators while kernel estimators are regarded as local estimators.

One may conclude from Theorem 4.1 that the results and inference procedures for sieve estimators carry over from iid data to the time series case without modifications. However, this is true only when the sample size is large. Whether the sample size is large enough so that we can ignore the temporal dependence depends on the functional of interest, the strength of the temporal dependence, and the sieve basis functions employed. So it is ultimately an empirical question. In any finite sample, the temporal dependence does affect the sampling distribution of the sieve estimator. In the next section, we design an inference procedure that is easy to use and at the same time captures the time series dependence in finite samples.

4.2 Sufficient Conditions for Assumption 4.1

In this subsection, we first provide sufficient conditions for Assumption 4.1 for sieve M estimation of irregular functionals $f(\alpha_0)$ of general semi-nonparametric models. We then present additional low-level sufficient conditions for sieve M estimation of real-valued functionals of purely nonparametric models. We show that these sufficient conditions are satisfied for sieve M estimation of the evaluation and the weighted integration functionals.

4.2.1 Irregular functionals of general semi-nonparametric models

Given the closed-form expressions of $\|v_T^*\|$ and $Var\left(\Delta(Z,\alpha_0)[v_T^*]\right)$ in Subsection 3.3, it is easy to see that the following assumption implies Assumption 4.1.(i).

Assumption 4.2 *(i) Assumption 3.5 holds and* $\lim_{k_T} ||\frac{\partial f(\alpha_0)}{\partial h}[P_{k_T}(\cdot)]||_E^2 = \infty$*; (ii) The smallest eigenvalue of* $E\left[S_{k_T}(Z_t)S_{k_T}(Z_t)'\right]$ *in (3.25) is bounded away from zero uniformly for all* k_T*.*

Next, we provide some sufficient conditions for Assumption 4.1.(ii). Let $f_{Z_1,Z_t}(\cdot,\cdot)$ be the joint density of (Z_1, Z_t) and $f_Z(\cdot)$ be the marginal density of Z. Let $p \in [1,\infty)$. Define

$$\|\Delta(Z,\alpha_0)[v_T^*]\|_p \equiv \left(E\left\{|\Delta(Z,\alpha_0)[v_T^*]|^p\right\}\right)^{1/p}. \tag{4.5}$$

By definition, $\|\Delta(Z,\alpha_0)[v_T^*]\|_2^2 = Var\left(\Delta(Z,\alpha_0)[v_T^*]\right)$. The following assumption implies Assumption 4.1.(ii)(a).

Assumption 4.3 *(i)* $\sup_{t\geq 2}\sup_{(z,z')\in\mathcal{Z}\times\mathcal{Z}}\left|f_{Z_1,Z_t}(z,z')/\left[f_{Z_1}(z)f_{Z_t}(z')\right]\right| \leq C$ *for some constant* $C > 0$*; (ii)* $\|\Delta(Z,\alpha_0)[v_T^*]\|_1 / \|\Delta(Z,\alpha_0)[v_T^*]\|_2 = o(1)$.

Assumption 4.3.(i) is mild. When Z_t is a continuous random variable, it is equivalent to assuming that the bivariate copula density of (Z_1, Z_t) is bounded uniformly in $t \geq 2$. For irregular functionals (i.e., $\|v_T^*\| \nearrow \infty$), the $L^2(f_Z)$ norm $\|\Delta(Z,\alpha_0)[v_T^*]\|_2$ diverges (under Assumption 4.1.(i) or Assumption 4.2), Assumption 4.3.(ii) requires that the $L^1(f_Z)$ norm $\|\Delta(Z,\alpha_0)[v_T^*]\|_1$ diverge at a slower rate than the $L^2(f_Z)$ norm $\|\Delta(Z,\alpha_0)[v_T^*]\|_2$ as $k_T \to \infty$. In many applications the $L^1(f_Z)$ norm $\|\Delta(Z,\alpha_0)[v_T^*]\|_1$ actually remains bounded as $k_T \to \infty$ and hence Assumption 4.3.(ii) is trivially satisfied.

The following assumption implies Assumption 4.1.(ii)(b).

Assumption 4.4 *(i) The process* $\{Z_t\}_{t=1}^{\infty}$ *is strictly stationary strong-mixing with mixing coefficients* $\alpha(t)$ *satisfying* $\sum_{t=1}^{\infty} t^{\gamma}[\alpha(t)]^{\frac{\eta}{2+\eta}} < \infty$ *for some* $\eta > 0$ *and* $\gamma > 0$*; (ii) As* $k_T \to \infty$,

$$\frac{\|\Delta(Z,\alpha_0)[v_T^*]\|_1^{\gamma}\,\|\Delta(Z,\alpha_0)[v_T^*]\|_{2+\eta}}{\left\|\Delta(Z,\alpha_0)[v_T^*]\right\|_2^{\gamma+1}} = o(1).$$

The α-mixing condition in Assumption 4.4.(i) with $\gamma > \frac{\eta}{2+\eta}$ becomes Condition 2.(iii) in Masry and Fan (1997) for the pointwise asymptotic normality of their local polynomial estimator of a conditional mean function. See also Fan and Yao (2003, Condition 1.(iii) in section 6.6.2). In the next subsection, we illustrate that $\gamma > \frac{\eta}{2+\eta}$ is also sufficient for sieve M estimation of evaluation functionals of nonparametric time series models to satisfy Assumption 4.4.(ii). Instead of the strong mixing condition, we could also use other notions of weak dependence, such as the near epoch dependence used in Lu and Linton (2007) for the pointwise asymptotic normality of their local linear estimation of a conditional mean function.

Proposition 4.2 *Let Assumptions 4.2, 4.3 and 4.4 hold. Then:* $\sum_{t=1}^{T-1}|\rho_T^*(t)| = o(1)$ *and Assumption 4.1 holds.*

4.2.2 Irregular functionals of purely nonparametric models

In this subsection, we provide additional low-level sufficient conditions for Assumptions 4.1.(i), 4.3.(ii) and 4.4.(ii) for purely nonparametric models where the true unknown parameter is a real-valued function $h_0(\cdot)$ that solves $\sup_{h\in\mathcal{H}} E[\ell(Z_t, h(X_t))]$. This includes as a special case the nonparametric conditional mean model: $Y_t = h_0(X_t) + u_t$ with $E[u_t|X_t] = 0$. Our results can be easily generalized to more general settings with only some notational changes.

Let $\alpha_0 = h_0(\cdot) \in \mathcal{H}$ and let $f(\cdot) : \mathcal{H} \to \mathbb{R}$ be any functional of interest. By the results in Subsection 3.3, $f(h_0)$ has its sieve Riesz representor given by:

$$v_T^*(\cdot) = P_{k_T}(\cdot)'\beta_T^* \in \mathcal{V}_T \quad \text{with } \beta_T^* = R_{k_T}^{-1}\frac{\partial f(h_0)}{\partial h}[P_{k_T}(\cdot)],$$

where R_{k_T} is such that

$$\beta' R_{k_T}\beta = E\left(-r\left(Z_t, h_0\right)\left[\beta' P_{k_T}, P_{k_T}'\beta\right]\right) = \beta' E\left\{-\widetilde{r}\left(Z_t, h_0\left(X_t\right)\right) P_{k_T}(X_t) P_{k_T}(X_t)'\right\}\beta$$

for all $\beta \in \mathbb{R}^{k_T}$. Also, the score process can be expressed as

$$\Delta(Z_t, h_0)[v_T^*] = \widetilde{\Delta}(Z_t, h_0\left(X_t\right)) v_T^*(X_t) = \widetilde{\Delta}(Z_t, h_0\left(X_t\right)) P_{k_T}(X_t)'\beta_T^*.$$

Here the notations $\widetilde{\Delta}(Z_t, h_0(X_t))$ and $\widetilde{r}(Z_t, h_0(X_t))$ indicate the standard first-order and second-order derivatives of $\ell(Z_t, h(X_t))$ instead of functional pathwise derivatives (for example, we have $-\widetilde{r}(Z_t, h_0(X_t)) = 1$ and $\widetilde{\Delta}(Z_t, h_0(X_t)) = [Y_t - h_0(X_t)]/2$ in the nonparametric conditional mean model). Thus,

$$\|v_T^*\|^2 = E\left\{E[-\widetilde{r}\left(Z, h_0\left(X\right)\right)|X](v_T^*(X))^2\right\} = \beta_T^{*\prime} R_{k_T}\beta_T^* = \frac{\partial f(h_0)}{\partial h}[P_{k_T}(\cdot)'] R_{k_T}^{-1}\frac{\partial f(h_0)}{\partial h}[P_{k_T}(\cdot)],$$

$$Var\left(\Delta(Z, h_0)[v_T^*]\right) = E\left\{E([\widetilde{\Delta}(Z, h_0\left(X\right))]^2|X)(v_T^*(X))^2\right\}.$$

It is then obvious that Assumption 4.1.(i) is implied by the following condition.

Assumption 4.5 *(i)* $\inf_{x\in\mathcal{X}} E[-\widetilde{r}(Z, h_0(X))|X = x] \geq c_1 > 0$*; (ii)* $\sup_{x\in\mathcal{X}} E[-\widetilde{r}(Z, h_0(X))|X = x] \leq c_2 < \infty$*; (iii) the smallest and largest eigenvalues of* $E\{P_{k_T}(X)P_{k_T}(X)'\}$ *are bounded and bounded away from zero uniformly for all* k_T*, and* $\lim_{k_T} ||\frac{\partial f(h_0)}{\partial h}[P_{k_T}(\cdot)]||_E^2 = \infty$*; (iv)* $\inf_{x\in\mathcal{X}} E([\widetilde{\Delta}(Z, h_0(X))]^2|X = x) \geq c_3 > 0$.

It is easy to see that Assumptions 4.3.(ii) and 4.4.(ii) are implied by the following assumption.

Assumption 4.6 *(i)* $E\{|v_T^*(X)|\} = O(1)$*; (ii)* $\sup_{x\in\mathcal{X}} E\left[\left|\widetilde{\Delta}(Z, h_0(X))\right|^{2+\eta}|X = x\right] \leq c_4 < \infty$*; (iii)* $\left(E\{|v_T^*(X)|^2\}\right)^{-(2+\eta)(\gamma+1)/2} E\{|v_T^*(X)|^{2+\eta}\} = o(1)$.

It actually suffices to use *ess*-$\inf_x$ (or *ess*-$\sup_x$) instead of $\inf_x$ (or $\sup_x$) in Assumptions 4.5 and 4.6. We immediately obtain the following results.

Remark 4.3 *(1) Let Assumptions 4.3.(i), 4.4.(i), 4.5 and 4.6 hold. Then:*

$$\sum_{t=1}^{T-1} |\rho_T^*(t)| = o(1) \quad \text{and} \quad \left| \frac{\|v_T^*\|_{sd}^2}{Var\left(\Delta(Z,\alpha_0)[v_T^*]\right)} - 1 \right| = o\left(1\right).$$

(2) Assumptions 4.5 and 4.6.(ii) imply that

$$Var\left(\Delta(Z,\alpha_0)[v_T^*]\right) \asymp E\left\{(v_T^*(X))^2\right\} \asymp \|v_T^*\|^2 \asymp ||\beta_T^*||_E^2 \asymp ||\frac{\partial f(h_0)}{\partial h}[P_{k_T}(\cdot)]||_E^2 \to \infty;$$

hence Assumption 4.6.(iii) is satisfied if $E\{|P_{k_T}(X)'\beta_T^*|^{2+\eta}\}/||\beta_T^*||_E^{(2+\eta)(\gamma+1)} = o(1)$.

Assumptions 4.3.(i), 4.4.(i), 4.5 and 4.6.(ii) are all very standard low level sufficient conditions. In the following, we illustrate that Assumptions 4.6.(i) and (iii) are easily satisfied by two typical functionals of nonparametric models: the evaluation functional and the weighted integration functional.

Evaluation functionals. For the evaluation functional $f(h_0) = h_0(\overline{x})$ with $\overline{x} \in \mathcal{X}$, we have $\frac{\partial f(h_0)}{\partial h}[P_{k_T}(\cdot)] = P_{k_T}(\overline{x})$, $v_T^*(\cdot) = P_{k_T}(\cdot)'\beta_T^* = P_{k_T}(\cdot)'R_{k_T}^{-1}P_{k_T}(\overline{x})$. Then $\|v_T^*\|^2 = P_{k_T}'(\overline{x})R_{k_T}^{-1}P_{k_T}(\overline{x}) = v_T^*(\overline{x})$, and $\|v_T^*\|^2 \asymp ||P_{k_T}(\overline{x})||_E^2 \to \infty$ under Assumption 4.5.(i)(ii)(iii).

We first verify Assumption 4.6.(i): $\int_{x\in\mathcal{X}} |v_T^*(x)|\, f_X(x)\, dx = O(1)$. For the evaluation functional, we have, for any $v_T \in \mathcal{V}_T$:

$$\begin{aligned} v_T(\bar{x}) &= \langle h_T, v_T^* \rangle = E\left\{E[-\widetilde{r}(Z, h_0(X))\,|X] v_T(X) v_T^*(X)\right\} \\ &\equiv \int_{x\in\mathcal{X}} v_T(x)\, \delta_T(\bar{x}, x)\, dx, \end{aligned} \tag{4.6}$$

where

$$\begin{aligned} \delta_T(\bar{x}, x) &= E[-\widetilde{r}(Z, h_0(X))\,|X = x] v_T^*(x)\, f_X(x) \\ &= E[-\widetilde{r}(Z, h_0(X))\,|X = x] P_{k_T}'(\overline{x}) R_{k_T}^{-1} P_{k_T}(x) f_X(x). \end{aligned} \tag{4.7}$$

By equation (4.6) $\delta_T(\bar{x}, x)$ has the reproducing property on $\mathcal{V}_T$, so it behaves like the Dirac delta function $\delta(x - \bar{x})$ on $\mathcal{V}_T$. See Appendix B for further discussions about the properties of $\delta_T(\bar{x}, x)$. A direct implication is that $v_T^*(x)$ concentrates in a neighborhood around $x = \bar{x}$ and maintains the same positive sign in this neighborhood.

Using the definition of $\delta_T(\cdot,\cdot)$ in (4.7), we have

$$\int_{x\in\mathcal{X}} |v_T^*(x)|\, f_X(x)\, dx = \int_{x\in\mathcal{X}} \frac{sign\left(v_T^*(x)\right)}{E[-\widetilde{r}(Z, h_0(X))\,|X = x]} \delta_T(\bar{x}, x)\, dx,$$

where $sign(v_T^*(x)) = 1$ if $v_T^*(x) > 0$ and $sign(v_T^*(x)) = -1$ if $v_T^*(x) \leq 0$. Denote $b_T(x) \equiv \frac{sign\left(v_T^*(x)\right)}{E[-\widetilde{r}(Z,h_0(X))|X=x]}$. Then $\sup_{x\in\mathcal{X}} |b_T(x)| \leq c_1^{-1} < \infty$ under Assumption 4.5.(i) and $\int_{x\in\mathcal{X}} |v_T^*(x)|\, f_X(x)\, dx = \int_{x\in\mathcal{X}} b_T(x)\delta_T(\bar{x}, x)\, dx$.

If $b_T(x) \in \mathcal{V}_T$, then we have, using equation (4.6):

$$\int_{x\in\mathcal{X}} |v_T^*(x)| f_X(x)\,dx = b_T(\bar{x}) = \frac{sign\,(v_T^*(\bar{x}))}{E[-\widetilde{r}\,(Z, h_0(X))\,|X=\bar{x}]} \le c_1^{-1} = O(1).$$

If $b_T(x) \notin \mathcal{V}_T$ but can be approximated by a bounded function $\tilde{v}_T(x) \in \mathcal{V}_T$ such that

$$\int_{x\in\mathcal{X}} [b_T(x) - \tilde{v}_T(x)]\,\delta_T(\bar{x}, x)\,dx = o(1),$$

then, also using equation (4.6), we obtain:

$$\begin{aligned}\int_{x\in\mathcal{X}} |v_T^*(x)| f_X(x)\,dx &= \int_{x\in\mathcal{X}} \tilde{v}_T(x)\,\delta_T(\bar{x}, x)\,dx + \int_{x\in\mathcal{X}} [b_T(x) - \tilde{v}_T(x)]\,\delta_T(\bar{x}, x)\,dx \\ &= \tilde{v}_T(\bar{x}) + o(1) = O(1).\end{aligned}$$

Thus Assumption 4.6.(i) is satisfied.

Next, we verify Assumption 4.6.(iii). Using the definition of $\delta_T(\cdot,\cdot)$ in (4.7), we have

$$E\left\{|v_T^*(X)|^{2+\eta}\right\} = \int_{x\in\mathcal{X}} \frac{|v_T^*(x)|^{1+\eta}\,sign\,(v_T^*(x))}{E[-\widetilde{r}\,(Z, h_0(X))\,|X=x]}\delta_T(\bar{x}, x)\,dx.$$

Using the same argument for proving $\int_{x\in\mathcal{X}} |v_T^*(x)| f_X(x)\,dx = O(1)$, we can show that under mild conditions:

$$E\left\{|v_T^*(X)|^{2+\eta}\right\} \le \frac{|v_T^*(\bar{x})|^{1+\eta}}{E[-\widetilde{r}\,(Z, h_0(X))\,|X=\bar{x}]}(1+o(1)) = O\left(|v_T^*(\bar{x})|^{1+\eta}\right).$$

On the other hand,

$$E\left\{|v_T^*(X)|^2\right\} = \int_{x\in\mathcal{X}} |v_T^*(x)|^2 f_X(x)\,dx = \int_{x\in\mathcal{X}} \frac{v_T^*(x)}{E[-\widetilde{r}\,(Z, h_0(X))\,|X=x]}\delta_T(\bar{x}, x)\,dx \asymp v_T^*(\bar{x}).$$

Therefore

$$\left(E\left\{|v_T^*(X)|^2\right\}\right)^{-(2+\eta)(\gamma+1)/2} E\left\{|v_T^*(X)|^{2+\eta}\right\} \asymp |v_T^*(\bar{x})|^{1+\eta-(2+\eta)(\gamma+1)/2} = o(1)$$

if $1+\eta-(2+\eta)(\gamma+1)/2 < 0$, which is equivalent to $\gamma > \eta/(2+\eta)$. That is, when $\gamma > \eta/(2+\eta)$, Assumption 4.6.(iii) holds.

The above arguments employ the properties of delta sequences, i.e. sequences of functions that converge to the delta distribution. It follows from (4.6) that $\widehat{h}_T(\bar{x}) = \int_{x\in\mathcal{X}} \widehat{h}_T(x)\,\delta_T(\bar{x}, x)\,dx$. When the sample size is large, the sieve estimator of the evaluation functional effectively entails taking a weighted average of observations with the weights given by a delta sequence viz. $\delta_T(\bar{x}, x)$. The average is taken over a small neighborhood around $\bar{x}$ in the domain of X where there is no time series dependence. The observations X_t that fall in this neighborhood are not necessarily close to each other in time. Therefore this subset of observations has low dependence, and the contribution of their joint dependence to the asymptotic variance is asymptotically negligible.

Weighted integration functionals. For the weighted integration functional $f(h_0) = \int_{\mathcal{X}} w(x)h_0(x)dx$ for a weighting function $w(x)$, we have $\frac{\partial f(h_0)}{\partial h}[P_{k_T}(\cdot)] = \int_{\mathcal{X}} w(x)P_{k_T}(x)dx$, $v_T^*(\cdot) = P_{k_T}(\cdot)'\beta_T^* = P_{k_T}(\cdot)'R_{k_T}^{-1}\int_{\mathcal{X}} w(x)P_{k_T}(x)dx$. Suppose that the smallest and largest eigenvalues of $R_{k_T}^{-1}$ are bounded and bounded away from zero uniformly for all k_T. Then $\|v_T^*\|^2 \asymp ||\int_{\mathcal{X}} w(x)P_{k_T}(x)dx||_E^2$, thus $f(h_0) = \int_{\mathcal{X}} w(x)h_0(x)dx$ is irregular if $\lim_{k_T} ||\int_{\mathcal{X}} w(x)P_{k_T}(x)dx||_E^2 = \infty$.

For the weighted integration functional, we have, for any $h_T \in \mathcal{V}_T$:

$$\int_{\mathcal{X}} w(a)h_T(a)\,da = \langle h_T, v_T^*\rangle = E\left\{E[-\widetilde{r}(Z,h_0(X))\,|X]h_T(X)v_T^*(X)\right\}$$
$$\equiv \int_{x\in\mathcal{X}} h_T(x)\,\{\int_{\mathcal{X}} w(a)\delta_T(a,x)\,da\}dx,$$

where

$$\delta_T(a,x) = E[-\widetilde{r}(Z,h_0(X))\,|X=x]P_{k_T}'(a)R_{k_T}^{-1}P_{k_T}(x)f_X(x)\,.$$

Thus,

$$\int_{x\in\mathcal{X}} |v_T^*(x)|\,f_X(x)\,dx = \int_{\mathcal{X}}\left|\{\int_{\mathcal{X}} w(a)P_{k_T}'(a)da\}R_{k_T}^{-1}P_{k_T}(x)\right| f_X(x)\,dx$$
$$= \int_{a\in\mathcal{X}}\int_{x\in\mathcal{X}} w(a)sign\left\{w(a)\delta_T(a,x)\right\}\frac{\delta_T(a,x)}{E[-\widetilde{r}(Z,h_0(X))\,|X=x]}dadx$$
$$= \int_{a\in\mathcal{X}}\int_{x\in\mathcal{X}} b(a,x)\delta_T(a,x)\,dadx,$$

where

$$b(a,x) \equiv \frac{w(a)\,sign\left\{w(a)\delta_T(a,x)\right\}}{E[-\widetilde{r}(Z,h_0(X))\,|X=x]}.$$

If $b(\cdot,x)\in\mathcal{V}_T$, then, under Assumption 4.5.(i),

$$\int_{x\in\mathcal{X}} |v_T^*(x)|\,f_X(x)\,dx = \int_{x\in\mathcal{X}} b(x,x)dx \le C\int_{x\in\mathcal{X}} |w(x)|\,dx$$

for some constant C. If $b(\cdot,x)\notin\mathcal{V}_T$, then under some mild conditions, it can be approximated by $\tilde{w}_T(\cdot,x)\in\mathcal{V}_T$ with $|\tilde{w}_T(a,x)|\le C\,|b(a,x)|$ for some constant C and

$$\int_{a\in\mathcal{X}}\int_{x\in\mathcal{X}}\left[b(a,x)-\tilde{w}_T(a,x)\right]\delta_T(a,x)\,dadx = o(1).$$

In this case, we also have

$$\int_{x\in\mathcal{X}} |v_T^*(x)|\,f_X(x)\,dx \le \int_{x\in\mathcal{X}} \tilde{w}_T(x,x)dx \le C\int_{x\in\mathcal{X}} |w(x)|\,dx.$$

So if $\int_{x\in\mathcal{X}} |w(x)|\,dx < \infty$, we have $\int_{x\in\mathcal{X}} |v_T^*(x)|\,f_X(x)\,dx = O(1)$. Hence Assumption 4.6.(i) holds.

It remains to verify Assumption 4.6.(iii). Note that

$$\frac{E\{|P_{k_T}(X)'\beta_T^*|^{2+\eta}\}}{||\beta_T^*||_E^{(2+\eta)(\gamma+1)}} \le \frac{E\left(\|P_{k_T}(X)\|_E^{2+\eta}\right)\|\beta_T^*\|_E^{2+\eta}}{||\beta_T^*||_E^{(2+\eta)(\gamma+1)}} = \frac{E\left(\|P_{k_T}(X)\|_E^{2+\eta}\right)}{||\beta_T^*||_E^{(2+\eta)\gamma}}$$
$$= O\left[\frac{E\left(\|P_{k_T}(X)\|_E^{2+\eta}\right)}{\left\|\int_{x\in\mathcal{X}} w(x)\,P_{k_T}(x)\,dx\right\|_E^{(2+\eta)\gamma}}\right] = o(1)$$

for sufficiently large $\gamma > 1$, as $|| \int_{x\in\mathcal{X}} w(x) P_{k_T}(x) dx||_E \to \infty$. The minimum value of γ may depend on the weighting function $w(x)$. If $\sup_{x\in\mathcal{X}} \|P_{k_T}(x)\|_E^2 = O(k_T)$, which holds for many basis functions, and $|| \int_{x\in\mathcal{X}} w(x) P_{k_T}(x) dx||_E^2 \asymp k_T$, then $E\{|P_{k_T}(X)'\beta_T^*|^{2+\eta}\}/||\beta_T^*||_E^{(2+\eta)(\gamma+1)} = o(1)$ for any $\gamma > 1$. It follows from Remark 4.3 that Assumption 4.6.(iii) holds for the weighted integration functional.

5 Autocorrelation Robust Inference

In order to apply the asymptotic normality Theorem 3.1, we need an estimator of the sieve variance $\|v_T^*\|_{sd}^2$. In this section we propose a simple estimator of $\|v_T^*\|_{sd}^2$ and establish the asymptotic distributions of the associated t statistic and Wald statistic.

The theoretical sieve Riesz representor v_T^* is not known and has to be estimated. Let $\|\cdot\|_T$ denote the empirical norm induced by the following empirical inner product

$$\langle v_1, v_2\rangle_T = -\frac{1}{T}\sum_{t=1}^{T} r(Z_t, \widehat{\alpha}_T)[v_1, v_2], \tag{5.1}$$

for any $v_1, v_2 \in \mathcal{V}_T$. We define an empirical sieve Riesz representor $\widehat{v}_{\rho,T}^*$ of the functional $\frac{\partial f(\widehat{\alpha}_T)}{\partial\alpha}[\cdot]$ with respect to the empirical norm $\|\cdot\|_T$, i.e.

$$\frac{\partial f(\widehat{\alpha}_T)}{\partial\alpha}[\widehat{v}_T^*] = \sup_{v\in\mathcal{V}_T, v\neq 0} \frac{|\frac{\partial f(\widehat{\alpha}_T)}{\partial\alpha}[v]|^2}{\|v\|_T^2} < \infty \tag{5.2}$$

and

$$\frac{\partial f(\widehat{\alpha}_T)}{\partial\alpha}[v] = \langle v, \widehat{v}_T^*\rangle_T \tag{5.3}$$

for any $v \in \mathcal{V}_T$. We next show that the theoretical sieve Riesz representor v_T^* can be consistently estimated by the empirical sieve Riesz representor $\widehat{v}_T^*$ under the norm $\|\cdot\|$. In the following we denote $\mathcal{W}_T \equiv \{v \in \mathcal{V}_T : \|v\| = 1\}$.

Assumption 5.1 *Let $\{\epsilon_T^*\}$ be a positive sequence such that $\epsilon_T^* = o(1)$.*
(i) $\sup_{\alpha\in\mathcal{B}_T, v_1,v_2\in\mathcal{W}_T} E\{r(Z,\alpha)[v_1,v_2] - r(Z,\alpha_0)[v_1,v_2]\} = O(\epsilon_T^*)$;
(ii) $\sup_{\alpha\in\mathcal{B}_T, v_1,v_2\in\mathcal{W}_T} \mu_T\{r(Z,\alpha)[v_1,v_2]\} = O_p(\epsilon_T^*)$;
(iii) $\sup_{\alpha\in\mathcal{B}_T, v\in\mathcal{W}_T} \left|\frac{\partial f(\alpha)}{\partial\alpha}[v] - \frac{\partial f(\alpha_0)}{\partial\alpha}[v]\right| = O(\epsilon_T^*)$.

Assumption 5.1.(i) is a smoothness condition on the second derivative of the criterion function with respect to α. In the nonparametric LS regression model, we have $r(Z,\alpha)[v_1,v_2] = r(Z,\alpha_0)[v_1,v_2]$ for all α and v_1, v_2. Hence Assumption 5.1.(i) is trivially satisfied. Assumption 5.1.(ii) is a stochastic equicontinuity condition on the empirical process $T^{-1}\sum_{t=1}^{T} r(Z_t,\alpha)[v_1,v_2]$ indexed by α in the shrinking neighborhood $\mathcal{B}_T$ uniformly in $v_1, v_2 \in \mathcal{W}_T$. Assumption 5.1.(iii) puts some smoothness condition on the functional $\frac{\partial f(\alpha)}{\partial\alpha}[v]$ with respect to α in the shrinking neighborhood $\mathcal{B}_T$ uniformly in $v \in \mathcal{W}_T$.

Lemma 5.1 *Let Assumption 5.1 hold, then*

$$\left|\frac{\|\widehat{v}_T^*\|}{\|v_T^*\|} - 1\right| = O_p(\epsilon_T^*) \text{ and } \frac{\|\widehat{v}_T^* - v_T^*\|}{\|v_T^*\|} = O_p(\epsilon_T^*). \tag{5.4}$$

With the empirical estimator $\widehat{v}_T^*$ satisfying Lemma 5.1, we can now construct an estimate of the $\|v_T^*\|_{sd}^2$, which is the LRV of the score process $\Delta(Z_t, \alpha_0)[v_T^*]$. Many nonparametric LRV estimators are available in the literature. For kernel LRV estimators, see Newey and West (1987), Andrews (1991), Jansson (2004), Kiefer and Vogelsang (2005), Sun (2011b) and the numerous references therein. Nevertheless, to be consistent with our focus on the method of sieves and to derive a simple and accurate asymptotic approximation, we use an orthonormal series LRV (OS-LRV) estimator in this paper. The OS-LRV estimator has already been used in constructing autocorrelation robust inference on regular functionals of parametric time series models; see, e.g., Phillips (2005), Müller (2007), and Sun (2011a). Let $\{\phi_m\}_{m=1}^{\infty}$ be a sequence of orthonormal basis functions in $L_2[0,1]$. Define the orthogonal series projection

$$\widehat{\Lambda}_m = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \phi_m(\frac{t}{T}) \Delta(Z_t, \widehat{\alpha}_T)[\widehat{v}_T^*] \tag{5.5}$$

and construct the direct series estimator $\widehat{\Omega}_m = \widehat{\Lambda}_m^2$ for each $m = 1, 2, ..., M$ where $M \in \mathbb{Z}^+$. Taking a simple average of these direct estimators yields our OS-LRV estimator $||\widehat{v}_T^*||_{sd,T}^2$ of $\|v_T^*\|_{sd}^2$:

$$\|\widehat{v}_T^*\|_{sd,T}^2 \equiv \frac{1}{M} \sum_{m=1}^{M} \widehat{\Omega}_m = \frac{1}{M} \sum_{m=1}^{M} \widehat{\Lambda}_m^2, \tag{5.6}$$

where M, the number of orthonormal basis functions used, is the smoothing parameter in the present setting.

For irregular functionals, our asymptotic result in Section 4 suggests that we can ignore the temporal dependence and estimate $\|v_T^*\|_{sd}^2$ by

$$\widehat{\sigma}_v^2 = T^{-1} \sum_{t=1}^{T} \{\Delta(Z_t, \alpha_0)[\widehat{v}_T^*]\}^2.$$

However, when the sample size is small, there may still be considerable autocorrelation in the time series $\Delta(Z_t, \alpha_0)[v_T^*]$. To capture the possibly large but diminishing autocorrelation, we propose treating $\Delta(Z_t, \alpha_0)[v_T^*]$ as a generic time series and using the same formula as in (5.6) to estimate the asymptotic variance of $T^{-1/2} \sum_{t=1}^{T} \Delta(Z_t, \alpha_0)[v_T^*]$. That is, we estimate the variance based on the finite sample variance expression without going into deep asymptotics. We call the estimator the "pre-asymptotic" variance estimator. With a data-driven smoothing parameter choice, the "pre-asymptotic" variance estimator $||\widehat{v}_T^*||_{sd,T}^2$ should be close to $\widehat{\sigma}_v^2$ when the sample size is large. On the other hand, when the sample size is small, the "pre-asymptotic" variance estimator may provide a more accurate measure of the sampling variation of the plug-in sieve M estimator of irregular functionals. An extra benefit of the "pre-asymptotic" idea is that it allows us to treat regular and irregular functionals in a unified framework. So we do not distinguish regular and irregular functionals in the rest of this section.

To make statistical inference on a scalar functional $f(\alpha_0)$, we construct a t statistic as follows:

$$t_T \equiv \frac{\sqrt{T}\left[f(\widehat{\alpha}_T) - f(\alpha_0)\right]}{\left\|\widehat{v}_T^*\right\|_{sd,T}}. \tag{5.7}$$

We proceed to establish the asymptotic distribution of t_T when M is a fixed constant. To facilitate our development, we make the assumption below.

Assumption 5.2 *Let $\sqrt{T}\epsilon_T^* \xi_T = o(1)$ and the following conditions hold:*

(i) $\sup_{\tau\in[0,1]} \sup_{v\in\mathcal{W}_T,\ \alpha\in\mathcal{B}_T} \frac{1}{\sqrt{T}} \sum_{t=1}^{[T\tau]} \left(\Delta(Z_t,\alpha)[v] - \Delta(Z_t,\alpha_0)[v] - E\{\Delta(Z_t,\alpha)[v]\}\right) = o_p(1);$

(ii) $\sup_{v\in\mathcal{W}_T,\ \alpha\in\mathcal{B}_T} E\left\{\Delta(Z,\alpha)[v] - \Delta(Z_t,\alpha_0)[v] - r(Z,\alpha_0)[v,\alpha-\alpha_0]\right\} = O\left(\epsilon_T^*\xi_T\right);$

(iii) $\sup_{\tau\in[0,1]} \sup_{v\in\mathcal{W}_T} \frac{1}{\sqrt{T}} \sum_{t=1}^{[T\tau]} \Delta(Z_t,\alpha_0)[v] = O_p(1);$

(iv) $\frac{1}{\sqrt{T}} \sum_{t=1}^{[T\tau]} \Delta(Z_t,\alpha_0)[u_T^*] \to_d W(\tau)$ *where $W(\tau)$ is the standard Brownian motion process.*

Assumption 5.2.(i), (iii) and (iv) can be verified by applying the sequential Donsker's Theorem. Assumption 5.2.(ii) imposes a smoothness condition on the criterion function $\ell(Z,\alpha)$ with respect to α, and it can be verified by taking the first order expansion of $E\{\Delta(Z,\alpha)[v]\}$ around α_0 and using the convergence rate ξ_T. Assumption 5.2.(iv) is a slightly stronger version of Assumption 3.4.

Theorem 5.1 *Let $\int_0^1 \phi_m(r)\,dr = 0$, $\int_0^1 \phi_m(r)\,\phi_n(r)\,dr = 1\{m=n\}$ and $\phi_m(\cdot)$ be continuously differentiable. Under Assumptions 3.2, 3.3, 5.1 and 5.2, we have, for a fixed finite integer M:*

$$\|v_T^*\|_{sd}^{-1}\,\widehat{\Lambda}_m \to_d \int_0^1 \phi_m(\tau)\,dW(\tau).$$

If further Assumption 3.1 holds, then

$$t_T \equiv \frac{\sqrt{T}\left[f(\widehat{\alpha}_T) - f(\alpha_0)\right]}{\left\|\widehat{v}_T^*\right\|_{sd,T}} \to_d t(M),$$

where $t(M)$ is the t distribution with degree of freedom M.

Theorem 5.1 shows that when M is fixed, the t_T statistic converges weakly to a standard t distribution. This result is very handy as critical values from the t distribution can be easily obtained from statistical tables or standard software packages. This is an advantage of using the OS-LRV estimator. When $M \to \infty$, $t(M)$ approaches the standard normal distribution. So critical values from $t(M)$ can be justified even if $M = M_T \to \infty$ slowly with the sample size T. Theorem 5.1 extends the result of Sun (2011a) on robust OS-LRV estimation for parametric trend regressions to the case of general semi-nonparametric models.

In some economic applications, we may be interested in a vector of functionals $\mathbf{f} = (f_1, \ldots, f_q)'$ for some fixed finite $q \in \mathbb{Z}^+$. If each f_j satisfies Assumptions 3.1–3.3 and their Riesz representor $\mathbf{v}_T^* = (v_{1,T}^*, \ldots, v_{q,T}^*)$ satisfies the multivariate version of Assumption 3.4:

$$\|\mathbf{v}_T^*\|_{sd}^{-1} \sqrt{T} \mu_T \{\Delta(Z, \alpha_0) [\mathbf{v}_T^*]\} \to_d N(0, I_q),$$

then

$$\|\mathbf{v}_T^*\|_{sd}^{-1} \sqrt{T} [\mathbf{f}(\widehat{\alpha}_T) - \mathbf{f}(\alpha_0)] \to_d N(0, I_q), \tag{5.8}$$

where $\|\mathbf{v}_T^*\|_{sd}^2 = Var\left(\sqrt{T}\mu_T \Delta(Z, \alpha_0)[\mathbf{v}_T^*]\right)$ is a $q \times q$ matrix. A direct implication is that

$$T [\mathbf{f}(\widehat{\alpha}_T) - \mathbf{f}(\alpha_0)]' \|\mathbf{v}_T^*\|_{sd}^{-2} [\mathbf{f}(\widehat{\alpha}_T) - \mathbf{f}(\alpha_0)] \to_d \chi_q^2. \tag{5.9}$$

To estimate $\|\mathbf{v}_T^*\|_{sd}^2$, we define the orthogonal series projection

$$\widehat{\mathbf{\Lambda}}_m = \left(\widehat{\Lambda}_m^{(1)}, ..., \widehat{\Lambda}_m^{(q)}\right)'$$

with

$$\widehat{\Lambda}_m^{(j)} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \phi_m(\frac{t}{T}) \Delta(Z_t, \widehat{\alpha}_T)[\hat{v}_{j,T}^*],$$

where $\hat{v}_{j,T}^*$ denotes the empirical sieve Riesz representor of the functional $\frac{\partial f_j(\widehat{\alpha}_T)}{\partial \alpha}[\cdot]$ $(j = 1, ..., q)$. The OS-LRV estimator $||\widehat{\mathbf{v}}_T^*||_{sd,T}^2$ of the sieve variance $\|\mathbf{v}_T^*\|_{sd}^2$ is

$$\|\widehat{\mathbf{v}}_T^*\|_{sd,T}^2 = \frac{1}{M} \sum_{m=1}^{M} \widehat{\mathbf{\Lambda}}_m \widehat{\mathbf{\Lambda}}_m'.$$

To make statistical inference on $\mathbf{f}(\alpha_0)$, we construct the F test version of the Wald statistic as follows:

$$F_T \equiv T [\mathbf{f}(\widehat{\alpha}_T) - \mathbf{f}(\alpha_0)]' \|\widehat{\mathbf{v}}_T^*\|_{sd,T}^{-2} [\mathbf{f}(\widehat{\alpha}_T) - \mathbf{f}(\alpha_0)] / q. \tag{5.10}$$

We maintain Assumption 5.2 but replace Assumption 5.2(iv) by its multivariate version:

$$\|\mathbf{v}_T^*\|_{sd}^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{[T\tau]} \Delta(Z_t, \alpha_0) [\mathbf{v}_T^*] \to_d \mathbf{W}(\tau)$$

where $\mathbf{W}(\tau)$ is the q-dimensional standard Brownian motion process. Using a proof similar to that for Theorem 5.1, we can prove the theorem below.

Theorem 5.2 *Let $\int_0^1 \phi_m(r)\, dr = 0$, $\int_0^1 \phi_m(r)\, \phi_n(r)\, dr = 1\{m = n\}$ and $\phi_m(\cdot)$ be continuously differentiable. Let Assumptions 3.1, 3.2, 3.3, 5.1 and the multivariate version of Assumption 5.2 hold. Then, for a fixed finite integer M:*

$$\frac{M - q + 1}{M} F_T \to_d F_{q, M-q+1},$$

where $F_{q,M-q+1}$ is the F distribution with degree of freedom $(q, M - q + 1)$.

For $\sqrt{T}$ estimable parameters in parametric time series models, Sun (2011b) points out that the multiplicative modification $(M-q+1)/M$ is a type of Bartlett correction in addition to a distributional correction (i.e. using an F approximation instead of the standard χ^2-approximation).

The weak convergence of the F statistic can be rewritten as

$$F_T \to_d \frac{\chi_q^2/q}{\chi_{M-q+1}^2/(M-q+1)} \frac{M}{M-q+1} =^d F_{q,M-q+1} \frac{M}{M-q+1}.$$

As $M \to \infty$, both $\chi_{M-q+1}^2/(M-q+1)$ and $M/(M-q+1)$ converge to one. As a result, the limiting distribution approaches the standard χ^2 distribution. That is, under the sequential limit theory in which $T \to \infty$ for a fixed M and then $M \to \infty$, we obtain the standard χ^2 distribution as the sequential limiting distribution. When M is not very large or the number of the restrictions q is large, the second stage approximation in the sequential limit is likely to produce a large approximation error. This explains why the F approximation is more accurate, especially when M is relatively small and q is relatively large.

6 Numerical Equivalence of Asymptotic Variance Estimators

To compute the OS-LRV estimator in the previous section, we have to first find the empirical Riesz representor $\widehat{v}_T^*$, which is not very appealing to applied researchers. In this section we show that in finite samples we can directly apply the formula of the OS-LRV estimation derived under parametric assumptions and ignore the semiparametric/nonparametric nature of the model.

For simplicity, let the sieve space be $\mathcal{A}_T = \Theta \times \mathcal{H}_T$ with Θ a compact subset of $\mathbb{R}^{d_\theta}$ and $\mathcal{H}_T = \{h(\cdot) = P_{k_T}(\cdot)'\beta : \beta \in \mathbb{R}^{k_T}\}$. Let $\alpha_{0,T} = (\theta_0, P_{k_T}(\cdot)'\beta_{0,T}) \in int(\Theta) \times \mathcal{H}_T$. For $\alpha \in \mathcal{A}_T = \Theta \times \mathcal{H}_T$, we write $\ell(Z_t, \alpha) = \ell(Z_t, \theta, h(\cdot)) = \ell(Z_t, \theta, P_{k_T}(\cdot)'\beta)$ and define $\tilde{\ell}(Z_t, \gamma) = \ell(Z_t, \theta, P_{k_T}(\cdot)'\beta)$ as a function of $\gamma = (\theta', \beta')' \in \mathbb{R}^{d_\gamma}$ where $d_\gamma = d_\theta + d_\beta$ and $d_\beta \equiv k_T$. For any given Z_t, we view $\ell(Z_t, \alpha)$ as a functional of α on the infinite dimensional function space $\mathcal{A}$, but $\tilde{\ell}(Z_t, \gamma)$ as a function of γ on the Euclidian space $\mathbb{R}^{d_\gamma}$ whose dimension d_γ grows with the sample size but could be regarded as fixed in finite samples. By definition, for any $\alpha_1 = (\theta_1', P_{k_T}(\cdot)'\beta_1)'$ and $\alpha_2 = (\theta_2', P_{k_T}(\cdot)'\beta_2)'$, we have

$$\frac{\partial \tilde{\ell}(Z_t, \gamma_1)}{\partial \gamma'} (\gamma_2 - \gamma_1) = \Delta \ell(Z_t, \alpha_1) [\alpha_2 - \alpha_1] \tag{6.1}$$

where the left hand side is the regular derivative and the right hand side is the pathwise functional derivative. By the consistency of the sieve M estimator $\widehat{\alpha}_T = (\widehat{\theta}_T', P_{k_T}(\cdot)'\widehat{\beta}_T)$ for $\alpha_{0,T} = (\theta_0, P_{k_T}(\cdot)'\beta_{0,T})$, we have that $\widehat{\gamma}_T' \equiv (\widehat{\theta}_T', \widehat{\beta}_T')$ is a consistent estimator of $\gamma_{0,T}' = (\theta_0', \beta_{0,T}')$, then the first order conditions for the sieve M estimation can be represented as

$$\frac{1}{T} \sum_{t=1}^{T} \frac{\partial \tilde{\ell}(Z_t, \widehat{\gamma}_T)}{\partial \gamma} \approx 0. \tag{6.2}$$

These first order conditions are exactly the same as what we would get for parametric models with d_γ-dimensional parameter space.

Next, we pretend that $\tilde{\ell}(Z_t, \gamma)$ is a parametric criterion function on a finite dimensional space $\mathbb{R}^{d_\gamma}$. Using the OS-LRV estimator for the parametric M estimator based on the sample criterion function $\frac{1}{T}\sum_{t=1}^{T}\tilde{\ell}(Z_t, \gamma)$, we can obtain the typical sandwich asymptotic variance estimator for $\sqrt{T}\left(\widehat{\gamma}_T - \gamma_{0,T}\right)$ as follows:

$$\widehat{\Sigma}_T = \widehat{R}_T^{-1}\widehat{B}_T\widehat{R}_T^{-1},$$

where

$$\widehat{R}_T = -\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2\tilde{\ell}(Z_t, \widehat{\gamma}_T)}{\partial\gamma\partial\gamma'},$$
$$\widehat{B}_T = \frac{1}{M}\sum_{m=1}^{M}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_m\left(\frac{t}{T}\right)\frac{\partial\tilde{\ell}(Z_t, \widehat{\gamma}_T)}{\partial\gamma}\right]\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_m\left(\frac{t}{T}\right)\frac{\partial\tilde{\ell}(Z_t, \widehat{\gamma}_T)}{\partial\gamma'}\right].$$

Now suppose we are interested in a real-valued functional $f_{0,T} = f\left(\alpha_{0,T}\right) = f\left(\theta_0, P_{k_T}(\cdot)'\beta_{0,T}\right)$, which is estimated by the plug-in sieve M estimator $\widehat{f} = f\left(\widehat{\alpha}_T\right) = f(\widehat{\theta}_T, P_{k_T}(\cdot)'\widehat{\beta}_T)$. We compute the asymptotic variance of $\widehat{f}$ mechanically via the Delta method. We can then estimate the asymptotic variance of $\sqrt{T}(\widehat{f} - f_{0,T})$ by

$$\widehat{Var}(\widehat{f}) = \widehat{F}_{k_T}'\widehat{\Sigma}_T\widehat{F}_{k_T} = \widehat{F}_{k_T}'\widehat{R}_T^{-1}\widehat{B}_T\widehat{R}_T^{-1}\widehat{F}_{k_T}$$

where $\widehat{F}_{k_T} \equiv \left(\frac{\partial f(\widehat{\alpha}_T)}{\partial\theta'}, \frac{\partial f(\widehat{\alpha}_T)}{\partial h}[P_{k_T}(\cdot)']\right)'$. The following proposition shows that $\widehat{Var}(\widehat{f})$ is numerically identical to $||\widehat{v}_T^*||_{sd,T}^2$ defined in (5.6). The same result also holds for vector-valued functionals.

Proposition 6.1 *For any sample size T, we have the numerical identity:*

$$\|\widehat{v}_T^*\|_{sd,T}^2 = \widehat{Var}(\widehat{f}) = \widehat{F}_{k_T}'\widehat{R}_T^{-1}\widehat{B}_T\widehat{R}_T^{-1}\widehat{F}_{k_T}.$$

The numerical equivalence in variance estimators and point estimators (i.e., $\widehat{\gamma}_T$) implies that the corresponding test statistics are also numerically identical. Hence, we can use standard statistical packages designed for (misspecified) parametric models to compute test statistics for semi-nonparametric models. However, depending on the magnitude of sieve approximation errors, statistical inference and interpretation may be different across these two classes of models. Finally, we wish to point out that these numerical equivalence results are established only when the same finite dimensional linear sieve basis $P_{k_T}(\cdot)$ is used in approximating both the unknown function $h_0(\cdot)$ and the sieve Riesz representor v_T^*.

7 Simulation Study

In this section, we examine the accuracy of our inference procedures in Section 5 via Monte Carlo experiments. We consider a partial linear model as given in (2.2):

$$Y_t = X_{1t}'\theta_0 + \tilde{h}_0(\tilde{X}_{2t}) + u_t, t = 1, ..., T$$

where $\tilde{X}_{2t}$ and u_t are scalar processes, X_{1t} is a d-dimensional vector process with independent component X_{1t}^j. More specifically, $X_{1t}' = \left(X_{1t}^1, ..., X_{1t}^d\right)$ for $d = 4$ and

$$X_{1t}^j = \rho X_{1,t-1}^j + \sqrt{1-\rho^2}\varepsilon_{1t}^j,\ \tilde{X}_{2t} = \frac{1}{\sqrt{2d}}\left(X_{1t}^1 + ... + X_{1t}^d\right) + \frac{e_t}{\sqrt{2}}$$
$$e_t = \rho e_{t-1} + \sqrt{1-\rho^2}\varepsilon_{et},\ u_t = \rho u_{t-1} + \sqrt{1-\rho^2}\varepsilon_{ut}$$

where $\left(\varepsilon_{1t}^1, ..., \varepsilon_{1t}^d, \varepsilon_{et}, \varepsilon_{ut}\right)'$ are iid $N(0, I_{d+2})$. Here we have normalized $X_{1t}^j, \tilde{X}_{2t}$, and u_t to have zero mean and unit variance. We take $\rho = -0.75 : 0.25 : 0.75$.

Without loss of generality, we set $\theta_0 = 0$. We consider $\tilde{h}_0(\tilde{X}_{2t}) = \sin(\tilde{X}_{2t})$ and $\cos(\tilde{X}_{2t})$. Such choices are qualitatively similar to that in Härdle, Liang and Gao (2000, pages 52 and 139) who employ $\sin(\pi\tilde{X}_{2t})$. We focus on $\tilde{h}_0(\tilde{X}_{2t}) = \cos(\tilde{X}_{2t})$ below as it is harder to be approximated by a linear function around the center of the distribution of $\tilde{X}_{2t}$, but the qualitative results are the same for $\tilde{h}_0(\tilde{X}_{2t}) = \sin(\tilde{X}_{2t})$.

To estimate the model using the method of sieves on the unit interval [0,1], we first transform $\tilde{X}_{2t}$ into $[0, 1]$:

$$X_{2t} = \frac{\exp(\tilde{X}_{2t})}{1 + \exp(\tilde{X}_{2t})} \text{ or } \tilde{X}_{2t} = \log\left(\frac{X_{2t}}{1 - X_{2t}}\right).$$

Then $\tilde{h}_0(\tilde{X}_{2t}) = \cos(\log[X_{2t}(1 - X_{2t})^{-1}]) := h_0(X_{2t})$ for $h_0(x_2) = \cos(\log[x_2(1 - x_2)^{-1}])$. Let $P_{k_T}(x_2) = [p_1(x_2), ..., p_{k_T}(x_2)]'$ be a $k_T \times 1$ vector, where $\{p_j(x_2) : j \geq 1\}$ is a set of basis functions on $[0, 1]$. We approximate $h_0(x_2)$ by $P_{k_T}(x_2)'\beta$ for some $\beta = (\beta_1, ..., \beta_{k_T})' \in \mathbb{R}^{k_T}$. Denote $\mathbf{X}_t = \left(X_{1t}', P_{k_T}(X_{2t})'\right)$ a $1 \times (d + k_T)$ vector and $\mathbf{X}$ a $T \times (d + k_T)$ matrix:

$$\mathbf{X} = \begin{pmatrix} X_{11}' & p_1(X_{21}) & ... & p_{k_T}(X_{21}) \\ X_{12}' & p_1(X_{22}) & ... & p_{k_T}(X_{22}) \\ ... & ... & ... & ... \\ X_{1T}' & p_1(X_{2T}) & ... & p_{k_T}(X_{2T}) \end{pmatrix} := \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ ... \\ \mathbf{X}_T \end{pmatrix},$$

$\mathbf{Y} = (Y_1, ..., Y_T)'$, $\mathbf{U} = (u_1, ..., u_T)'$ and $\gamma = (\theta', \beta')'$. Then the sieve LS estimator of γ is

$$\widehat{\gamma}_T = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}.$$

In our simulation experiment, we use AIC and BIC to select k_T.

We employ our asymptotic theory to construct confidence regions for $\theta_{1:j} = (\theta_{01}, ..., \theta_{0j})'$. Equivalently, we test the null of $H_{0j} : \theta_{1:j} = 0$ against the alternative H_{1j} : at least one element of $\theta_{1:j}$ is not zero. Depending on the value of j, the number of joint hypotheses under consideration ranges from 1 to d. Let $\mathcal{R}_\theta(j)$ be the first j rows of the identity matrix I_{d+k_T}, then the sieve estimator of $\theta_{1:j} = \mathcal{R}_\theta(j)\gamma$ is

$$\widehat{\theta}_{1:j} = \mathcal{R}_\theta(j)\,\widehat{\gamma}_T, \tag{7.1}$$

and so

$$\begin{aligned}\sqrt{T}\left(\widehat{\theta}_{1:j} - \theta_{1:j}\right) &= \sqrt{T}\mathcal{R}_\theta(j)\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{U} + o_p(1) \\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathcal{R}_\theta(j)\left(\frac{\mathbf{X}'\mathbf{X}}{T}\right)^{-1}\mathbf{X}_t'u_t + o_p(1).\end{aligned}$$

Let $(\widehat{u}_1, ..., \widehat{u}_T)' = \widehat{\mathbf{U}} = \mathbf{Y} - \mathbf{X}\widehat{\gamma}_T$, $\widehat{\Delta}_{\theta t} = \mathcal{R}_\theta(j)\left(\mathbf{X}'\mathbf{X}/T\right)^{-1}\mathbf{X}_t'\widehat{u}_t \in \mathbb{R}^j$ and

$$\widehat{\Omega}_{\theta M} = \frac{1}{M}\sum_{m=1}^{M}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_m(\frac{t}{T})\widehat{\Delta}_{\theta t}\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_m(\frac{t}{T})\widehat{\Delta}_{\theta t}\right)'$$

be the OS-LRV estimator of the asymptotic variance Ω of $\sqrt{T}\left(\widehat{\theta}_{1:j} - \theta_{1:j}\right)$. Using the numerical equivalence result in Section 6, we can construct the F-test version of the Wald statistic as:

$$F_\theta(j) = \left(\sqrt{T}\mathcal{R}_\theta(j)\,\widehat{\gamma}_T\right)'\widehat{\Omega}_{\theta M}^{-1}\left(\sqrt{T}\mathcal{R}_\theta(j)\,\widehat{\gamma}_T\right)/j.$$

We refer to the test using critical values from the χ_j^2/j distribution as the chi-square test. We refer to the test using critical value $M(M-j+1)^{-1}\mathcal{F}_{j,M-j+1}^{\tau}$ as the F test, where $\mathcal{F}_{j,M-j+1}^{\tau}$ is the $(1-\tau)$ quantile of the F distribution $F_{j,M-j+1}$. Throughout the simulation, we use $\phi_{2m-1}(x) = \sqrt{2}\cos 2m\pi x$, $\phi_{2m}(x) = \sqrt{2}\sin 2m\pi x, m = 1, ..., M/2$ as the orthonormal basis functions for the OS-LRV estimation.

To perform either the chi-square test or the F test, we employ two different rules for selecting the smoothing parameter M. Under the first rule, we choose M to minimize the asymptotic mean square error of $\widehat{\Omega}_{\theta M}$. See Phillips (2005). The MSE-optimal M is given by

$$M_{MSE} = \left\lceil\left(\frac{tr\left[\left(I_{j^2} + \mathbb{K}_{jj}\right)(\Omega\otimes\Omega)\right]}{4vec(B)'vec(B)}\right)^{1/5}T^{4/5}\right\rceil,$$

where B is the asymptotic bias of $\widehat{\Omega}_M$, $\mathbb{K}_{jj}$ is the $j^2\times j^2$ commutation matrix, and $\lceil\cdot\rceil$ is the ceiling function. Under the second rule, we choose M to minimize the coverage probability error (CPE) of the confidence region based on the conventional chi-square test. The CPE-optimal M can be derived in the same way as that in Sun (2011b) where kernel LRV estimation is considered, with his kernel bandwidth $b = M^{-1}$. Setting $q = 2, c_1 = 0, c_2 = 1, p = j$ in Sun (2011b)'s formula, we obtain:

$$M_{CPE} = \left\lceil\left(\frac{j\left(\mathcal{X}_j^{\tau} + j\right)}{4\left|tr\left(B\Omega^{-1}\right)\right|}\right)^{\frac{1}{3}}T^{\frac{2}{3}}\right\rceil,$$

where $\mathcal{X}_j^\tau$ is the $(1-\tau)$ quantile of χ_j^2 distribution.

The parameters B and Ω in M_{MSE} and M_{CPE} are unknown but could be estimated by a standard plug-in procedure as in Andrews (1991). We fit an approximating VAR(1) model to the vector process $\widehat{\Delta}_{\theta t}$ and use the fitted model to estimate Ω and B.

We are also interested in making inference on $h_0(x)$. For each given x, let $\mathcal{R}_x = [0_{1\times d},\ P_{k_T}(x)']$. Then the sieve estimator of $h_0(x) = \mathcal{R}_x \gamma$ is

$$\widehat{h}(x) = \mathcal{R}_x \widehat{\gamma}_T. \tag{7.2}$$

We test $H_0 : h(x) = h_0(x)$ against $H_1 : h(x) \neq h_0(x)$ for $x = [1 + \exp(-\tilde{x}_2)]^{-1}$ and $\tilde{x}_2 = -2 : 0.1 : 2$. Since $\tilde{X}_{2t}$ is standard normal, this range of $\tilde{x}_2$ largely covers the support of $\tilde{X}_{2t}$. Like the estimator for the parametric part in (7.1), the above nonparametric estimator is also a linear combination of $\widehat{\gamma}_T$. As a result, we can follow exactly the same testing procedure as described above. To be more specific, we let

$$\widehat{\Delta}_{xt} = \mathcal{R}_x \left(\frac{\mathbf{X}'\mathbf{X}}{T}\right)^{-1} \mathbf{X}_t' \widehat{u}_t$$

and

$$\widehat{\Omega}_{xM} = \frac{1}{M}\sum_{m=1}^{M}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_m(\frac{t}{T})\widehat{\Delta}_{xt}\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_m(\frac{t}{T})\widehat{\Delta}_{xt}\right)',$$

which is the pre-asymptotic LRV estimator of $\sqrt{T}\left[\mathcal{R}_x\widehat{\gamma}_T - h_0(x)\right]$. Then the test statistic is

$$F_x = \left(\sqrt{T}\left[\mathcal{R}_x\widehat{\gamma}_T - h_0(x)\right]\right)' \widehat{\Omega}_{xM}^{-1}\left(\sqrt{T}\left[\mathcal{R}_x\widehat{\gamma}_T - h_0(x)\right]\right). \tag{7.3}$$

As in the inference for the parametric part, we select the smoothing parameter M based on the MSE and CPE criteria. It is important to point out that the approximating model and hence the data-driven smoothing parameter M are different for different hypotheses under consideration.

In Section 4, we have shown that, for evaluation functionals, the asymptotic variance does not depend on the time series dependence. So from an asymptotic point of view, we could also use

$$\widehat{\Omega}_{xM}^{*} = \frac{1}{T}\sum_{t=1}^{T}\widehat{\Delta}_{xt}\left(\widehat{\Delta}_{xt}\right)'$$

as the estimator for the asymptotic variance of $\sqrt{T}\left[\mathcal{R}_x\widehat{\gamma}_T - h_0(x)\right]$ and construct the F_x^* statistic accordingly. Here F_x^* is the same as F_x given in (7.3) but with $\widehat{\Omega}_{xM}$ replaced by $\widehat{\Omega}_{xM}^*$.

For the nonparametric part, we have three different inference procedures. The first two are both based on the F_x statistic with pre-asymptotic variance estimator, except that one uses χ_1^2 approximation and the other uses $F_{1,M}$ approximation. The third one is based on the F_x^* statistic and uses the χ_1^2 approximation. For ease of reference, we call the first two tests the pre-asymptotic χ^2 test and the pre-asymptotic F test, respectively. We call the test based on F_x^* and the χ_1^2 approximation the asymptotic χ^2 test.

Table 8.1 gives the empirical null rejection probabilities for testing $\theta_{1:j} = 0$ for $j = 1, 2, 3, 4$ for $\rho \geq 0$ under the CPE criterion. The number of simulation replications is 10,000. We consider two types of sieve basis functions to approximate $h(\cdot)$: the sine/cosine bases and the cubic spline bases with evenly spaced knots. The nominal rejection probability is $\tau = 5\%$ and k_T is selected by AIC. Results for BIC are qualitatively similar. Several patterns emerge from the table. First, the F test has a more accurate size than the chi-square test. This is especially true when the processes are persistent and the number of joint hypotheses being tested is large. Second, the size properties of the tests are not sensitive to the different sieve basis functions used for $h(\cdot)$. Finally, as the sample size increases, the size distortion of both the F test and the chi-square test decreases. It is encouraging that the size advantage of the F test remains even when $T = 500$.

Figures 8.2-8.4 present the empirical rejection probabilities for testing $H_0 : h(x) = h_0(x)$ against $H_0 : h(x) \neq h_0(x)$ for $x = [1 + \exp(-\tilde{x}_2)]^{-1}$ and $\tilde{x}_2 = -2 : 0.1 : 2$. As in Table 8.1, the CPE criterion is used to select the smoothing parameter M. It is clear that the asymptotic χ^2 test that ignores the time series dependence has a large size distortion when the process is persistent. This is true for both sample sizes $T = 100$ and $T = 500$ and for both sieve bases considered. Nevertheless, the asymptotic χ^2 test becomes less size-distorted as the sample size increases. This is consistent with our asymptotic theory. Compared to the pre-asymptotic χ^2 test, the pre-asymptotic F test has more accurate size when the sample size is not large and the processes are persistent. This, combined with the evidence for parametric inference, suggests that the pre-asymptotic F test is preferred for both parametric and nonparametric inference in practical situations.

For brevity, we do not report the simulation results when the MSE criterion is used to select the smoothing parameter M. We note that the superior performance of the pre-asymptotic F test relative to the pre-asymptotic χ^2 test and the conventional asymptotic χ^2 test remains. This is true for inference on both parametric and nonparametric components.

8 Conclusion

In this paper, we first establish the asymptotic normality of general plug-in sieve M estimators of possibly irregular functionals of semi-nonparametric time series models. We then obtain a surprising result that weak dependence does not affect the asymptotic variances of sieve M estimators of many irregular functionals including evaluation functionals and some weighted average derivative functionals.

Our theoretical result suggests that temporal dependence can be ignored in making inference on irregular functionals when the time series sample size is large. However, for small and moderate time series sample sizes, we find that it is better to conduct inference using the "pre-asymptotic" sieve variance estimation that accounts for temporal dependence.

We provide an accurate, autocorrelation robust inference procedure for sieve M estimators of both regular and irregular functionals. Our procedure is based on the "pre-asymptotic" scaled Wald statistic using the OS-LRV estimation and F approximation. The "pre-asymptotic" F approximations are asymptotically valid regardless of (1) whether the functional of interest is regular or not; (2) whether there is temporal dependence or not; and (3) whether the orthonormal series number of terms in computing the OS-LRV estimator is held fixed or not. Our scaled Wald statistics for possibly irregular functionals of semi-nonparametric models are shown to be numerically equivalent to the corresponding test statistics for regular functionals of parametric models, and hence are very easy to compute.

Table 8.1: Empirical Null Rejection Probabilities for the 5% F test and Chi-square Test

	$j=1$		$j=2$		$j=3$		$j=4$	
	F test	χ^2 Test	F test	χ^2 Test	F test	χ^2 Test	F test	χ^2 Test
$T=100$, Cosine and Sine Basis								
$\rho=0$	0.0687	0.0882	0.0723	0.0921	0.0885	0.1151	0.1329	0.1905
$\rho=0.25$	0.0706	0.1032	0.0825	0.1193	0.1085	0.1715	0.1679	0.2923
$\rho=0.50$	0.0717	0.1250	0.0884	0.1485	0.1255	0.2214	0.2012	0.3880
$\rho=0.75$	0.0744	0.1525	0.0973	0.1814	0.1458	0.2765	0.2338	0.4918
$T=100$, Spline Basis								
$\rho=0$	0.0680	0.0844	0.0711	0.0887	0.0830	0.1126	0.1212	0.1791
$\rho=0.25$	0.0647	0.0967	0.0743	0.1133	0.1011	0.1635	0.1518	0.2729
$\rho=0.50$	0.0668	0.1174	0.0799	0.1392	0.1176	0.2138	0.1880	0.3726
$\rho=0.75$	0.0655	0.1418	0.0867	0.1736	0.1358	0.2675	0.2137	0.4754
$T=500$, Cosine and Sine Basis								
$\rho=0$	0.0549	0.0596	0.0578	0.0621	0.0605	0.0695	0.0699	0.0898
$\rho=0.25$	0.0527	0.0593	0.0554	0.0646	0.0602	0.0798	0.0699	0.1145
$\rho=0.50$	0.0536	0.0628	0.0576	0.0720	0.0621	0.0898	0.0736	0.1354
$\rho=0.75$	0.0529	0.0659	0.0583	0.0789	0.0613	0.1003	0.0773	0.1651
$T=500$, Spline Basis								
$\rho=0$	0.0524	0.0552	0.0559	0.0607	0.0567	0.0683	0.0648	0.0858
$\rho=0.25$	0.0507	0.0582	0.0539	0.0625	0.0552	0.0743	0.0659	0.1078
$\rho=0.50$	0.0485	0.0584	0.0537	0.0663	0.0573	0.0850	0.0686	0.1327
$\rho=0.75$	0.0500	0.0614	0.0547	0.0739	0.0570	0.0964	0.0724	0.1581

Note: j is the number of joint hypotheses.

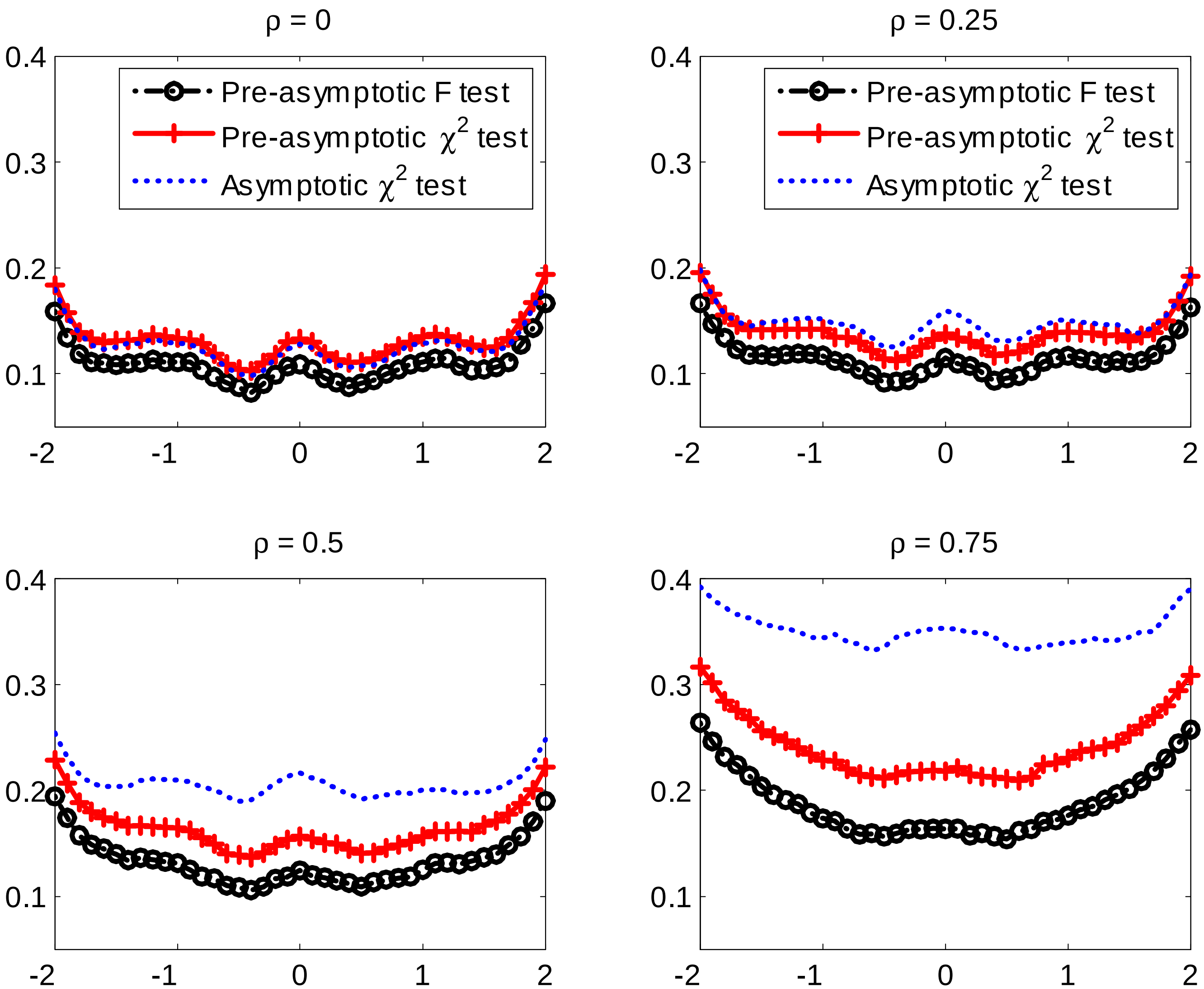

Figure 8.1: Plot of Empirical Rejection Probabilities Against the value of X_{2t} with Sine and Cosine Basis Functions and $T = 100$

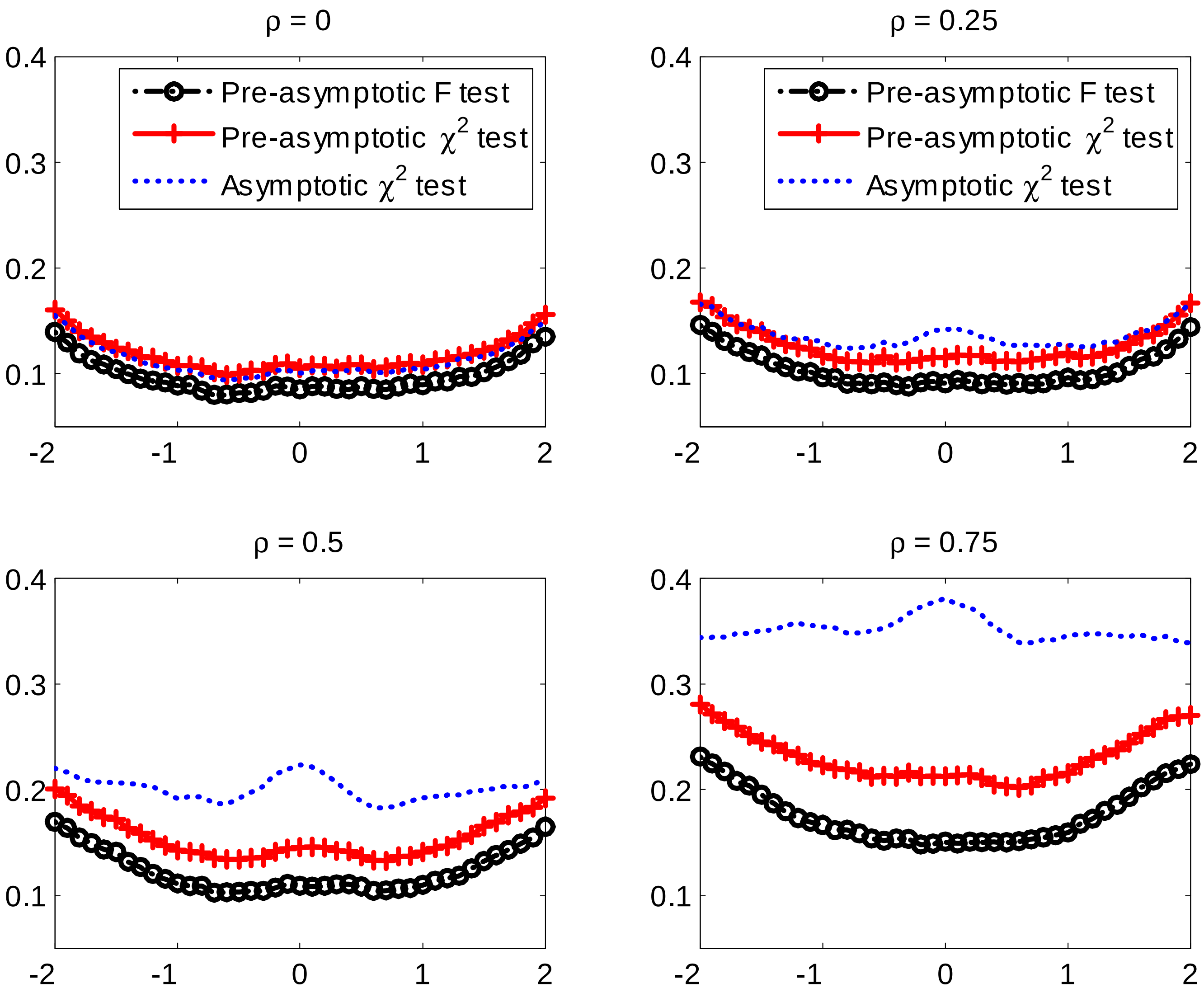

Figure 8.2: Plot of Empirical Rejection Probabilities Against the value of X_{2t} with Spline Basis Functions and $T = 100$

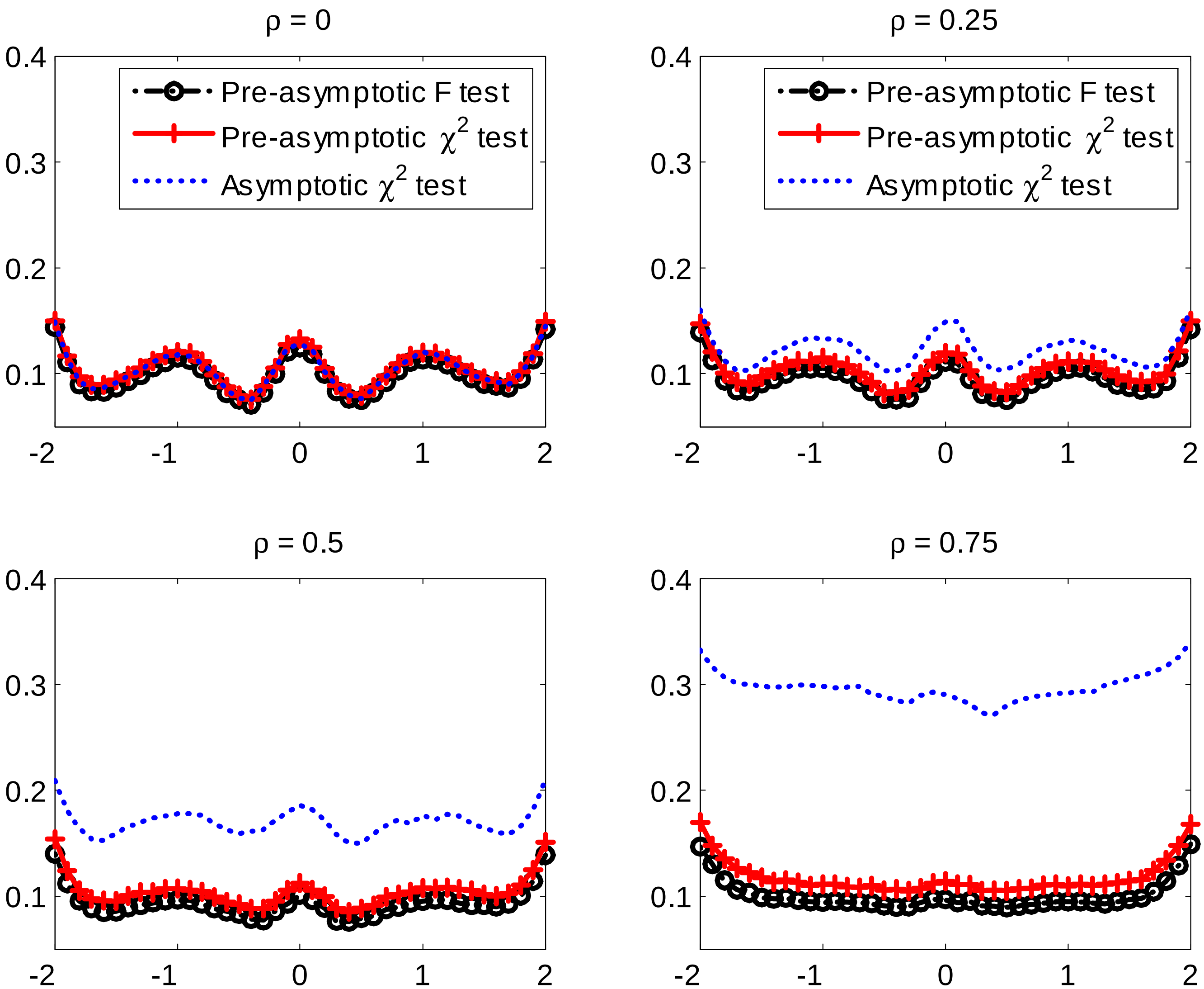

Figure 8.3: Plot of Empirical Rejection Probabilities Against the value of X_{2t} with Sine and Cosine Basis Functions and $T = 500$

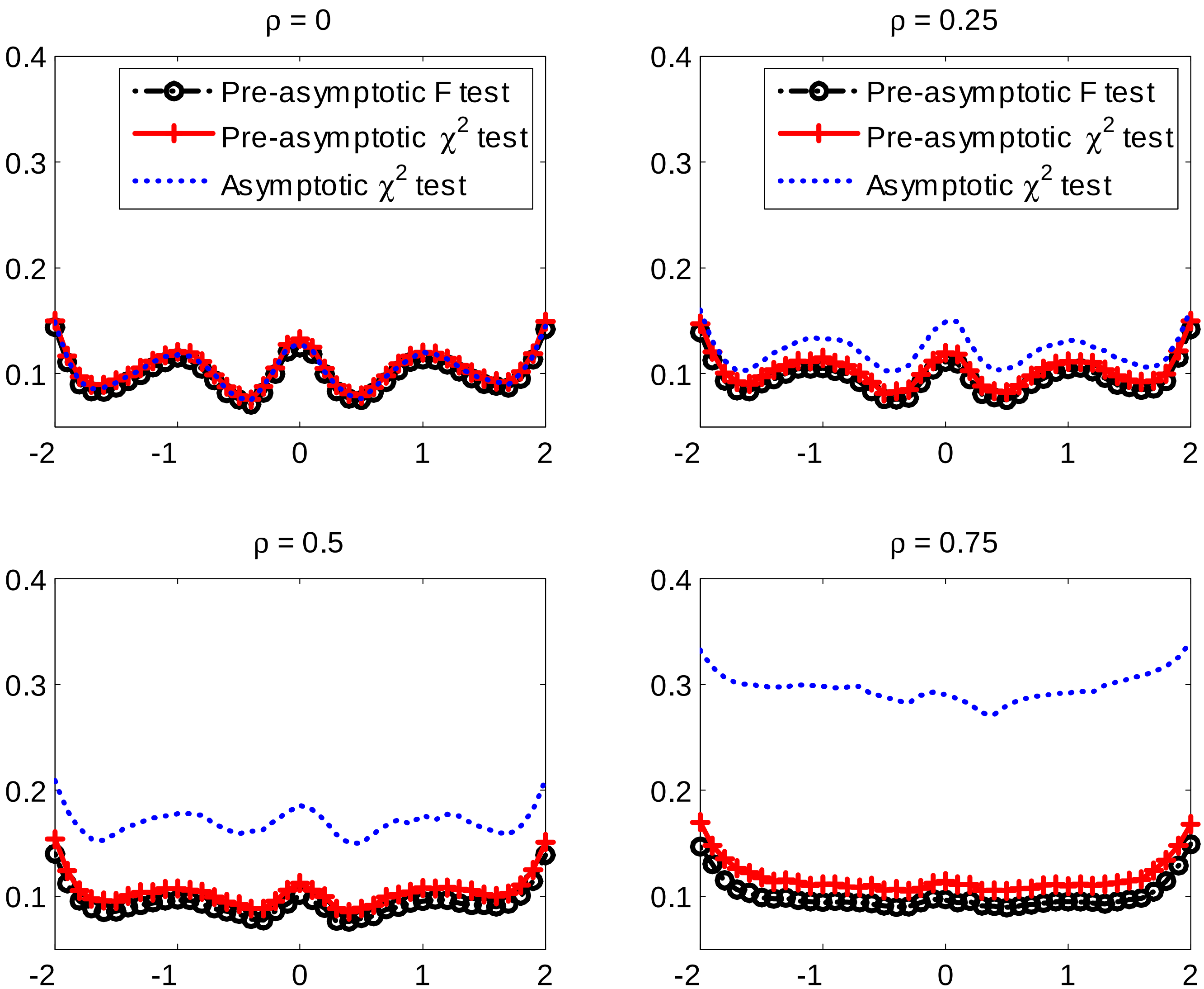

Figure 8.4: Plot of Empirical Rejection Probabilities Against the value of X_{2t} with Spline Basis Functions and $T = 500$

9 Appendix A: Mathematical Proofs

Proof of Theorem 3.1. For any $\alpha \in \mathcal{B}_T$, denote local alternative α_u^* of α as

$$\alpha_u^* = \alpha \pm \varepsilon_T u_T^* = \alpha \pm \frac{\varepsilon_T v_T^*}{\left\|v_T^*\right\|_{sd}},$$

where $\varepsilon_T = o(T^{-\frac{1}{2}})$. It is clear that if $\alpha \in \mathcal{B}_T$, then by the definition of α_u^*, Assumption 3.2 ($\|u_T^*\| = O(1)$), and the triangle inequality, we have $\alpha_u^* \in \mathcal{B}_T$. Since $\widehat{\alpha}_T \in \mathcal{B}_T$ with probability approaching one, we have that $\widehat{\alpha}_{u,T}^* = \widehat{\alpha}_T \pm \varepsilon_T u_T^* \in \mathcal{B}_T$ with probability approaching one. By the definition of $\widehat{\alpha}_T$, we have

$$\begin{aligned}
-O_p(\varepsilon_T^2) &\le \frac{1}{T}\sum_{t=1}^{T}\ell(Z_t,\widehat{\alpha}_T) - \frac{1}{T}\sum_{t=1}^{T}\ell(Z_t,\widehat{\alpha}_{u,T}^*) \\
&= E[\ell(Z_t,\widehat{\alpha}_T) - \ell(Z_t,\widehat{\alpha}_{u,T}^*)] + \mu_T\left\{\Delta(Z,\alpha_0)\left[\widehat{\alpha}_T - \widehat{\alpha}_{u,T}^*\right]\right\} \\
&+ \mu_T\left\{\ell(Z,\widehat{\alpha}_T) - \ell(Z,\widehat{\alpha}_{u,T}^*) - \Delta(Z,\alpha_0)\left[\widehat{\alpha}_T - \widehat{\alpha}_{u,T}^*\right]\right\} \\
&= E[\ell(Z_t,\widehat{\alpha}_T) - \ell(Z_t,\widehat{\alpha}_{u,T}^*)] \mp \mu_T\left\{\Delta(Z,\alpha_0)[\varepsilon_T u_T^*]\right\} + O_p(\varepsilon_T^2) \qquad (9.1)
\end{aligned}$$

by Assumption 3.3.(i)(ii). Next, by Assumptions 3.2 and 3.3.(iii) we have:

$$\begin{aligned}
&E[\ell(Z_t,\widehat{\alpha}_T) - \ell(Z_t,\widehat{\alpha}_{u,T}^*)] \\
&= \frac{||\widehat{\alpha}_T \pm \varepsilon_T u_T^* - \alpha_0||^2 - ||\widehat{\alpha}_T - \alpha_0||^2}{2} + O_p(\varepsilon_T^2) \\
&= \pm\varepsilon_T\langle\widehat{\alpha}_T - \alpha_0, u_T^*\rangle + \frac{1}{2}\varepsilon_T^2||u_T^*||^2 + O_p(\varepsilon_T^2) \\
&= \pm\varepsilon_T\langle\widehat{\alpha}_T - \alpha_0, u_T^*\rangle + O_p(\varepsilon_T^2).
\end{aligned}$$

Combining these with the definition of $\widehat{\alpha}_{u,T}^*$ and the inequality in (9.1), we deduce that

$$-O_p(\varepsilon_T^2) \le \pm\varepsilon_T\langle\widehat{\alpha}_T - \alpha_0, u_T^*\rangle \mp \varepsilon_T\mu_T\left\{\Delta(Z,\alpha_0)[u_T^*]\right\} + O_p(\varepsilon_T^2),$$

which further implies that

$$\langle\widehat{\alpha}_T - \alpha_0, u_T^*\rangle - \mu_T\left\{\Delta(Z,\alpha_0)[u_T^*]\right\} = O_p(\varepsilon_T) = o_p\left(T^{-1/2}\right). \qquad (9.2)$$

By the definition of $\alpha_{0,T}$, we have $\langle\alpha_{0,T} - \alpha_0, v\rangle = 0$ for any $v \in \mathcal{V}_T$, and hence

$$\langle\alpha_{0,T} - \alpha_0, u_T^*\rangle = 0.$$

Thus

$$\left|\sqrt{T}\left\langle\widehat{\alpha}_T - \alpha_{0,T}, u_T^*\right\rangle - \sqrt{T}\mu_T\left\{\Delta(Z,\alpha_0)\left[u_T^*\right]\right\}\right| = o_p(1). \qquad (9.3)$$

Next, we link $\sqrt{T}\frac{f(\widehat{\alpha}_T)-f(\alpha_{0,T})}{\left\|v_T^*\right\|_{sd}}$ with $\sqrt{T}\mu_T\left\{\Delta(Z,\alpha_0)\left[u_T^*\right]\right\}$ through the inner product $\sqrt{T}\left\langle\widehat{\alpha}_T - \alpha_{0,T}, u_T^*\right\rangle$.

By Assumptions 3.1.(i) and 3.2, and the Riesz representation theorem,

$$\begin{aligned}
&\sqrt{T}\frac{f(\widehat{\alpha}_T)-f(\alpha_{0,T})}{\left\|v_T^*\right\|_{sd}}\\
&=\sqrt{T}\frac{f(\widehat{\alpha}_T)-f(\alpha_0)-\frac{\partial f(\alpha_0)}{\partial\alpha}[\widehat{\alpha}_T-\alpha_0]}{\left\|v_T^*\right\|_{sd}}\\
&-\sqrt{T}\frac{f(\alpha_{0,T})-f(\alpha_0)-\frac{\partial f(\alpha_0)}{\partial\alpha}[\alpha_{0,T}-\alpha_0]}{\left\|v_T^*\right\|_{sd}}\\
&+\sqrt{T}\frac{\frac{\partial f(\alpha_0)}{\partial\alpha}\left[\widehat{\alpha}_T-\alpha_0\right]-\frac{\partial f(\alpha_0)}{\partial\alpha}\left[\alpha_{0,T}-\alpha_0\right]}{\left\|v_T^*\right\|_{sd}}\\
&=\sqrt{T}\frac{\frac{\partial f(\alpha_0)}{\partial\alpha}\left[\widehat{\alpha}_T-\alpha_{0,T}\right]}{\left\|v_T^*\right\|_{sd}}+o_p\left(1\right)=\sqrt{T}\frac{\langle\widehat{\alpha}_T-\alpha_{0,T},v_T^*\rangle}{\left\|v_T^*\right\|_{sd}}+o_p\left(1\right)\\
&=\sqrt{T}\left\langle\widehat{\alpha}_T-\alpha_{0,T},u_T^*\right\rangle+o_p\left(1\right).
\end{aligned} \tag{9.4}$$

It follows from (9.3) and (9.4) that

$$\left|\sqrt{T}\frac{f(\widehat{\alpha}_T)-f(\alpha_{0,T})}{\left\|v_T^*\right\|_{sd}}-\sqrt{T}\mu_T\left\{\Delta(Z,\alpha_0)\left[u_T^*\right]\right\}\right|=o_p(1), \tag{9.5}$$

which establishes the first result of the theorem.

For the second result, under Assumption 3.1.(ii), we get

$$\begin{aligned}
&\sqrt{T}\frac{f(\alpha_{0,T})-f(\alpha_0)}{\left\|v_T^*\right\|_{sd}}\\
&=\sqrt{T}\frac{f(\alpha_{0,T})-f(\alpha_0)-\frac{\partial f(\alpha_0)}{\partial\alpha}[\alpha_{0,T}-\alpha_0]}{\left\|v_T^*\right\|_{sd}}+\sqrt{T}\frac{\frac{\partial f(\alpha_0)}{\partial\alpha}[\alpha_{0,T}-\alpha_0]}{\left\|v_T^*\right\|_{sd}}=o_p(1).
\end{aligned}$$

This, (9.5) and Assumption 3.4 immediately imply that

$$\sqrt{T}\frac{f(\widehat{\alpha}_T)-f(\alpha_0)}{\left\|v_T^*\right\|_{sd}}=\sqrt{T}\mu_T\left\{\Delta(Z,\alpha_0)\left[u_T^*\right]\right\}+o_p(1)\to_d N(0,1).$$

■

Proof of Theorem 4.1. By Assumption 4.1.(i), we have: $0<Var\left(\Delta(Z,\alpha_0)[v_T^*]\right)\to\infty$. By equation (4.1) and definition of $\rho_T^*(t)$, we have:

$$\frac{||v_T^*||_{sd}^2}{Var\left(\Delta(Z,\alpha_0)[v_T^*]\right)}-1=J_{1,T}+J_{2,T}.$$

where

$$\begin{aligned}
J_{1,T}&=2\sum_{t=1}^{d_T}\frac{\left(1-\frac{t}{T}\right)E\left\{\Delta(Z_1,\alpha_0)[v_T^*]\Delta(Z_{t+1},\alpha_0)[v_T^*]\right\}}{Var\{\Delta(Z,\alpha_0)[v_T^*]\}}\text{ and}\\
J_{2,T}&=2\sum_{t=d_T+1}^{T-1}\left(1-\frac{t}{T}\right)\rho_T^*(t).
\end{aligned}$$

By Assumption 4.1.(ii)(a), we have:

$$|J_{1,T}| \leq \frac{2d_T C_T}{Var\{\Delta(Z,\alpha_0)[v_T^*]\}} = o(1). \tag{9.6}$$

Assumption 4.1.(ii)(b) immediately gives $|J_{2,T}| = o(1)$. Thus

$$\left| \frac{||v_T^*||_{sd}^2}{Var\left(\Delta(Z,\alpha_0)[v_T^*]\right)} - 1 \right| \leq |J_{1,T}| + |J_{2,T}| = o(1), \tag{9.7}$$

which establishes the first claim. This, Assumption 4.1.(i) and Theorem 3.1 together imply the asymptotic normality result in (4.4). ■

Proof of Proposition 4.2. For Assumption 4.1.(i), we note that Assumption 4.2.(i) implies $\|v_T^*\| \to \infty$ by Remark 3.2. Also under Assumption 4.2, we have:

$$\frac{\|v_T^*\|^2}{Var\left\{\Delta(Z,\alpha_0)[v_T^*]\right\}} = \frac{\gamma_T^{*\prime} R_{k_T} \gamma_T^*}{\gamma_T^{*\prime} E\left[S_{k_T}(Z) S_{k_T}(Z)'\right] \gamma_T^*} \leq \frac{\lambda_{\max}\left(R_{k_T}\right)}{\lambda_{\min}\left(E\left[S_{k_T}(Z) S_{k_T}(Z)'\right]\right)} = O(1),$$

where $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ denote the largest and the smallest eigenvalues of a matrix A. Hence $\|v_T^*\|^2 / Var\left\{\Delta(Z,h_0)[v_T^*]\right\} = O(1)$. For Assumption 4.1.(ii)(a), we have, under Assumption 4.3.(i),

$$\begin{aligned}
&\left|E\left\{\Delta(Z_1,\alpha_0)[v_T^*]\Delta(Z_t,\alpha_0)[v_T^*]\right\}\right| \\
&= \left| \int_{z_1 \in \mathcal{Z}} \int_{z_t \in \mathcal{Z}} \Delta(z_1,\alpha_0)\left[v_T^*\right] \Delta(z_t,\alpha_0)\left[v_T^*\right] f_{Z_1,Z_t}\left(z_1,z_t\right) dz_1 dz_t \right| \\
&= \left| \int_{z_1 \in \mathcal{Z}} \int_{z_t \in \mathcal{Z}} \Delta(z_1,\alpha_0)\left[v_T^*\right] \Delta(z_t,\alpha_0)\left[v_T^*\right] \frac{f_{Z_1,Z_t}\left(z_1,z_t\right)}{f_Z\left(z_1\right) f_Z\left(z_t\right)} f_Z\left(z_1\right) f_Z\left(z_t\right) dz_1 dz_t \right| \\
&\leq C \left(\int_{z_1 \in \mathcal{Z}} \left|\Delta(z_1,\alpha_0)\left[v_T^*\right]\right| f_Z\left(z_1\right) dz_1 \right)^2 = C \left\|\Delta(Z,\alpha_0)[v_T^*]\right\|_1^2,
\end{aligned}$$

which implies that

$$C_T \leq C \left\|\Delta(Z,\alpha_0)[v_T^*]\right\|_1^2.$$

This and Assumption 4.3.(ii) imply the existence of a growing $d_T \to \infty$ such that

$$\frac{d_T C_T}{\left\|\Delta(Z,\alpha_0)[v_T^*]\right\|_2^2} \to 0,$$

thus Assumption 4.1.(ii)(a) is satisfied. Under Assumption 4.4.(ii), we could further choose $d_T \to \infty$ to satisfy

$$\frac{\|\Delta(Z,\alpha_0)[v_T^*]\|_1^2 \times d_T}{\left\|\Delta(Z,\alpha_0)[v_T^*]\right\|_2^2} = o(1) \quad \text{and} \quad d_T^{\gamma} \asymp \frac{\|\Delta(Z,\alpha_0)[v_T^*]\|_{2+\eta}^2}{\left\|\Delta(Z,\alpha_0)[v_T^*]\right\|_2^2} \to \infty \text{ for some } \gamma > 0.$$

It remains to verify that such a choice of d_T and Assumption 4.4.(i) together imply Assumption 4.1.(ii)(b). Under Assumption 4.4.(i), $\{Z_t\}$ is a strictly stationary and strong-mixing process, $\{\Delta(Z_t,\alpha_0)[v_T^*] :$

$t \geq 1\}$ forms a triangular array of strong-mixing processes with the same decay rate. We can then apply Davydov's Lemma (Hall and Heyde 1980, Corollary A2) and obtain:

$$|E\{\Delta(Z_1,\alpha_0)[v_T^*]\Delta(Z_{t+1},\alpha_0)[v_T^*]\}| \leq 8[\alpha(t)]^{\frac{\eta}{2+\eta}} \|\Delta(Z,\alpha_0)[v_T^*]\|_{2+\eta}^2 .$$

Then:

$$\begin{aligned}
&\sum_{t=d_T}^{T-1} \left| \frac{E\{\Delta(Z_1,\alpha_0)[v_T^*]\Delta(Z_{t+1},\alpha_0)[v_T^*]\}}{\left\|\Delta(Z,\alpha_0)[v_T^*]\right\|_2^2} \right| \\
&\leq 8\frac{\|\Delta(Z,\alpha_0)[v_T^*]\|_{2+\eta}^2}{\left\|\Delta(Z,\alpha_0)[v_T^*]\right\|_2^2} \sum_{t=d_T}^{T-1} [\alpha(t)]^{\frac{\eta}{2+\eta}} \\
&\leq 8\frac{\|\Delta(Z,\alpha_0)[v_T^*]\|_{2+\eta}^2}{\left\|\Delta(Z,\alpha_0)[v_T^*]\right\|_2^2} d_T^{-\gamma} \sum_{t=d_T}^{T-1} t^{\gamma}[\alpha(t)]^{\frac{\eta}{2+\eta}} = o(1)
\end{aligned}$$

provided that

$$\frac{\|\Delta(Z,\alpha_0)[v_T^*]\|_{2+\eta}^2}{\left\|\Delta(Z,\alpha_0)[v_T^*]\right\|_2^2} d_T^{-\gamma} = O(1) \text{ and } \sum_{t=1}^{\infty} t^{\gamma}[\alpha(t)]^{\frac{\eta}{2+\eta}} < \infty \text{ for some } \gamma > 0,$$

which verifies Assumption 4.1.(ii)(b). Actually, we have established the stronger result: $\sum_{t=1}^{T-1}|\rho_T^*(t)| = o(1)$. ■

Proof of Lemma 5.1. First, using Assumptions 5.1.(i)-(ii) and the triangle inequality, we have

$$\begin{aligned}
&\sup_{\alpha\in\mathcal{B}_T}\sup_{v_1,v_2\in\mathcal{V}_T} \frac{\left|T^{-1}\sum_{t=1}^{T} r(Z_t,\alpha)[v_1,v_2] - E\{r(Z_t,\alpha_0)[v_1,v_2]\}\right|}{\|v_1\|\,\|v_2\|} \\
&\leq \sup_{\alpha\in\mathcal{B}_T}\sup_{v_1,v_2\in\mathcal{W}_T} \left|T^{-1}\sum_{t=1}^{T} r(Z_t,\alpha)[v_1,v_2] - E\{r(Z_t,\alpha)[v_1,v_2]\}\right| \\
&+ \sup_{\alpha\in\mathcal{B}_T}\sup_{v_1,v_2\in\mathcal{W}_T} |E\{r(Z,\alpha)[v_1,v_2] - r(Z,\alpha_0)[v_1,v_2]\}| = O_p(\epsilon_T^*).
\end{aligned} \tag{9.8}$$

Let $\alpha = \widehat{\alpha}_T$, $v_1 = \widehat{v}_T^*$ and $v_2 = v$. Then it follows from (9.8), the definitions of $\langle\cdot,\cdot\rangle$ and $\langle\cdot,\cdot\rangle_T$ that

$$\begin{aligned}
&\frac{\left|T^{-1}\sum_{t=1}^{T} r(Z_t,\widehat{\alpha}_T)[\widehat{v}_T^*,v] - E\{r(Z_t,\alpha_0)[\widehat{v}_T^*,v]\}\right|}{\left\|\widehat{v}_T^*\right\|\|v\|} \\
&= \left|\frac{\langle\widehat{v}_T^*,v\rangle_T - \langle\widehat{v}_T^*,v\rangle}{\left\|\widehat{v}_T^*\right\|\|v\|}\right| = O_p(\epsilon_T^*).
\end{aligned} \tag{9.9}$$

Combining this result with Assumption 5.1.(iii) and using

$$\frac{\partial f(\widehat{\alpha}_T)}{\partial\alpha}[v] = \langle\widehat{v}_T^*,v\rangle_T \text{ and } \frac{\partial f(\alpha_0)}{\partial\alpha}[v] = \langle v_T^*,v\rangle,$$

we can deduce that

$$O_p(\epsilon_T^*) = \sup_{v\in\mathcal{V}_T}\left|\frac{\frac{\partial f(\widehat{\alpha}_T)}{\partial\alpha}[v] - \frac{\partial f(\alpha_0)}{\partial\alpha}[v]}{\|v\|}\right|$$
$$= \sup_{v\in\mathcal{V}_T}\left|\frac{\langle\widehat{v}_T^*, v\rangle_T - \langle\widehat{v}_T^*, v\rangle}{\left\|\widehat{v}_T^*\right\|\|v\|}\|\widehat{v}_T^*\| + \frac{\langle\widehat{v}_T^* - v_T^*, v\rangle}{\|v\|}\right|$$
$$= \sup_{v\in\mathcal{V}_T}\left|\frac{\langle\widehat{v}_T^* - v_T^*, v\rangle}{\|v\|}\right| + O_p(\epsilon_T^*\|\widehat{v}_T^*\|). \tag{9.10}$$

This implies that

$$\sup_{v\in\mathcal{V}_T}\left|\frac{\langle\widehat{v}_T^* - v_T^*, v\rangle}{\|v\|}\right| = O_p(\epsilon_T^*\|\widehat{v}_T^*\|). \tag{9.11}$$

Letting $v = \widehat{v}_T^* - v_T^*$ in (9.11), we get

$$\frac{||\widehat{v}_T^* - v_T^*||}{\left\|v_T^*\right\|} = O_p\left(\epsilon_T^*\frac{||\widehat{v}_T^*||}{\left\|v_T^*\right\|}\right). \tag{9.12}$$

It follows from this result that

$$\left|\frac{||\widehat{v}_T^*||}{\left\|v_T^*\right\|} - 1\right| \leq \left\|\frac{\widehat{v}_T^*}{\left\|v_T^*\right\|} - \frac{v_T^*}{\left\|v_T^*\right\|}\right\|$$
$$= \frac{||\widehat{v}_T^* - v_T^*||}{\left\|v_T^*\right\|} = O_p\left(\epsilon_T^*\frac{||\widehat{v}_T^*||}{\left\|v_T^*\right\|}\right)$$
$$= O_p\left(\epsilon_T^*\left|\frac{||\widehat{v}_T^*||}{\left\|v_T^*\right\|} - 1\right|\right) + O_p\left(\epsilon_T^*\right) \tag{9.13}$$

from which we deduce that

$$\left|\frac{||\widehat{v}_T^*||}{\left\|v_T^*\right\|} - 1\right| = O_p(\epsilon_T^*). \tag{9.14}$$

Combining the results in (9.12), (9.13), and (9.14), we get

$$\frac{||\widehat{v}_T^* - v_T^*||}{\left\|v_T^*\right\|} = O_p(\epsilon_T^*), \tag{9.15}$$

as desired. ■

Proof of Theorem 5.1. Part (i) Define $\mathcal{S}_0 = 0$ and

$$\mathcal{S}_t = \frac{1}{\sqrt{T}}\sum_{\tau=1}^{t}\Delta(Z_\tau, \widehat{\alpha}_T)[\widehat{v}_T^*].$$

Then

$$\mathcal{S}_{[T\tau]} = \frac{1}{\sqrt{T}}\sum_{t=1}^{[T\tau]}\Delta(Z_t, \alpha_0)[\widehat{v}_T^*] + \tau\sqrt{T}E\left\{r(Z, \alpha_0)\left[\widehat{v}_T^*, \widehat{\alpha}_T - \alpha_0\right]\right\}$$
$$+ \frac{1}{\sqrt{T}}\sum_{t=1}^{[T\tau]}\left\{\Delta(Z_t, \widehat{\alpha}_T)[\widehat{v}_T^*] - \Delta(Z_t, \alpha_0)[\widehat{v}_T^*] - E\left[\Delta(Z_t, \widehat{\alpha}_T)[\widehat{v}_T^*]\right]\right\}$$
$$+ \tau\sqrt{T}E\left\{\Delta(Z_t, \widehat{\alpha}_T)[\widehat{v}_T^*] - r(Z, \alpha_0)\left[\widehat{v}_T^*, \widehat{\alpha}_T - \alpha_0\right]\right\}.$$

Using Assumption 5.2.(i)-(ii), we have

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[T\tau]}\left\{\Delta(Z_t,\widehat{\alpha}_T)[\widehat{v}_T^*]-\Delta(Z_t,\alpha_0)[\widehat{v}_T^*]-E\left[\Delta(Z_t,\widehat{\alpha}_T)[\widehat{v}_T^*]\right]\right\}=o_p(\|\widehat{v}_T^*\|),$$

and

$$\sqrt{T}E\left\{\Delta(Z_t,\widehat{\alpha}_T)[\widehat{v}_T^*]-r(Z,\alpha_0)\left[\widehat{v}_T^*,\widehat{\alpha}_T-\alpha_0\right]\right\}=O_p\left(\sqrt{T}\epsilon_T^*\xi_T\left\|\widehat{v}_T^*\right\|\right).$$

So

$$\begin{aligned}\mathcal{S}_{[T\tau]}&=\frac{1}{\sqrt{T}}\sum_{t=1}^{[T\tau]}\Delta(Z_t,\alpha_0)[\widehat{v}_T^*]+\tau\sqrt{T}E\left\{r(Z,\alpha_0)\left[\widehat{v}_T^*,\widehat{\alpha}_T-\alpha_0\right]\right\}\\&+o_p(\|\widehat{v}_T^*\|)+O_p\left(\sqrt{T}\xi_T^2\left\|\widehat{v}_T^*\right\|\right)\\&=\frac{1}{\sqrt{T}}\sum_{t=1}^{[T\tau]}\Delta(Z_t,\alpha_0)[v_T^*]+\frac{1}{\sqrt{T}}\sum_{t=1}^{[T\tau]}\Delta(Z_t,\alpha_0)[\widehat{v}_T^*-v_T^*]\\&-\tau\sqrt{T}\left\langle v_T^*,\widehat{\alpha}_T-\alpha_0\right\rangle-\tau\sqrt{T}\left\langle \widehat{v}_T^*-v_T^*,\widehat{\alpha}_T-\alpha_0\right\rangle\\&+o_p(\|\widehat{v}_T^*\|)+O_p\left(\sqrt{T}\epsilon_T^*\xi_T\left\|\widehat{v}_T^*\right\|\right).\end{aligned}\tag{9.16}$$

Under Assumptions 3.2 and 3.3, we can invoke equation (9.2) in the proof of Theorem 3.1 to deduce that

$$\sqrt{T}\left\|v_T^*\right\|_{sd}^{-1}\left\langle v_T^*,\widehat{\alpha}_T-\alpha_0\right\rangle=\frac{1}{\sqrt{T}}\left\|v_T^*\right\|_{sd}^{-1}\sum_{t=1}^{T}\Delta(Z_t,\alpha_0)[v_T^*]+o_p(1).\tag{9.17}$$

Using Lemma 5.1 and the Hölder inequality, we get

$$\left|\sqrt{T}\left\langle\widehat{v}_T^*-v_T^*,\widehat{\alpha}_T-\alpha_0\right\rangle\right|\le\sqrt{T}\left\|\widehat{v}_T^*-v_T^*\right\|\left\|\widehat{\alpha}_T-\alpha_0\right\|=O_p(\sqrt{T}\left\|v_T^*\right\|\epsilon_T^*\xi_T).\tag{9.18}$$

Next, by Assumption 5.2.(iii) and Lemma 5.1,

$$\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{[T\tau]}\Delta(Z_t,\alpha_0)[\widehat{v}_T^*-v_T^*]\right|\le\left\|\widehat{v}_T^*-v_T^*\right\|\sup_{v\in\mathcal{W}_T}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{[T\tau]}\Delta(Z_t,\alpha_0)[v]\right|=O_p(\left\|v_T^*\right\|\epsilon_T^*).\tag{9.19}$$

Now, using Lemma 5.1, (9.16)-(9.19), Assumption 3.2 ($\|v_T^*\|=O\left(\|v_T^*\|_{sd}\right)$), Assumption 5.2.(iv) and $\sqrt{T}\epsilon_T^*\xi_T=o(1)$, we can deduce that

$$\begin{aligned}\left\|v_T^*\right\|_{sd}^{-1}\mathcal{S}_{[T\tau]}&=\frac{1}{\sqrt{T}}\left\|v_T^*\right\|_{sd}^{-1}\sum_{t=1}^{[T\tau]}\Delta(Z_t,\alpha_0)[v_T^*]-\frac{\tau}{\sqrt{T}}\left\|v_T^*\right\|_{sd}^{-1}\sum_{t=1}^{T}\Delta(Z_t,\alpha_0)[v_T^*]+o_p(1)\\&\to_d W(\tau)-\tau W(1):=B(\tau).\end{aligned}$$

We use the above result to finish the proof. Note that

$$
\begin{aligned}
\|v_T^*\|_{sd}^{-1} \widehat{\Lambda}_m &= \|v_T^*\|_{sd}^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \phi_m\left(\frac{t}{T}\right) \Delta(Z_t, \widehat{\alpha}_T)[\widehat{v}_T^*] \\
&= \|v_T^*\|_{sd}^{-1} \left[\sum_{t=1}^{T} \phi_m\left(\frac{t}{T}\right) \mathcal{S}_t - \sum_{t=1}^{T} \phi_m\left(\frac{t}{T}\right) \mathcal{S}_{t-1} \right] \\
&= \|v_T^*\|_{sd}^{-1} \left[\sum_{t=1}^{T} \phi_m\left(\frac{t}{T}\right) \mathcal{S}_t - \sum_{t=0}^{T-1} \phi_m\left(\frac{t+1}{T}\right) \mathcal{S}_t \right] \\
&= \frac{1}{T} \sum_{t=1}^{T-1} \left[\frac{\phi_m(t/T) - \phi_m((t+1)/T)}{1/T} \right] \|v_T^*\|_{sd}^{-1} \mathcal{S}_t + \phi_m(1) \|v_T^*\|_{sd}^{-1} \mathcal{S}_T.
\end{aligned}
$$

Since ϕ_m is continuously differentiable, we can invoke the continuous mapping theorem to obtain

$$
\|v_T^*\|_{sd}^{-1} \widehat{\Lambda}_m \to_d - \int_0^1 \phi_m'(\tau) B(\tau) d\tau.
$$

Using integration by parts, we can show that

$$
- \int_0^1 \phi_m'(\tau) B(\tau) d\tau = - \int_0^1 B(\tau) d\phi_m(\tau) = \int_0^1 \phi_m(\tau) dB(\tau).
$$

Hence

$$
\|v_T^*\|_{sd}^{-1} \widehat{\Lambda}_m \to_d \int_0^1 \phi_m(\tau) dB(\tau) = \int_0^1 \phi_m(\tau) dW(\tau)
$$

where the equality follows from the assumption that $\int_0^1 \phi_m(\tau) d\tau = 0$.

Part (ii) It follows from part (i) that

$$
\|v_T^*\|_{sd}^{-1} \|\widehat{v}_T^*\|_{sd,T}^2 \|v_T^*\|_{sd}^{-1} = \frac{1}{M} \sum_{m=1}^{M} \left(\|v_T^*\|_{sd}^{-1} \widehat{\Lambda}_m \right)^2 \to_d \frac{1}{M} \sum_{m=1}^{M} \left[\int_0^1 \phi_m(\tau) dB(\tau) \right]^2. \tag{9.20}
$$

which, combining Theorem 3.1 and Slutsky's Theorem, further implies that

$$
\begin{aligned}
t_T &= \frac{\sqrt{T}[f(\widehat{\alpha}_T) - f(\alpha_0)]}{\|v_T^*\|_{sd}} \Bigg/ \frac{\|\widehat{v}_T^*\|_{sd,T}}{\|v_T^*\|_{sd}} \\
&= \frac{\sqrt{T}[f(\widehat{\alpha}_T) - f(\alpha_0)]}{\|v_T^*\|_{sd}} \Bigg/ \sqrt{\frac{1}{M} \sum_{m=1}^{M} \left(\|v_T^*\|_{sd}^{-1} \widehat{\Lambda}_m \right)^2} \\
&\to_d \frac{W(1)}{\sqrt{M^{-1} \sum_{m=1}^{M} \left[\int_0^1 \phi_m(\tau) dW(\tau) \right]^2}}.
\end{aligned} \tag{9.21}
$$

Note that both $\int_0^1 \phi_m(\tau) dW(\tau)$ and $W(1)$ are normal. In addition, $\text{cov}\Big(W(1), \int_0^1 \phi_m(\tau) dW(\tau)\Big) = \int_0^1 \phi_m(\tau) d\tau = 0$ and

$$
cov\left(\int_0^1 \phi_m(\tau) dW(\tau), \int_0^1 \phi_n(\tau) dW(\tau) \right) = \int_0^1 \phi_m(\tau) \phi_n(\tau) d\tau = 1\{m = n\}.
$$

So $W(1)$ is independent of $M^{-1}\sum_{m=1}^{M}\left[\int_0^1 \phi_m(\tau)\,dW(\tau)\right]^2$ and $\sum_{m=1}^{M}\left[\int_0^1 \phi_m(\tau)\,dW(\tau)\right]^2 \sim \chi^2_M$. This implies that

$$\frac{W(1)}{\sqrt{M^{-1}\sum_{m=1}^{M}\left[\int_0^1 \phi_m(\tau)\,dW(\tau)\right]^2}} \sim t(M) \text{ and } t_T \to_d t(M).$$

■

Proof of Theorem 5.2. Using similar arguments as in deriving Theorem 5.1, we can show that

$$\|\mathbf{v}_T^*\|_{sd}^{-1}\,\widehat{\Lambda}_m \to_d \int_0^1 \phi_m(\tau)\,d\mathbf{W}(\tau). \tag{9.22}$$

It then follows that

$$\begin{aligned}
&\|\mathbf{v}_T^*\|_{sd}^{-1}\,\|\widehat{\mathbf{v}}_T^*\|_{sd,T}^2\left(\|\mathbf{v}_T^*\|_{sd}^{-1}\right)' \\
&= \frac{1}{M}\sum_{m=1}^{M}\left(\|\mathbf{v}_T^*\|_{sd}^{-1}\,\widehat{\Lambda}_m\right)\left(\|\mathbf{v}_T^*\|_{sd}^{-1}\,\widehat{\Lambda}_m\right)' \\
&\to_d \frac{1}{M}\sum_{m=1}^{M}\left[\int_0^1 \phi_m(\tau)\,d\mathbf{W}(\tau)\right]\left[\int_0^1 \phi_m(\tau)\,d\mathbf{W}(\tau)\right]'.
\end{aligned} \tag{9.23}$$

Using the results in (5.8), (9.23) and Slutsky's Theorem, we have

$$\begin{aligned}
F_T &= T\left[\mathbf{f}(\widehat{\alpha}_T) - \mathbf{f}(\alpha_0)\right]'\,\|\widehat{\mathbf{v}}_T^*\|_{sd,T}^{-2}\left[\mathbf{f}(\widehat{\alpha}_T) - \mathbf{f}(\alpha_0)\right]/q \\
&= \left\{\|\mathbf{v}_T^*\|_{sd}^{-1}\sqrt{T}\left[\mathbf{f}(\widehat{\alpha}_T) - \mathbf{f}(\alpha_0)\right]\right\}'\left(\|\mathbf{v}_T^*\|_{sd}'\,\|\widehat{\mathbf{v}}_T^*\|_{sd,T}^{-2}\,\|\mathbf{v}_T^*\|_{sd}\right) \\
&\times \left\{\|\mathbf{v}_T^*\|_{sd}^{-1}\sqrt{T}\left[\mathbf{f}(\widehat{\alpha}_T) - \mathbf{f}(\alpha_0)\right]\right\}/q \\
&\to_d \mathbf{W}(1)'\left\{\frac{1}{M}\sum_{m=1}^{M}\left[\int_0^1 \phi_m(\tau)\,d\mathbf{W}(\tau)\right]\left[\int_0^1 \phi_m(\tau)\,d\mathbf{W}(\tau)\right]'\right\}^{-1}\mathbf{W}(1)/q.
\end{aligned} \tag{9.24}$$

Since $\phi_m(\tau)$, $m = 1, 2, ..., M$ are orthonormal and integrate to zero,

$$\mathbf{W}(1)'\left\{\frac{1}{M}\sum_{m=1}^{M}\left[\int_0^1 \phi_m(\tau)\,d\mathbf{W}(\tau)\right]\left[\int_0^1 \phi_m(\tau)\,d\mathbf{W}(\tau)\right]'\right\}^{-1}\mathbf{W}(1) =^d \zeta_0'\left(\frac{1}{M}\sum_{m=1}^{M}\zeta_m\zeta_m'\right)^{-1}\zeta_0$$

where $\zeta_j \sim i.i.d\ N(0, I_q)$ for $j = 0, ..., q$. This is exactly the same distribution as Hotelling (1931)'s T^2 distribution. Using the well-known relationship between the T^2 distribution and F distribution, we have

$$\frac{M-q+1}{M}F_T \to_d F_{q,M-q+1}$$

as desired. ■

Proof of Proposition 6.1. We first find the empirical Riesz representor

$$\widehat{v}_T^* = \left(\widehat{\theta}_T^{*\prime}, P_{k_T}(\cdot)'\,\widehat{\beta}_T^*\right)' \tag{9.25}$$

of $f(\alpha_0)$ on the sieve space by solving

$$||\widehat{v}_T^*||_T^2 = \sup_{v\in\mathcal{V}_T, v\neq 0} \frac{|\frac{\partial f(\widehat{\alpha}_T)}{\partial\alpha}[v]|^2}{-\frac{1}{T}\sum_{t=1}^T\{r(Z_t,\widehat{\alpha}_T)[v,v]\}}.$$

By definition

$$-\frac{1}{T}\sum_{t=1}^T\{r(Z_t,\widehat{\alpha}_T)[v,v]\} = -\gamma'\frac{1}{T}\sum_{t=1}^T\frac{\partial^2\tilde{\ell}(Z_t,\widehat{\gamma}_T)}{\partial\gamma\partial\gamma'}\gamma = \gamma'\widehat{R}_T\gamma.$$

Denote $\widehat{F}_{k_T} \equiv \left(\frac{\partial f(\widehat{\alpha}_T)}{\partial\theta'}, \frac{\partial f(\widehat{\alpha}_T)}{\partial h}[P_{k_T}(\cdot)']\right)'$. Then:

$$||\widehat{v}_T^*||_T^2 = \sup_{\gamma,\gamma\neq\gamma_{0,T}} \frac{\gamma'\widehat{F}_{k_T}\widehat{F}_{k_T}'\gamma}{-\gamma'\left(\frac{1}{T}\sum_{t=1}^T\frac{\partial^2\tilde{\ell}(Z_t,\widehat{\gamma}_T)}{\partial\gamma\partial\gamma'}\right)\gamma}.$$

The sup is achieved at

$$\widehat{\gamma}_T^* = \left(\widehat{\theta}_T^{*\prime}, \widehat{\beta}_T^{*\prime}\right)' = -\left[\frac{1}{T}\sum_{t=1}^T\frac{\partial^2\tilde{\ell}(Z_t,\widehat{\gamma}_T)}{\partial\gamma\partial\gamma'}\right]^{-1}\widehat{F}_{k_T} = \widehat{R}_T^{-1}\widehat{F}_{k_T}.$$

Substituting this into (9.25) gives us an alternative representation of $\widehat{v}_T^*$. Using this representation and (6.1), we can rewrite $\widehat{\Lambda}_m$ defined in (5.5) as

$$\begin{aligned}\widehat{\Lambda}_m &= \frac{1}{\sqrt{T}}\sum_{t=1}^T\phi_m(\frac{t}{T})\Delta(Z_t,\widehat{\alpha}_T)[\widehat{v}_T^*] = \frac{1}{\sqrt{T}}\sum_{t=1}^T\phi_m(\frac{t}{T})\frac{\partial\tilde{\ell}(Z_t,\widehat{\gamma}_T)}{\partial\gamma'}\widehat{\gamma}_T^* \\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^T\phi_m(\frac{t}{T})\frac{\partial\tilde{\ell}(Z_t,\widehat{\gamma}_T)}{\partial\gamma'}\widehat{R}_T^{-1}\widehat{F}_{k_T} = \widehat{F}_{k_T}'\widehat{R}_T^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^T\phi_m(\frac{t}{T})\frac{\partial\tilde{\ell}(Z_t,\widehat{\gamma}_T)}{\partial\gamma},\end{aligned}$$

and also rewrite $\|\widehat{v}_T^*\|_{sd,T}^2$ defined in (5.6) as

$$\begin{aligned}\|\widehat{v}_T^*\|_{sd,T}^2 &\equiv \frac{1}{M}\sum_{m=1}^M\widehat{\Lambda}_m\widehat{\Lambda}_m' \\ &= \widehat{F}_{k_T}'\widehat{R}_T^{-1}\frac{1}{M}\sum_{m=1}^M\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^T\phi_m(\frac{t}{T})\frac{\partial\tilde{\ell}(Z_t,\widehat{\gamma}_T)}{\partial\gamma}\right\}\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^T\phi_m(\frac{t}{T})\frac{\partial\tilde{\ell}(Z_t,\widehat{\gamma}_T)}{\partial\gamma'}\right\}\widehat{R}_T^{-1}\widehat{F}_{k_T} \\ &= \widehat{F}_{k_T}'\widehat{R}_T^{-1}\widehat{B}_T\widehat{R}_T^{-1}\widehat{F}_{k_T} = \widehat{Var}(\widehat{f}),\end{aligned}$$

which concludes the proof. ■

10 Appendix B: Delta Functions of Sieve M Estimation of Evaluation Functionals

Let $\alpha_0 = h_0(\cdot) \in \mathcal{H}$ be the unique maximizer of the criterion $E[\ell(Z_t, h(\cdot))]$ on $\mathcal{H}$. By the results in Subsection 3.3, the sieve Riesz representor of any linear functional $\frac{\partial f(h_0)}{\partial h}[v]$ takes the form $v_T^* = P_{k_T}{}'\beta_T^* = P_{k_T}{}'R_{k_T}^{-1}\frac{\partial f(h_0)}{\partial h}[P_{k_T}(\cdot)] \in \mathcal{V}_T$.

10.1 Delta functions for nonparametric regression type models

Let $h_0(\cdot) \in \mathcal{H}$ be the unique maximizer of the regression type criterion $E[\ell(Z_t, h(X_t))]$ on $\mathcal{H}$. For the evaluation functional $f(h_0) = h_0(\bar{x})$ with $\bar{x} \in \mathcal{X}$, we have $v_T^*(\cdot) = P_{k_T}(\cdot)'\beta_T^* = P_{k_T}(\cdot)' R_{k_T}^{-1} P_{k_T}(\bar{x})$. We also have, for any $v \in \mathcal{V}_T$:

$$v(\bar{x}) = E\left\{E[-\widetilde{r}(Z, h_0(X))|X] v(X) v_T^*(X)\right\} \equiv \int_{x\in\mathcal{X}} v(x)\,\delta_T(\bar{x}, x)\,dx$$

where

$$\begin{aligned}\delta_T(\bar{x}, x) &= E[-\widetilde{r}(Z, h_0(X))|X = x] v_T^*(x) f_X(x) \\ &= E[-\widetilde{r}(Z, h_0(X))|X = x] P_{k_T}'(\bar{x}) R_{k_T}^{-1} P_{k_T}(x) f_X(x).\end{aligned}$$

Thus $\delta_T(\bar{x}, x)$ has the reproducing property on $\mathcal{V}_T$: $v(\bar{x}) = \int_{\mathcal{X}} \delta_T(\bar{x}, x)\, v(x)\, dx$. In particular, if $1 \in \mathcal{V}_T$ then $\int_{\mathcal{X}} \delta_T(\bar{x}, x)\, dx = 1$. So $\delta_T(\bar{x}, x)$ behaves like the Dirac delta function $\delta(x - \bar{x})$.

To illustrate the properties of $\delta_T(\bar{x}, x)$, we consider the sieve (or series) LS regression of the nonparametric conditional mean model: $Y = h_0(X) + u_t$ with $E[u_t|X_t] = 0$. Then $-\widetilde{r}(Z, h_0(X)) = 1$ and

$$\delta_T(\bar{x}, x) = P_{k_T}'(x)\left\{E\left(P_{k_T} P_{k_T}'\right)\right\}^{-1} P_{k_T}(\bar{x}) f_X(x).$$

We compute and graph $\delta_T(\bar{x}, x)$ explicitly for three different sieve basis functions:

Example 1: X is uniform on $[0, 1]$ and $p_j(x) = \sqrt{2}\sin(j - 1/2)\pi x$, $j = 1, ..., k_T$. Some algebra shows that

$$\delta_T(\bar{x}, x) = \sum_{j=1}^{k_T} p_j(x)\, p_j(\bar{x}) = \frac{\sin k_T \pi (x - \bar{x})}{2\sin\frac{\pi}{2}(x - \bar{x})} - \frac{\sin k_T \pi (x + \bar{x})}{2\sin\frac{\pi}{2}(x + \bar{x})}.$$

The first part is the familiar Dirichlet kernel, which converges weakly to the Dirac delta function $\delta(x - \bar{x})$. By the Riemann-Lebesgue lemma, the $L^2([0, 1])$ inner product of the second part with any function in $L^1([0, 1]) \cap L^2([0, 1])$ converges to zero as $k_T \to \infty$. So $\delta_T(\bar{x}, x)$ does converge weakly to $\delta(x - \bar{x})$ in the sense that $\lim_{T\to\infty} \int_0^1 \delta_T(\bar{x}, x)\, v(x)\, dx = v(\bar{x}) := \int_0^1 \delta(x - \bar{x})\, v(x)\, dx$ for any $v \in L^1([0, 1]) \cap L^2([0, 1])$. It is also easy to see that

$$\delta_T(\bar{x}, \bar{x}) = k_T(1 + o(1)) \text{ and } \delta_T(\bar{x}, x) = O(1) \text{ when } x \neq \bar{x}, \tag{10.1}$$

Figure 10.1 displays the graph of $\delta_T(\bar{x}, x)/\sqrt{k_T}$, a scaled version of $\delta_T(\bar{x}, x)$ when $\bar{x} = 0.5$, the center of the distribution. The figure supports our asymptotic result in (10.1) and our qualitative observation that $\delta_T(\bar{x}, x)$ approaches the Dirac delta function as k_T increases.

Example 2: X is uniform on $[0, 1]$ and $P_{k_T}(x)$ consists of cubic B-splines with k_T evenly spaced knots. Using the property that the B-splines have compact supports, we can show that

$$\delta_T(\bar{x}, x) = \begin{cases} O(k_T) & \text{if } |x - \bar{x}| \leq C/k_T \\ o(1), & \text{if } |x - \bar{x}| > C/k_T \end{cases} \tag{10.2}$$

for some constant C. Figure 10.2 displays $\delta_T(\bar{x}, x)/\sqrt{k_T}$ for this case. As before, both the asymptotic result in (10.2) and graphical illustration show that $\delta_T(\bar{x}, x)$ collapses at $\bar{x}$ as k_T increases.

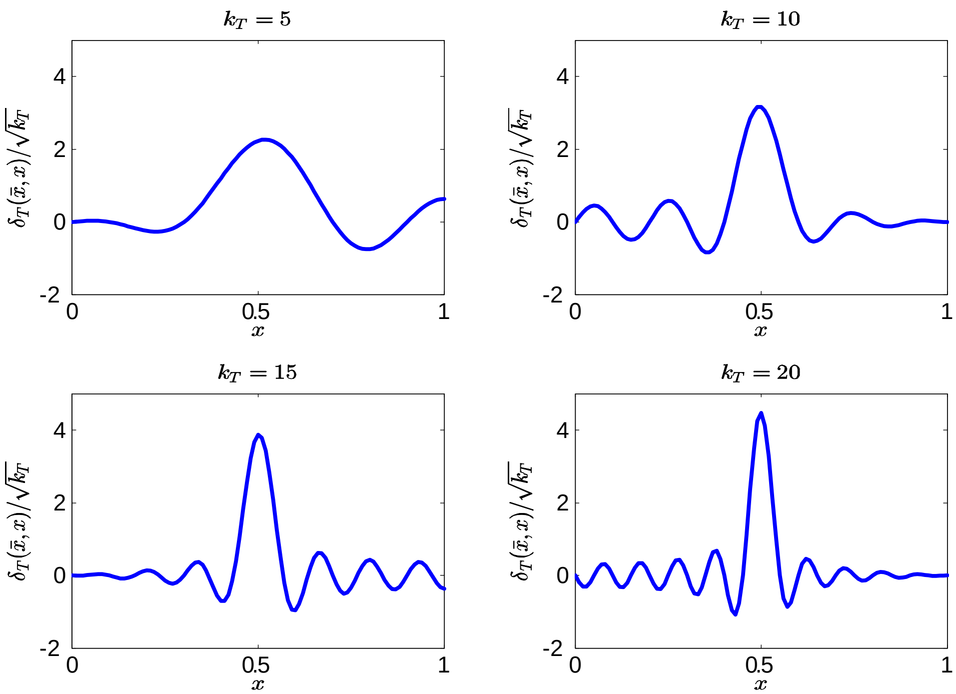

Figure 10.1: Graph of $\delta_T\left(\bar{x}, x\right)/\sqrt{k_T}$ when $\bar{x} = 0.5$, $p_j\left(x\right) = \sqrt{2}\sin\left(j - \frac{1}{2}\right)\pi x$ for different values of k_T

Example 3: X is uniform on $[-1, 1]$ and $P_{k_T}(x)$ consists of orthonormal Legendre polynomials. Then:

$$\begin{aligned}
\delta_T\left(\bar{x}, x\right) &= P'_{k_T}(x)P_{k_T}(\bar{x}) = \sum_{j=0}^{k_T} p_j\left(\bar{x}\right)p_j\left(x\right) \\
&= \frac{\left(k_T + 1\right)}{\sqrt{\left(2k_T + 1\right)\left(2k_T + 3\right)}} \frac{p_{k_T+1}\left(\bar{x}\right)p_{k_T}\left(x\right) - p_{k_T+1}\left(x\right)p_{k_T}\left(\bar{x}\right)}{\left(\bar{x} - x\right)},
\end{aligned}$$

where the last line follows from the Christoffel-Darboux formula (c.f. Szegö, 1975). Based on this result, Lebedev and Silverman (1972, sec 4.7, Theorem 1) has shown that $\int_{-1}^{1}\delta_T\left(\bar{x}, x\right)v\left(x\right)dx \rightarrow \left[\lim_{x\rightarrow\bar{x}+} v\left(x\right) + \lim_{x\rightarrow\bar{x}-} v\left(x\right)\right]/2$ for any piecewise smooth function v such that $\int_{-1}^{1} v^2\left(x\right)dx < \infty$. This is entirely analogous to the result for Fourier series expansions. Figure 10.3 graphs $\delta_T\left(\bar{x}, x\right)/\sqrt{k_T}$ for $\bar{x} = 0$. Again, as k_T increases, $\delta_T\left(\bar{x}, x\right)$ clearly becomes more concentrated at $\bar{x}$.

10.2 Delta functions for nonparametric likelihood models

To shed further light on the hidden delta sequence, we consider sieve ML estimation of a probability density function $f_X\left(\cdot\right)$ on the real line. Let $\{X_t\}_{t=1}^{T}$ be a strictly stationary weakly dependent sample with marginal density $f_X\left(\cdot\right)$, we estimate $h_0\left(\cdot\right) = \sqrt{f_X\left(\cdot\right)}$ by the sieve MLE $\widehat{h}_T$ that solves:

$$\max_{h\in\mathcal{H}_T} \frac{1}{T}\sum_{t=1}^{T}\ell(Z_t, h), \quad \text{with} \quad \ell(Z_t, h) = \frac{1}{2}\left\{\log h^2\left(X_t\right) - \log\left[\int_{-\infty}^{\infty} h^2(x)dx\right]\right\},$$

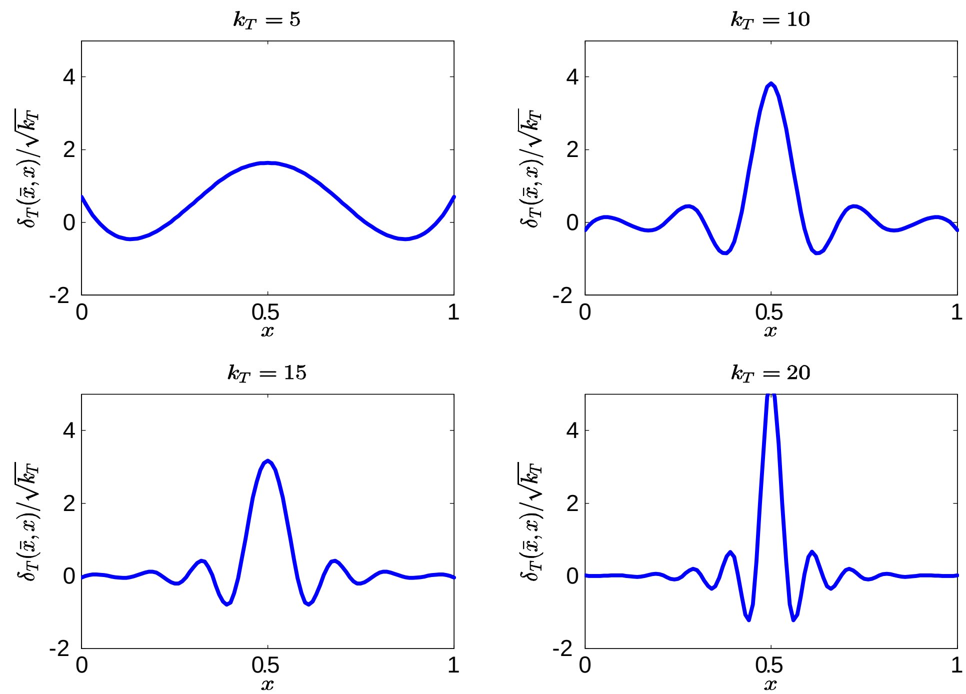

Figure 10.2: Graph of $\delta_T(\bar{x}, x)/\sqrt{k_T}$ when $\bar{x} = 0.5$ and $p_j(\cdot)$ are spline bases with k_T evenly spaced knots on [0,1]

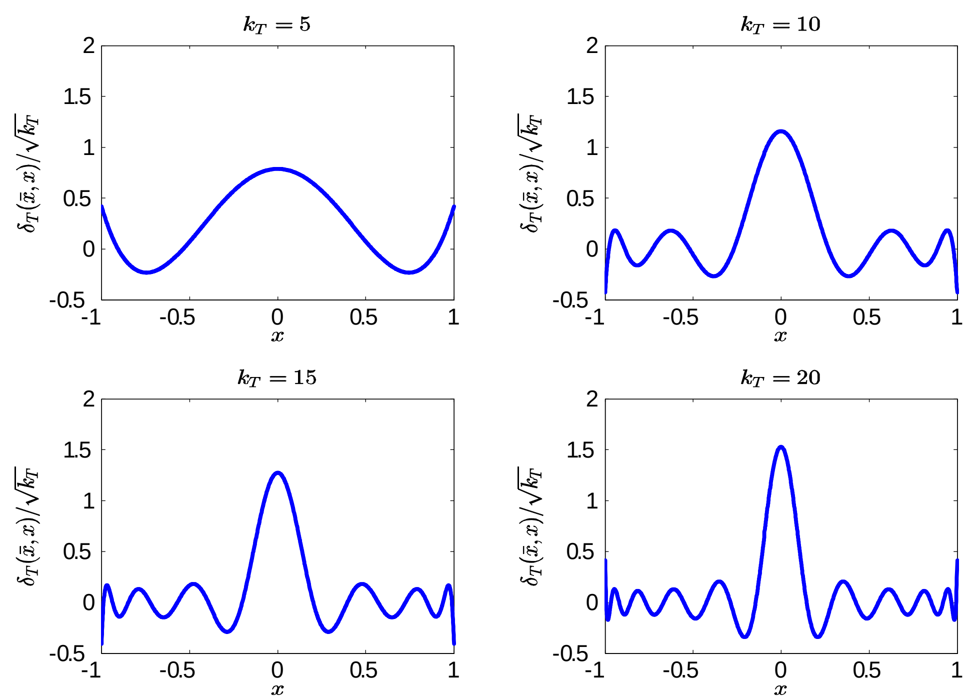

Figure 10.3: Graph of $\delta_T(\bar{x}, x)/\sqrt{k_T}$ when $\bar{x} = 0$ and $p_j(\cdot)$ are Legendre polynomials.

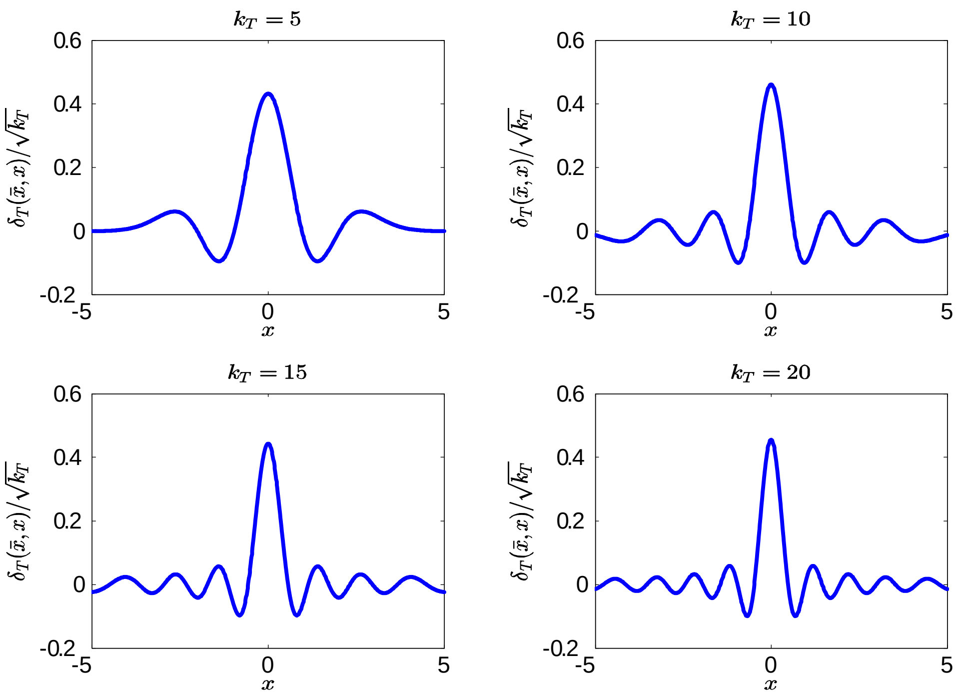

Figure 10.4: Graph of $\delta_T(\bar{x},x)/\sqrt{k_T}$ when $\bar{x}=0$ and $p_j(\cdot)$ are Hermite polynomials.

where

$$\mathcal{H}_T = \left\{ h(\cdot) = p_0(\cdot) + \sum_{j=1}^{k_T} \beta_j p_j(\cdot) : \beta_j \in \mathbb{R} \right\} \text{ with } p_j(x) = H_j(x) \exp\left(-\frac{x^2}{2}\right),$$

and $\{H_j : j = 0, 1, 2, ...\}$ is an orthonormal Hermite basis in $L^2(\mathbb{R}, \exp(-x^2))$ with the inner product $\langle h, g \rangle_w = \int_{-\infty}^{\infty} h(x) g(x) \exp(-x^2)\, dx$; see, e.g., Gallant and Tauchen (1989).

Suppose the functional of interest is $f(h_0) = h_0^2(\bar{x}) = f_X(\bar{x})$ for some $\bar{x} \in \mathbb{R}$. For any square integrable function $h(\cdot)$, define the functional $f(h) = h^2(\bar{x}) / \int_{-\infty}^{\infty} h^2(x)\, dx$. We estimate $f(h_0)$ using the plug-in sieve MLE $f(\widehat{h}_T)$. For any $v_T \in \mathcal{V}_T = \{v(\cdot) = \sum_{j=1}^{k_T} \beta_j p_j(\cdot) : \beta_j \in \mathbb{R}\}$, we have

$$\begin{aligned} \frac{\partial f(h_0)}{\partial h}[v_T] &= \frac{\partial}{\partial \tau} \left\{ \frac{[h_0(\bar{x}) + \tau v_T(\bar{x})]^2}{\int_{-\infty}^{\infty} [h_0(x) + \tau v_T(x)]^2\, dx} \right\} \Bigg|_{\tau=0} \\ &= 2[h_0(\bar{x})]\, v_T(\bar{x}) - 2[h_0(\bar{x})]^2 E\left[\frac{v_T(X)}{h_0(X)}\right] \\ &= 2h_0(\bar{x}) \left(P_{k_T}(\bar{x}) - h_0(\bar{x}) \int_{-\infty}^{\infty} h_0(x) P_{k_T}(x) dx \right)' \beta. \end{aligned}$$

It is not hard to show that

$$\begin{aligned}\|v_T\|^2 &= 2\int_{-\infty}^{\infty} [v_T(a)]\, v_T(a)\, da - 2\left(\int_{-\infty}^{\infty} [h_0(a)]\, v_T(a)\, da\right)^2 \\ &= 2\left[E\left(\frac{v_T(X)}{h_0(X)}\right)^2 - \left(E\frac{v_T(X)}{h_0(X)}\right)^2\right]\end{aligned}$$

with the corresponding inner product

$$\langle g_1, g_2\rangle = 2cov\left(\frac{g_1}{h_0}, \frac{g_2}{h_0}\right).$$

Using the results in Subsection 3.3, we can show that the sieve Riesz representor of $\frac{\partial f(h_0)}{\partial h}[v_T]$ is $v_T^*(x) = v_{T1}^*(x) - v_{T2}^*(x)$ with $v_{T1}^*(x) = P_{k_T}(x)' \beta_{T1}^*$ and $v_{T2}^*(x) = P_{k_T}(x)' \beta_{T2}^*$ where

$$\begin{aligned}\beta_{T1}^* &= \left[cov\left(\frac{P_{k_T}(X)}{h_0(X)}, \frac{P_{k_T}(X)'}{h_0(X)}\right)\right]^{-1} h_0(\bar{x})\, P_{k_T}(\bar{x}), \\ \beta_{T2}^* &= \left[cov\left(\frac{P_{k_T}(X)}{h_0(X)}, \frac{P_{k_T}(X)'}{h_0(X)}\right)\right]^{-1} [h_0(\bar{x})]^2\, E\left[\frac{P_{k_T}(X)}{h_0(X)}\right].\end{aligned}$$

Define two linear functionals $\frac{\partial f_1(h_0)}{\partial h}[v_T] = 2h_0(\bar{x})\, v_T(\bar{x})$ and $\frac{\partial f_2(h_0)}{\partial h}[v_T] = 2h_0^2(\bar{x}) \int_{-\infty}^{\infty} v_T(x)\, h_0(x)\, dx$. Then $\frac{\partial f(h_0)}{\partial h}[v_T]$ is the difference of these two functionals, and $v_{T1}^*(x)$ and $v_{T2}^*(x)$ are their respective sieve Riesz representors. While the first functional $\frac{\partial f_1(h_0)}{\partial h}[\cdot]$ is an evaluation functional and hence is irregular, the second functional $\frac{\partial f_2(h_0)}{\partial h}[\cdot]$ is a weighted integration functional with a square integrable weight function and hence is regular. Since the regular functional is $\sqrt{T}$ estimable and the irregular functional is slower than $\sqrt{T}$ estimable, the asymptotic variance of the plug-in sieve estimator $f(\hat{h}_T)$ is determined by the irregular functional $\frac{\partial f_1(h_0)}{\partial h}[\cdot]$. So for the purpose of the asymptotic variance calculation, we can focus on $\frac{\partial f_1(h_0)}{\partial h}[\cdot]$ from now on.

By definition $2h_0(\bar{x})\, v_T(\bar{x}) = \frac{\partial f_1(h_0)}{\partial h}[v_T] = \langle v_T, v_{T1}^*\rangle$. It follows by direction calculations that

$$\begin{aligned}v_T(\bar{x}) &= \frac{1}{2}\left\langle v_T(x), \frac{v_{T1}^*(x)}{h_0(\bar{x})}\right\rangle \\ &= \int_{-\infty}^{\infty} v_T(x)\, \frac{\left[v_{T1}^*(x) - h_0(x)\left(\int_{-\infty}^{\infty} v_{T1}^*(a)\, h_0(a)\, da\right)\right]}{h_0(\bar{x})} dx \\ &\equiv \int_{-\infty}^{\infty} v_T(x)\, \tilde{\delta}_T(\bar{x}, x)\, dx,\end{aligned}$$

where

$$\tilde{\delta}_T(\bar{x}, x) = \frac{v_{T1}^*(x)}{h_0(\bar{x})} - \frac{h_0(x)\left(\int_{-\infty}^{\infty} v_{T1}^*(a)\, h_0(a)\, da\right)}{h_0(\bar{x})} = \frac{v_{T1}^*(x)}{h_0(\bar{x})} + O(1).$$

Here we have used the square integrability of $h_0(\cdot)$ and so $\int_{-\infty}^{\infty} v_{T1}^*(a)\, h_0(a)\, da = O(1)$.

Using the orthonormality of $H_j(x)$ with respect to the weighting function $\exp(-x^2)$ and the matrix inverse formula:

$$\left(I_{k_T} + bb'\right)^{-1} = I_{k_T} - \frac{bb'}{1 + b'b} \text{ for any vector } b \in \mathbb{R}^{k_T},$$

we have

$$\begin{aligned}\frac{v_{T1}^{*}(x)}{h_0(\bar{x})} &= P_{k_T}(x)'\left\{E\left(\frac{P_{k_T}(X)}{h_0(X)}\frac{P'_{k_T}(X)}{h_0(X)}\right)-\left[E\frac{P_{k_T}(X)}{h_0(X)}\right]\left[E\frac{P'_{k_T}(X)}{h_0(X)}\right]\right\}^{-1}P_{k_T}(\bar{x}) \\ &= P_{k_T}(x)'\left\{I_{k_T}-\left[E\frac{P_{k_T}(X)}{h_0(X)}\right]\left[E\frac{P'_{k_T}(X)}{h_0(X)}\right]\right\}^{-1}P_{k_T}(\bar{x}) \\ &= P_{k_T}(x)'P_{k_T}(\bar{x})-\frac{P_{k_T}(x)'\left[E\frac{P_{k_T}(X)}{h_0(X)}\right]\left[E\frac{P'_{k_T}(X)}{h_0(X)}\right]P_{k_T}(\bar{x})}{1+\left[E\frac{P'_{k_T}(X)}{h_0(X)}\right]\left[E\frac{P_{k_T}(X)}{h_0(X)}\right]} \\ &= P_{k_T}(x)'P_{k_T}(\bar{x})+O(1)\end{aligned}$$

using the square integrability of $h_0(\cdot)$.

Let $\delta_T(\bar{x},x)=P_{k_T}(x)'P_{k_T}(\bar{x})+p_0(x)p_0(\bar{x})$, then $\tilde{\delta}_T(\bar{x},x)=\delta_T(\bar{x},x)+O(1)$. But

$$\delta_T(\bar{x},x)=\sum_{j=0}^{k_T}H_j(\bar{x})H_j(x)\exp\left[-\left(\frac{x^2}{2}+\frac{\bar{x}^2}{2}\right)\right].$$

It is known that $\sum_{j=0}^{\infty}H_j(\bar{x})H_j(x)\exp\left[-\left(\frac{x^2}{2}+\frac{\bar{x}^2}{2}\right)\right]=\delta(x-\bar{x})$ in the sense of distributions. This follows by letting $u\to1$ in Mehler's formula, valid for $u\in(-1,1)$:

$$\sum_{j=0}^{\infty}u^jH_j(\bar{x})H_j(x)\exp\left[-\left(\frac{x^2}{2}+\frac{\bar{x}^2}{2}\right)\right]=\frac{1}{\sqrt{\pi(1-u^2)}}\exp\left(-\frac{1-u}{1+u}\frac{(x+\bar{x})^2}{4}-\frac{1+u}{1-u}\frac{(x-\bar{x})^2}{4}\right).$$

See Lebedev and Silverman (1972, Sec 4.11). Figure 10.4 demonstrates the convergence of $\delta_T(\bar{x},x)$ to the delta function.

It is important to point out that in all the examples the asymptotic behavior $\delta_T(\bar{x},x)$ remains more or less the same for other nonboundary values $\bar{x}\in\mathcal{X}$. So implicitly in the method of sieves, there are delta sequences, i.e. sequences of functions that converge to the delta distribution. The Dirichlet and Fejer kernels of Fourier series and the Christoffel-Darboux and Mehler formulae for orthogonal polynomials are examples of these delta sequences. When the sample size is large, the method of sieves effectively entails taking a weighted average of observations with the weights given by a delta sequence.

References

[1] Ai, C., and X. Chen (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795–1843.

[2] Andrews, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–854.

[3] Chen, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," In: James J. Heckman and Edward E. Leamer, Editor(s), *Handbook of Econometrics*, Volume 6, Part 2, Pages 5549–5632.

[4] Chen, X. (2011): "Penalized Sieve Estimation and Inference of Semi-nonparametric Dynamic Models: A Selective Review," Cowles Foundation Discussion Paper 1804. Yale University.

[5] Chen, X. and Z. Liao (2008): "On Limiting Distributions of Sieve M Estimators of Irregular Functionals," working paper, Yale and UCLA.

[6] Chen, X. and X. Shen (1998): "Sieve Extremum Estimates for Weakly Dependent Data," *Econometrica*, 66, 289–314.

[7] Chen, X. , W. Wu and Y. Yi (2009): "Efficient Estimation of Copula-based Semiparametric Markov Models," *Annals of Statistics*, Vol. 37(6B), 4214–4253.

[8] Conley, T.G., L. P. Hansen, W. Liu (1997): "Bootstrapping the Long Run," *Macroeconomic Dynamics*, 1, 1997, 279–311.

[9] Fan, J. and Q. Yao (2003): *Nonlinear Time Series: Nonparametric and Parametric Methods*, New York: Springer.

[10] Gallant, A.R. and G. Tauchen (1989): "Semiparametric Estimation of Conditional Constrained Heterogenous Processes: Asset Pricing Applications," *Econometrica*, 57, 1091–1120.

[11] Gao, J. (2007): *Nonlinear Time Series: Semiparametric and Nonparametric Methods.* London: Chapman & Hall/CRC.

[12] Granger, C.W.J. (2003): "Time Series Concepts for Conditional Distributions," *Oxford Bulletin of Economics and Statistics*, 65, supplement 689–701.

[13] Grenander, U. (1981): *Abstract Inference*, New York: Wiley Series.

[14] Hall, P. and C. C. Heyde (1980): *Martingale Limit Theory and Its Applications*, New York: Academic Press.

[15] Hansen, L.P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

[16] Härdle, W., H. Liang and J. Gao (2000): *Partially Linear Models.* Springer.

[17] Hotelling H. (1931): "The Generalization of Student's Ratio," *Annals of Mathematical Statistics*, 2, 360-378.

[18] Jansson, M. (2004): "The Error in Rejection Probability of Simple Autocorrelation Robust Tests," *Econometrica*, 72, 937–946.

[19] Kiefer, N.M. and T.J. Vogelsang (2005): "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests," *Econometric Theory*, 21, 1130-1164.

[20] Lebedev, N.N. and R.A. Silverman (1972): *Special Functions and Their Applications*. Revised edition, Dover Publications.

[21] Li, Q. and J. Racine (2007): *Nonparametric Econometrics Theory and Practice*. Princeton: Princeton University Press.

[22] Lu, Z. and O. Linton (2007) "Local linear fitting under near epoch dependence," *Econometric Theory*, 23, 37–70.

[23] Masry, E. and J. Fan (1997): "Local Polynomial Estimation of Regression Functions for Mixing Processes," *Scandinavian Journal of Statistics*, 24, 165–179.

[24] Müller, U.K. (2007): "A Theory of Robust Long-Run Variance Estimation," *Journal of Econometrics*, 141(2), 1331–1352.

[25] Newey, W. K. and K. D. West (1987): "A Simple, Positive Semidefinite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

[26] Newey, W.K. (1997): "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics*, 79, 147–168.

[27] Pritsker, M. (1998): "Nonparametric Density Estimation and Tests of Continuous Time Interest Rate Models," *Review of Financial Studies*, 11 449–487.

[28] Phillips, P. C. B. (2005): "HAC Estimation by Automated Regression," *Econometric Theory*, 21, 116–142.

[29] Robinson, P. (1983): "Nonparametric Estimators for Time Series," *Journal of Time Series Analysis*, 4, 185–207.

[30] Sun, Y. (2011a): "Robust Trend Inference with Series Variance Estimator and Testing-optimal Smoothing Parameter," *Journal of Econometrics*, 164(2), 345–366.

[31] Sun, Y. (2011b): "Asymptotic F Test in a GMM Framework," Working paper, Department of Economics, UC San Diego.

[32] Szegö, G. (1975): *Orthogonal Polynomials*, 4th edition. Providence, RI: American Mathematical Society, 42–44.

[33] White, H. (2004): *Asymptotic Theory for Econometricians*. Academic Press.

[34] White, H. and J. Wooldridge (1991): "Some Results on Sieve Estimation with Dependent Observations," in Barnett, W.A., J. Powell and G. Tauchen (eds.), *Non-parametric and Semi-parametric Methods in Econometrics and Statistics*, 459–493, Cambridge: Cambridge University Press.