SIEVE INFERENCE ON POSSIBLY MISSPECIFIED SEMI-NONPARAMETRIC TIME SERIES MODELS

By Xiaohong Chen* , Zhipeng Liao and Yixiao Sun†

Yale University, UC Los Angeles and UC San Diego

This paper provides a general theory on the asymptotic normality of plug-in sieve M estimators of possibly irregular functionals of semi-nonparametric time series models. We show that, even when the sieve score process is not a martingale difference, the asymptotic variances of plug-in sieve M estimators of irregular (i.e., slower than root-T estimable) functionals are the same as those for independent data. Nevertheless, ignoring the temporal dependence in finite 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 of unknown forms. 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 conventional Wald test with chi-square critical values.

1. Introduction. Many economic and financial time series are nonlinear and non-Gaussian; see, e.g., Granger (2003). For policy analysis, it is important to uncover complicated nonlinear economic relations in structural models. Unfortunately, it is difficult to correctly parameterize all aspects of nonlinear dynamic functional relations. Due to the well-known problem of "curse of dimensionality" it is also impractical to estimate a general nonlinear time series model fully nonparametrically. These issues motivate the growing popularity of semiparametric and semi-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 statistics, economics, finance, biostatistics and other disciplines. In this paper, we focus on sieve M estimation, which optimizes a sample average of a random criterion 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 Shen and Wong (1994), Chen (2007) and the references therein for many examples of sieve

*Supported by the National Science Foundation SES-0838161 and Cowles Foundation

†Supported by the National Science Foundation SES-0752443

AMS 2000 subject classifications: Primary 62G10; secondary 62M10

Keywords and phrases: Dynamic misspecification, sieve M estimation, sieve Riesz representer, irregular functional, pre-asymptotic variance, orthogonal series long run variance estimation, F distribution

M estimation, including sieve (quasi) maximum likelihood, sieve (nonlinear) least squares, sieve generalized least squares, and sieve quantile regression.

We consider inference on possibly misspecified semi-nonparametric time series models via the method of sieve M estimation. 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. The key insight is to examine the functional of interest on a sieve tangent space where a Riesz representer always exists regardless of whether the functional is regular or irregular. The asymptotic normality result is rate-adaptive in the sense that applied researchers do not need to know *a priori* whether the functional of interest is $\sqrt{T}$ estimable or not.

For possibly misspecified semi-nonparametric models with weakly dependent data, Chen and Shen (1998) establish that the asymptotic variance of a sieve M estimator of any regular functional depends on the temporal dependence and is equal to the long run variance (LRV) of a scaled score (or moment) process. In this paper, we show a new result that, regardless of whether the score process is martingale difference or not, the asymptotic variance of a sieve M estimator of an irregular functional for weakly dependent data is the same as that for independent data.

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. However, simulation studies indicate that inference procedures based on asymptotic variance estimates ignoring autocorrelation do 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.

To deal with this problem, for inference on both regular and irregular functionals, we propose to use a "pre-asymptotic" sieve variance that captures temporal dependence of an unknown form. 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 novel "pre-asymptotic" sieve approach enables us to develop a unified inference framework that can accommodate both regular and irregular functionals.

To derive a simple and more accurate asymptotic approximation under weak conditions, we compute a "pre-asymptotic" Wald statistic using an orthonormal series LRV (OS-LRV) estimator. For both regular and irregular functionals, we show that the "pre-asymptotic" t statistic and a scaled 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 of unknown form or not; and (3) whether the series number of terms in the OS-LRV estimator is held fixed or not.

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 that the asymptotic variances of plug-in sieve M estimators of irregular functionals for weakly dependent 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 describes a simple computation method and reports a simulation study using a partially linear regression model. Appendix contains all the proofs.

Notation. We denote $f_A(a)$ $(F_A(a))$ as the marginal probability density (cdf) of 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 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. For simplicity, 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. 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 $\mathcal{Z} \subseteq \mathbb{R}^{d_z}, 1 \leq d_z < \infty$, $\mathcal{Y} \subseteq \mathbb{R}^{d_y}$ and $\mathcal{X} \subseteq \mathbb{R}^{d_x}$ be the supports of Z_t, Y_t and X_t 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 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:

$$d(\widehat{\alpha}_T,\alpha_0) = O_p\left\{\max\left[d(\widehat{\alpha}_T,\Pi_T\alpha_0), d\left(\Pi_T\alpha_0,\alpha_0\right)\right]\right\} = o_p(1).$$

Given the consistency, 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 $c>0$), 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: $\|\widehat{\alpha}_T-\alpha_0\| = O_p(\xi_T) = o_p(T^{-1/4})$, where $\xi_T = \max\left[\|\widehat{\alpha}_T-\Pi_T\alpha_0\|, \|\Pi_T\alpha_0-\alpha_0\|\right]$.

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., Shen and Wong (1994) and Chen (2007) for additional examples.

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

$$Y_t = X_t'\theta_0 + h_{01}\left(Y_{t-1}\right) + h_{02}\left(Y_{t-2}\right) + u_t, \tag{2.2}$$

with

$$E\left[u_t|X_t,Y_{t-1},Y_{t-2}\right] = 0,$$

where X_t *is a* d_x-*dimensional random vector, and could include finitely many lagged* Y_t*'s. Let* $\theta_0\in\Theta\subset\mathbb{R}^{d_x}$ *and* $h_{0j}\in\mathcal{H}_j$ *for* $j=1,2$*. Let* $\alpha_0 = (\theta_0',h_{01},h_{02})' \in \mathcal{A} = \Theta\times\mathcal{H}_1\times\mathcal{H}_2$*. Examples of functionals of interest could be* $f(\alpha_0) = \lambda'\theta_0$ *or* $\nabla h_{0j}(\overline{y}_j)$ *where* $\lambda\in\mathbb{R}^{d_x}$ *and* $\overline{y}_j \in int(\mathcal{Y})$ *for* $j=1,2$*.*

For the sake of concreteness we assume that $\mathcal{Y}$ is a bounded interval of $\mathbb{R}$ and $\mathcal{H}_j = \Lambda^{s_j}(\mathcal{Y})$ (a Hölder space) for $s_j > 0.5$, $j=1,2$, where

$$\Lambda^s(\mathcal{Y}) = \left\{h\in C^{[s]}(\mathcal{Y}) : \sup_{k\leq[s]}\sup_{y\in\mathcal{Y}}\left|\nabla^k h(y)\right| < \infty, \sup_{y,y'\in\mathcal{Y}}\frac{\left|\nabla^{[s]}h(y)-\nabla^{[s]}h\left(y'\right)\right|}{|y-y'|^{s-[s]}} < \infty\right\},$$

where $[s]$ is the largest integer that is strictly smaller than s. The Hölder space $\Lambda^s(\mathcal{Y})$ (with $s>0.5$) is a smooth function space that is widely assumed in the semi-nonparametric literature. We can then approximate $\mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2$ by a sieve $\mathcal{H}_T = \mathcal{H}_{1,T} \times \mathcal{H}_{2,T}$, where for $j = 1, 2$,

$$\mathcal{H}_{j,T} = \left\{ h(\cdot) : h(\cdot) = \sum_{k=1}^{k_{j,T}} \beta_k p_{j,k}(\cdot) = \beta' P_{k_{j,T}}(\cdot),\ \beta \in \mathbb{R}^{k_{j,T}} \right\}, \tag{2.3}$$

where the known sieve basis $P_{k_{j,T}}(\cdot)$ could be polynomial splines, B-splines, wavelets, Fourier series and others.

Let $\ell(Z_t, \alpha) = -\left[Y_t - X_t'\theta - h_1(Y_{t-1}) - h_2(Y_{t-2})\right]^2 /4$ with $\alpha = (\theta', h_1, h_2)' \in \mathcal{A} = \Theta \times \mathcal{H}_1 \times \mathcal{H}_2$. Let $\mathcal{A}_T = \Theta \times \mathcal{H}_{1,T} \times \mathcal{H}_{2,T}$ be a sieve for $\mathcal{A}$. We can estimate $\alpha_0 \in \mathcal{A}$ by the sieve least squares (LS) estimator $\widehat{\alpha}_T \equiv (\widehat{\theta}_T', \widehat{h}_{1,T}, \widehat{h}_{2,T})' \in \mathcal{A}_T$:

$$\widehat{\alpha}_T = \arg \max_{(\theta, h_1, h_2) \in \mathcal{A}_T} \frac{1}{T} \sum_{t=1}^{T} \ell(Z_t, \theta, h_1, h_2). \tag{2.4}$$

A functional of interest $f(\alpha_0)$ (such as $\lambda'\theta_0$ or $\nabla h_{0j}(\overline{y}_j)$) is then estimated by the plug-in sieve LS estimator $f(\widehat{\alpha}_T)$ (such as $\lambda'\widehat{\theta}_T$ or $\nabla \widehat{h}_{j,T}(\overline{y}_j)$).

This example is very similar to Example 2 in Chen and Shen (1998), except that we allow for dynamic mispecification in the sense that $E\left[u_t | X_t, Y_{t-1}, Y_{t-2}; Y_{t-j} \text{ for } j \geq 3\right]$ may not equal to zero. 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 that paper does not provide 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, robust 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 $Y^{t-1} = (Y_{t-1}, \ldots, Y_1)$ 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:

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

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 $Q^Y_{0.01}(\overline{y}) = F_Y^{-1}\left(C^{-1}_{2|1}[\tau|F_Y(y);\theta_0]\right)$ 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

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

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:

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

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 $Q^Y_{0.01}(\overline{y})$) 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 $\widehat{Q}^Y_{0.01}(\overline{y}) = \widehat{F}_Y^{-1}(C^{-1}_{2|1}[\tau|\widehat{F}_Y(y);\widehat{\theta}])$).

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 $Q^Y_{0.01}(\overline{y})$. 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 representer that appears in our asymptotic normality result.

3.1. *Local Geometry.* The convergence rate result of Chen and Shen (1998) implies that $\widehat{\alpha}_T \in \mathcal{B}_T \subset \mathcal{B}_0$ with probability approaching one, where

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

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

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

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), there is lots of freedom to choose such a norm $\|\alpha - \alpha_0\|$ that 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 $v = \alpha - \alpha_0 \in \mathcal{V}$:

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

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

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

We call v_T^* the *sieve Riesz representer* 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. This crucial observation enables us to develop a general and unified theory that is currently lacking in the literature.

- 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 representer of the functional $\frac{\partial f(\alpha_0)}{\partial \alpha}[\cdot]$ on $\mathcal{V}$. See, e.g., Shen (1997).
- 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) $\sup_{\alpha\in\mathcal{B}_T}\left|f(\alpha)-f(\alpha_0)-\frac{\partial f(\alpha_0)}{\partial\alpha}[\alpha-\alpha_0]\right|=o\left(T^{-\frac{1}{2}}\left\|v_T^*\right\|\right);$

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

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

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

Note that $\|v_T^*\|_{sd}^2$ is the finite dimensional sieve version of the long run variance of the score process $\Delta(Z_t,\alpha_0)[v_T^*]$, and $\|v_T^*\|_{sd}^2=Var\left(\Delta(Z,\alpha_0)[v_T^*]\right)$ if the score process $\{\Delta(Z_t,\alpha_0)[v_T^*]\}_{t\leq T}$ is a martingale difference array.

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

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

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

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

Let $\mu_T\left\{g\left(Z\right)\right\}\equiv T^{-1}\sum_{t=1}^{T}\left[g\left(Z_t\right)-Eg\left(Z_t\right)\right]$ 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* $v \in \mathcal{V}$*;*

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

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

Assumptions 3.3.(ii) and (iii) are simplified versions of those in Chen and Shen (1998), and can be verified in the same way.

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

Assumption 3.4 is a very mild one, and can be easily verified by applying any existing triangular array CLT for weakly dependent data (see, e.g., Hall and Heyde, 1980).

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*

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

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

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

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.

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

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

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

3.3. *Sieve Riesz Representer.* To apply the asymptotic normality Theorem 3.1 one needs to verify Assumptions 3.1–3.4. Once we compute the sieve Riesz representer $v_T^* \in \mathcal{V}_T$, 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 representer $v^* \in \mathcal{V}$ in a closed form for a regular functional on the infinite dimensional space $\mathcal{V}$, we can always compute the sieve Riesz representer $v_T^* \in \mathcal{V}_T$ 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 representers 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 representer $v_T^* = (v_{\theta,T}^{*\prime}, v_{h,T}^*(\cdot))' = (v_{\theta,T}^{*\prime}, P_{k_T}(\cdot)'\beta_T^*)' \in \mathcal{V}_T$ of $\frac{\partial f(\alpha_0)}{\partial \alpha}[\cdot]$ solves the following optimization problem:

$$\begin{aligned}\frac{\partial f(\alpha_0)}{\partial \alpha}[v_T^*] = \|v_T^*\|^2 &= \sup_{v=(v_\theta', v_h)' \in \mathcal{V}_T, v \neq 0} \frac{\left|\frac{\partial f(\alpha_0)}{\partial \theta'} v_\theta + \frac{\partial f(\alpha_0)}{\partial h}[v_h(\cdot)]\right|^2}{E\left(-r\left(Z_t, \theta_0, h_0(\cdot)\right)[v, v]\right)} \\ &= \sup_{\gamma=(v_\theta', \beta')' \in \mathbb{R}^{d_\theta + k_T}, \gamma \neq 0} \frac{\gamma' F_{k_T} F_{k_T}' \gamma}{\gamma' R_{k_T} \gamma},\end{aligned} \tag{3.18}$$

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,[1] 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

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

[1] 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 $-r(Z_t, \alpha_0)[\cdot, \cdot]$ in 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 $v = (v'_\theta, P_{k_T}(\cdot)'\beta)' \in \mathcal{V}_T$,

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

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

$$\gamma_T^* = \left(v_{\theta,T}^{*\prime}, \beta_T^{*\prime}\right)' = R_{k_T}^{-1} F_{k_T}. \tag{3.23}$$

The sieve Riesz representer is then given by

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

Consequently,

$$\|v_T^*\|^2 = \gamma_T^{*\prime} R_{k_T}\gamma_T^* = F'_{k_T} R_{k_T}^{-1} F_{k_T}, \tag{3.24}$$

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

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

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

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.

REMARK 3.2. *Assumption 3.5 implies that*

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

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

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 $\mathbf{0}'_{k_T} = [0, \ldots, 0]_{1\times k_T}$, and hence $v_T^* = (v_{\theta,T}^{*\prime}, P_{k_T}(\cdot)'\beta_T^*)' \in \mathcal{V}_T$ with $v_{\theta,T}^* = I_T^{11}\lambda$, $\beta_T^* = I_T^{21}\lambda$, and

$$\|v_T^*\|^2 = F'_{k_T} R_{k_T}^{-1} F_{k_T} = \lambda' I_T^{11}\lambda.$$

If the largest eigenvalue of I_T^{11}, $\lambda_{\max}(I_T^{11})$, is bounded above by a finite constant uniformly in k_T, then $\|v_T^*\|^2 \le \lambda_{\max}(I_T^{11}) \times \lambda'\lambda < \infty$ 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 $v_T^* = (v_{\theta,T}^{*\prime}, P_{k_T}(\cdot)'\beta_T^*)' \in \mathcal{V}_T$ with $v_{\theta,T}^* = I_T^{12}P_{k_T}(\overline{x})$, $\beta_T^* = I_T^{22}P_{k_T}(\overline{x})$, and

$$\|v_T^*\|^2 = F'_{k_T} R_{k_T}^{-1} F_{k_T} = P'_{k_T}(\overline{x}) I_T^{22} P_{k_T}(\overline{x}).$$

So if the smallest eigenvalue of I_T^{22}, $\lambda_{\min}(I_T^{22})$, is bounded away from zero uniformly in k_T, then $\|v_T^*\|^2 \ge \lambda_{\min}(I_T^{22})||P_{k_T}(\overline{x})||_E^2 \to \infty$, 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 $v_T^* = (v_{\theta,T}^{*\prime}, P_{k_T}(\cdot)'\beta_T^*)'$ with $v_{\theta,T}^* = I_T^{12}\int_{\mathcal{X}} w(x)P_{k_T}(x)dx$, $\beta_T^* = I_T^{22}\int_{\mathcal{X}} w(x)P_{k_T}(x)dx$, and

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

Suppose that the smallest and largest eigenvalues of I_T^{22} 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 $f(\alpha) = \int_{\mathcal{X}} w(x)h(x)dx$ is regular if $\lim_{k_T} ||\int_{\mathcal{X}} w(x)P_{k_T}(x)dx||_E^2 < \infty$; is irregular if $\lim_{k_T} ||\int_{\mathcal{X}} w(x)P_{k_T}(x)dx||_E^2 = \infty$.

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 $h_0^2 = f_Y$ 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 $f(\alpha_0) = f_Y(\overline{y}) = h_0^2(\overline{y})$ 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 $v_T^* = (v_{\theta,T}^{*\prime}, P_{k_T}(\cdot)'\beta_T^*)' \in \mathcal{V}_T$ with $v_{\theta,T}^* = I_T^{12}\frac{\partial f(\alpha_0)}{\partial h}[P_{k_T}(\cdot)]$, $\beta_T^* = I_T^{22}\frac{\partial f(\alpha_0)}{\partial h}[P_{k_T}(\cdot)]$, and

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

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

4. Asymptotic Variances 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.

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:

$$||v_T^*||_{sd}^2 = Var\left(\Delta(Z,\alpha_0)[v_T^*]\right) \times \left[1 + 2\sum_{t=1}^{T-1}\left(1-\frac{t}{T}\right)\rho_T^*(t)\right], \tag{4.1}$$

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

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

Denote

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

The following high-level assumption captures the essence of the problem.

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*

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

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

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

$$\frac{\sqrt{T}\left[f(\widehat{\alpha}_T)-f(\alpha_0)\right]}{\sqrt{Var\left(\Delta(Z,\alpha_0)[v_T^*]\right)}} \to_d N(0,1). \tag{4.3}$$

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 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 easily 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

$$\|\Delta(Z,\alpha_0)[v_T^*]\|_p \equiv \left(E\left\{|\Delta(Z,\alpha_0)[v_T^*]|^p\right\}\right)^{1/p}. \tag{4.4}$$

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 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)* $\{Z_t\}_{t=1}^{\infty}$ *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 1.(iii) in section 6.6.2 of Fan and Yao (2003) for the pointwise asymptotic normality of their local polynomial estimator of a conditional mean function. 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).

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.*

Theorem 4.1 and Proposition 4.2 show 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), Fan and Yao (2003) and Gao (2007). However, whether this is the case for general sieve M estimators of unknown functionals has been a long standing question. Theorem 4.1 and Proposition 4.2 give 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.

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 representer 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

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

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

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

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,

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

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

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:*

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

(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. Assumptions 4.6.(i) and (iii) are easily satisfied by two typical functionals of nonparametric models: the evaluation functional and the weighted integration functional.

Consider as an example 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). Furthermore, we have, for any $v_T\in\mathcal{V}_T$:

$$v_T\left(\bar{x}\right)=E\left\{E[-\widetilde{r}\left(Z,h_0\left(X\right)\right)|X]v_T(X)v_T^*(X)\right\}\equiv\int_{x\in\mathcal{X}}v_T\left(x\right)\delta_T\left(\bar{x},x\right)dx, \tag{4.5}$$

where

$$\begin{aligned}\delta_T\left(\bar{x},x\right)&=E[-\widetilde{r}\left(Z,h_0\left(X\right)\right)|X=x]v_T^*\left(x\right)f_X\left(x\right)\\&=E[-\widetilde{r}\left(Z,h_0\left(X\right)\right)|X=x]P_{k_T}'(\overline{x})R_{k_T}^{-1}P_{k_T}(x)f_X\left(x\right).\end{aligned} \tag{4.6}$$

By equation (4.5) $\delta_T\left(\bar{x},x\right)$ has the reproducing property on $\mathcal{V}_T$, so it behaves like the Dirac delta function $\delta\left(x-\bar{x}\right)$ on $\mathcal{V}_T$. Therefore $v_T^*\left(x\right)$ concentrates in a neighborhood around $x=\bar{x}$ and maintains the same positive sign in this neighborhood.

We first verify Assumption 4.6.(i). By equation (4.6), we have

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

where $sign(v_T^*\left(x\right))=1$ if $v_T^*\left(x\right)>0$ and $sign(v_T^*\left(x\right))=-1$ if $v_T^*\left(x\right)\leq 0$, and $\sup_{x\in\mathcal{X}}|b_T(x)|\leq c_1^{-1}<\infty$ under Assumption 4.5.(i). If $b_T(x)\in\mathcal{V}_T$, then by equation (4.5) we have:

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

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 (4.5), 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.

Similarly 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) is satisfied.

One may conclude from Theorem 4.1 and Proposition 4.2 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 and the dependence is weak. Whether the sample size is large enough so that one 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.

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 representer 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 representer $\widehat{v}_T^*$ 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 representer v_T^* can be consistently estimated by the empirical sieve Riesz representer $\widehat{v}_T^*$ 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. To be consistent with our focus on the method of sieves and to derive a simple and robust 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) and Sun (2011a). Let $\{\phi_m\}_{m=0}^\infty$ be a sequence of orthonormal basis functions in $L^2([0,1])$ with $\phi_0(\cdot) \equiv 1$. 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 $m = 1, 2, \ldots, M$ 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 LRV estimation.

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 sieve score process $\{\Delta(Z_t,\alpha_0)[v_T^*]\}_{t=1}^T$. To capture the possibly large but diminishing autocorrelation in a finite sample, we propose treating $\{\Delta(Z_t,\alpha_0)[v_T^*]\}_{t=1}^T$ as a generic time series and using the same formula as in (5.6) to estimate the asymptotic variance of $T^{-1/2}\sum_{t=1}^{T}\Delta(Z_t,\alpha_0)[v_T^*]$. We call the estimator the "pre-asymptotic" variance estimator. With a data-driven smoothing parameter choice of M, 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:

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

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) $\sup_{v\in\mathcal{W}_T,\alpha\in\mathcal{B}_T} T^{-1/2}\sum_{t=1}^{T}\phi_m\left(t/T\right)\left(\Delta(Z_t,\alpha)\left[v\right] - \Delta(Z_t,\alpha_0)\left[v\right] - E\{\Delta(Z_t,\alpha)\left[v\right]\}\right) = o_p(1)$ *for* $m = 0, 1, \ldots, M$;

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

(iii) $\sup_{v\in\mathcal{W}_T}\left|T^{-1/2}\sum_{t=1}^{T}\phi_m(t/T)\Delta(Z_t,\alpha_0)[v]\right| = O_p(1)$ *for* $m = 0, 1, \ldots, M$;

(iv) For $e_t \sim iid\ N(0,1)$, *we have for any* $x = (x_1, \ldots, x_M)' \in \mathbb{R}^M$,

$$\begin{aligned}
&P\left(T^{-1/2}\sum_{t=1}^{T}\phi_m(t/T)\Delta(Z_t,\alpha_0)\left[u_T^*\right] < x_m,\ m = 0, 1, \ldots, M\right)\\
&= P\left(T^{-1/2}\sum_{t=1}^{T}\phi_m(t/T)e_t < x_m,\ m = 0, 1, \ldots, M\right) + o\left(1\right).
\end{aligned}$$

Assumption 5.2.(iv) is a slightly stronger version of Assumption 3.4. It is equivalent to assuming that $T^{-1/2}\sum_{t=1}^{T}[\phi_0(t/T),\ldots,\phi_m(t/T)]'\,\Delta(Z_t,\alpha_0)\,[u_T^*]$ follows a multivariate CLT. When $\phi_m(x)$ is continuously differentiable in x, Assumption 5.2.(iv) is weaker than a FCLT of the form:

$$T^{-1/2}\sum_{t=1}^{[T\tau]}\Delta(Z_t,\alpha_0)\,[u_T^*]\to^d W(\tau)$$

where $W(\tau)$ is the standard Brownian motion process. A FCLT of the above type is often assumed in parametric time series analysis. When Assumption 5.2.(iv) holds, we write

$$T^{-1/2}\sum_{t=1}^{T}\phi_m(\frac{t}{T})\Delta(Z_t,\alpha_0)\,[u_T^*]\overset{a}{\sim}T^{-1/2}\sum_{t=1}^{T}\phi_m(\frac{t}{T})e_t$$

where $\overset{a}{\sim}$ signifies that the two sides are asymptotically equivalent in distribution.

Theorem 5.1. *Let $\{\phi_m\}_{m=0}^{M}$ be a sequence of orthonormal basis functions in $L^2([0,1])$. Under Assumptions 3.2, 3.3, 5.1 and 5.2, we have, for $m=1,\ldots,M$,*

$$\|v_T^*\|_{sd}^{-1}\,\widehat{\Lambda}_m\overset{a}{\sim}iid\ N(0,1).$$

If further Assumption 3.1 holds, then

$$t_T\equiv\sqrt{T}\,[f(\widehat{\alpha}_T)-f(\alpha_0)]\,/\,\|\widehat{v}_T^*\|_{sd,T}\overset{a}{\sim}t(M),$$

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 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 representer $\mathbf{v}_T^*=(v_{1,T}^*,\ldots,v_{q,T}^*)$ 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 $\widehat{\mathbf{\Lambda}}_m = (\widehat{\Lambda}_m^{(1)}, \ldots, \widehat{\Lambda}_m^{(q)})'$ with

$$\widehat{\Lambda}_m^{(j)} = T^{-1/2} \sum_{t=1}^{T} \phi_m(t/T) \Delta(Z_t, \widehat{\alpha}_T)[\hat{v}_{j,T}^*],$$

where $\hat{v}_{j,T}^*$ denotes the empirical sieve Riesz representer of the functional $\frac{\partial f_j(\widehat{\alpha}_T)}{\partial \alpha}[\cdot]$ ($j = 1, \ldots, q$). 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:

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

We maintain Assumption 5.2 but replace Assumption 5.2(iv) by its multivariate version: for $\mathbf{e}_t \sim iid\ N(0, I_q)$, we have

$$\begin{aligned} &P\left(T^{-1/2} \sum_{t=1}^{T} \phi_m(t/T) \Delta(Z_t, \alpha_0) \left[\|\mathbf{v}_T^*\|_{sd}^{-1} \mathbf{v}_T^*\right] < \mathbf{x}_m,\ m = 0, 1, \ldots, M\right) \\ &= P\left(T^{-1/2} \sum_{t=1}^{T} \phi_m(t/T) \mathbf{e}_t < \mathbf{x}_m,\ m = 0, 1, \ldots, M\right) + o(1) \end{aligned}$$

for $\mathbf{x}_m \in \mathbb{R}^q$.

Using a proof similar to that for Theorem 5.1, we can prove the theorem below.

THEOREM 5.2. *Let $\{\phi_m\}_{m=0}^M$ be a sequence of orthonormal basis functions in $L^2([0,1])$. 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 $(q, M - q + 1)$.

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, and hence $F_T \to_d \chi_q^2/q$. When M is not very large or the number of the restrictions q is large, the asymptotic distribution χ_q^2/q 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. Computation and Simulation.

6.1. *Computation.* To compute the OS-LRV estimator in the previous section, we have to first find the empirical Riesz representer $\widehat{v}_T^*$, which is not very appealing to applied researchers. In this subsection 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 $\mathcal{H}_T = \left\{ h(\cdot) = P_{k_T}(\cdot)'\beta : \beta \in \mathbb{R}^{k_T} \right\}$. 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 $\alpha_j = \left(\theta_j', P_{k_T}(\cdot)'\beta_j\right)'$, $j = 1, 2$, 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 $T^{-1} \sum_{t=1}^{T} \tilde{\ell}(Z_t, \gamma)$, we obtain the 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 $f_{0,T} = f(\alpha_{0,T}) = f(\theta_0, P_{k_T}(\cdot)'\beta_{0,T})$, which is estimated by the plug-in sieve M estimator $\widehat{f} = f(\widehat{\alpha}_T) = f(\widehat{\theta}_T, P_{k_T}(\cdot)'\widehat{\beta}_T)$. 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

$$\widehat{Var}(\widehat{f})=\widehat{F}_{k_T}'\widehat{\Sigma}_T\widehat{F}_{k_T},\quad \text{with } \widehat{F}_{k_T}=\left(\frac{\partial f(\widehat{\alpha}_T)}{\partial\theta'},\frac{\partial f(\widehat{\alpha}_T)}{\partial h}[P_{k_T}(\cdot)']\right)'.$$

It is easy to verify that for any sample size T, $\widehat{Var}(\widehat{f})$ is numerically identical to $\|\widehat{v}_T^*\|_{sd,T}^2$, our asymptotic variance estimator given in (5.6). 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.

6.2. *Simulation.* To examine the accuracy of our inference procedures in Section 5, we consider a partially linear regression model in our simulation study:

$$Y_t=X_{1t}'\theta_0+\tilde{h}_0(\tilde{X}_{2t})+u_t,\ E[u_t|X_{1t},\tilde{X}_{2t}]=0,\ t=1,\dots,T,$$

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

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

where $(\varepsilon_{1t}^1,\dots,\varepsilon_{1t}^d,\varepsilon_{et},\varepsilon_{ut})'$ 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 $\rho\in\{0,0.25,0.5,0.75\}$.

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 $[0,1]$ via $\tilde{X}_{2t}=\log\left(X_{2t}/(1-X_{2t})\right)$. Then $\tilde{h}_0(\tilde{X}_{2t})=\cos(\log[X_{2t}\left(1-X_{2t}\right)^{-1}])\equiv h_0\left(X_{2t}\right)$. Let $P_{k_T}\left(x_2\right)=[p_1\left(x_2\right),\dots,p_{k_T}\left(x_2\right)]'$ be a $k_T\times 1$ vector, where $\{p_j\left(x_2\right):j\geq 1\}$ is a set of basis functions on $[0,1]$. We approximate $h_0\left(X_{2t}\right)$ by $P_{k_T}\left(X_{2t}\right)'\beta$ for some $\beta=(\beta_1,\dots,\beta_{k_T})'\in\mathbb{R}^{k_T}$. Let $\mathbf{X}_t=\left(X_{1t}',P_{k_T}\left(X_{2t}\right)'\right)$ a $1\times(d+k_T)$ vector and $\mathbf{X}'=(\mathbf{X}_1',...,\mathbf{X}_T')$ a $(d+k_T)\times T$ matrix. Let $\mathbf{Y}=\left(Y_1,\dots,Y_T\right)'$, $\mathbf{U}=(u_1,\dots,u_T)'$ 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 $\theta_{1:j}=(\theta_{01},\dots,\theta_{0j})'$. 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 $\mathcal{R}_\theta\left(j\right)$ be the first j rows of the identity matrix I_{d+k_T}, then the sieve estimator of $\theta_{1:j}=\mathcal{R}_\theta\left(j\right)\gamma$ is $\widehat{\theta}_{1:j}=\mathcal{R}_\theta\left(j\right)\widehat{\gamma}_T$, and so

$$\sqrt{T}\left(\widehat{\theta}_{1:j}-\theta_{1:j}\right)=T^{-1/2}\sum_{t=1}^{T}\mathcal{R}_\theta\left(j\right)\left(\mathbf{X}'\mathbf{X}/T\right)^{-1}\mathbf{X}_t'u_t+o_p\left(1\right).$$

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

$$\widehat{\Omega}_{\theta M} = \frac{1}{M}\sum_{m=1}^{M}\left(T^{-1/2}\sum_{t=1}^{T}\phi_m(\frac{t}{T})\widehat{\Delta}_{\theta t}\right)\left(T^{-1/2}\sum_{t=1}^{T}\phi_m(\frac{t}{T})\widehat{\Delta}_{\theta t}\right)'$$

be the OS-LRV estimator of the asymptotic variance Ω of $\sqrt{T}\left(\widehat{\theta}_{1:j} - \theta_{1:j}\right)$. 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 $\phi_{2m-1}(x) = \sqrt{2}\cos(2m\pi x)$, $\phi_{2m}(x) = \sqrt{2}\sin(2m\pi x), m = 1, \ldots, M/2$ as the orthonormal basis functions for the OS-LRV estimation.

To perform either the chi-square test or the F test, we need to choose M. Here 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) for parametric models, with his kernel bandwidth $b = M^{-1}$, $q = 2, c_1 = 0, c_2 = 1, p = j$. 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 B is the asymptotic bias of $\widehat{\Omega}$, $\mathcal{X}_j^{\tau}$ is the $(1-\tau)$ quantile of χ_j^2 distribution, and $\lceil\cdot\rceil$ is the ceiling function. The parameters B and Ω in 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 have also implemented choosing M based on the mean square criterion and the simulation results are qualitatively similar.

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 $\widehat{h}(x) = \mathcal{R}_x\widehat{\gamma}_T$. 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 $\tilde{x}_2 \in \{-2, 0.1, 2\}$. Since $\tilde{X}_{2t}$ is standard normal, this range of $\tilde{x}_2$ largely covers the support of $\tilde{X}_{2t}$. Let $\widehat{\Delta}_{xt} = \mathcal{R}_x(\mathbf{X}'\mathbf{X}/T)^{-1}\mathbf{X}_t'\widehat{u}_t$ and

$$\widehat{\Omega}_{xM} = M^{-1}\sum_{m=1}^{M}\left(T^{-1/2}\sum_{t=1}^{T}\phi_m(\frac{t}{T})\widehat{\Delta}_{xt}\right)\left(T^{-1/2}\sum_{t=1}^{T}\phi_m(\frac{t}{T})\widehat{\Delta}_{xt}\right)',$$

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 6.1, we can construct the F-test version of the Wald statistic as:

$$F_x = \left(\sqrt{T}\left[\mathcal{R}_x\widehat{\gamma}_T - h_0(x)\right]\right)'\widehat{\Omega}_{xM}^{-1}\left(\sqrt{T}\left[\mathcal{R}_x\widehat{\gamma}_T - h_0(x)\right]\right). \tag{6.3}$$

As in the inference for the parametric part, we select the smoothing parameter M based on the CPE criterion. 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

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

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 (6.3) 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 1 gives the empirical null rejection probabilities for testing $\theta_{1:j} = 0$ for $j = 1, 2, 3, 4$ for $\rho \geq 0$. 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 $T = 500$.

Figure 1 presents 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 $\tilde{x}_2 \in \{-2, 0.1, 2\}$. 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. To save space, this figure only reports the case with $T = 100$ and spline sieve basis, but the pattern remains the same for both sample sizes and for both sieve bases. 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.

TABLE 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							
$\rho=0$	0.0633	0.0860	0.0685	0.0935	0.0825	0.1287	0.1115	0.2008
$\rho=0.25$	0.0621	0.1069	0.0677	0.1225	0.0806	0.1696	0.0973	0.2922
$\rho=0.50$	0.0588	0.1307	0.0635	0.1494	0.0815	0.2225	0.0997	0.3955
$\rho=0.75$	0.0521	0.1549	0.0640	0.1764	0.0874	0.2767	0.1016	0.4922
	$T=500$, Cosine and Sine Basis							
$\rho=0$	0.0597	0.0848	0.0649	0.0900	0.0760	0.1187	0.0992	0.1896
$\rho=0.25$	0.0570	0.1028	0.0648	0.1138	0.0752	0.1611	0.0886	0.2786
$\rho=0.50$	0.0539	0.1240	0.0621	0.1383	0.0706	0.2093	0.0850	0.3778
$\rho=0.75$	0.0440	0.1472	0.0574	0.1716	0.0794	0.2647	0.0904	0.4738
	$T=500$, Cosine and Sine Basis							
$\rho=0$	0.0517	0.0566	0.0521	0.0576	0.0500	0.0641	0.0607	0.0901
$\rho=0.25$	0.0531	0.0633	0.0522	0.0650	0.0513	0.0786	0.0578	0.1171
$\rho=0.50$	0.0545	0.0678	0.0527	0.0713	0.0498	0.0932	0.0512	0.1402
$\rho=0.75$	0.0511	0.0676	0.0499	0.0749	0.0447	0.1003	0.0431	0.1636
	$T=500$, Spline Basis							
$\rho=0$	0.0487	0.0544	0.0470	0.0547	0.0475	0.0606	0.0562	0.0844
$\rho=0.25$	0.0527	0.0608	0.0479	0.0629	0.0494	0.0745	0.0528	0.1108
$\rho=0.50$	0.0518	0.0656	0.0498	0.0703	0.0472	0.0900	0.0474	0.1339
$\rho=0.75$	0.0491	0.0637	0.0465	0.0688	0.0425	0.0959	0.0410	0.1555

Note: j is the number of joint hypotheses.

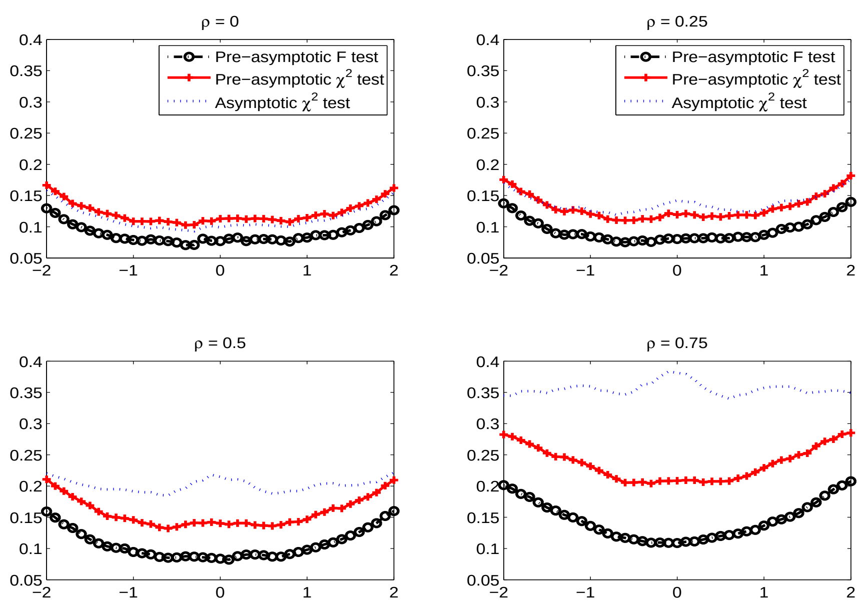

FIG 1. *Empirical Rejection Probabilities Against the value of X_{2t} with Spline Basis and $T=100$*

7. Appendix A: Mathematical Proofs.

PROOF OF THEOREM 3.1. For any $\alpha \in \mathcal{B}_T$, denote $\alpha_u^* = \alpha \pm \varepsilon_T u_T^*$ as a local alternative of α for some $\varepsilon_T = o(T^{-\frac{1}{2}})$. It is clear that if $\alpha \in \mathcal{B}_T$, then $\alpha_u^* \in \mathcal{B}_T$. Since $\widehat{\alpha}_T \in \mathcal{B}_T$ with probability approaching one (wpa1), we have that $\widehat{\alpha}_{u,T}^* = \widehat{\alpha}_T \pm \varepsilon_T u_T^* \in \mathcal{B}_T$ wpa1. By the definition of $\widehat{\alpha}_T$, we have

$$\begin{aligned}
-O_p(\varepsilon_T^2) &\leq \frac{1}{T}\sum_{t=1}^{T}\ell(Z_t,\widehat{\alpha}_T) - \frac{1}{T}\sum_{t=1}^{T}\ell(Z_t,\widehat{\alpha}_{u,T}^*) \\
&= E[\ell(Z_t,\widehat{\alpha}_T) - \ell(Z_t,\widehat{\alpha}_{u,T}^*)] + \mu_T\left\{\Delta(Z,\alpha_0)\left[\widehat{\alpha}_T - \widehat{\alpha}_{u,T}^*\right]\right\} \\
&\quad + \mu_T\left\{\ell(Z,\widehat{\alpha}_T) - \ell(Z,\widehat{\alpha}_{u,T}^*) - \Delta(Z,\alpha_0)\left[\widehat{\alpha}_T - \widehat{\alpha}_{u,T}^*\right]\right\} \\
&= E[\ell(Z_t,\widehat{\alpha}_T) - \ell(Z_t,\widehat{\alpha}_{u,T}^*)] \mp \mu_T\left\{\Delta(Z,\alpha_0)[\varepsilon_T u_T^*]\right\} + O_p(\varepsilon_T^2) \qquad (7.1)
\end{aligned}$$

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

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

Combining these with the definition of $\widehat{\alpha}_{u,T}^*$ and the inequality in (7.1), we deduce that

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

which further implies that

$$\langle\widehat{\alpha}_T - \alpha_0, u_T^*\rangle - \mu_T\left\{\Delta(Z,\alpha_0)[u_T^*]\right\} = O_p(\varepsilon_T) = o_p\left(T^{-1/2}\right). \qquad (7.2)$$

By definition of $\alpha_{0,T}$, we have $\langle\alpha_{0,T} - \alpha_0, v\rangle = 0$ for any $v \in \mathcal{V}_T$. Thus $\langle\alpha_{0,T} - \alpha_0, u_T^*\rangle = 0$, and

$$\left|\sqrt{T}\,\langle\widehat{\alpha}_T - \alpha_{0,T}, u_T^*\rangle - \sqrt{T}\mu_T\left\{\Delta(Z,\alpha_0)\left[u_T^*\right]\right\}\right| = o_p(1). \qquad (7.3)$$

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

$$\begin{aligned}
&\frac{f(\widehat{\alpha}_T) - f(\alpha_{0,T})}{\left\|v_T^*\right\|_{sd}} \\
&= \frac{f(\widehat{\alpha}_T) - f(\alpha_0) - \frac{\partial f(\alpha_0)}{\partial\alpha}[\widehat{\alpha}_T - \alpha_0]}{\left\|v_T^*\right\|_{sd}} - \frac{f(\alpha_{0,T}) - f(\alpha_0) - \frac{\partial f(\alpha_0)}{\partial\alpha}[\alpha_{0,T} - \alpha_0]}{\left\|v_T^*\right\|_{sd}} \\
&\quad + \frac{\frac{\partial f(\alpha_0)}{\partial\alpha}\left[\widehat{\alpha}_T - \alpha_0\right] - \frac{\partial f(\alpha_0)}{\partial\alpha}\left[\alpha_{0,T} - \alpha_0\right]}{\left\|v_T^*\right\|_{sd}} \\
&= \langle\widehat{\alpha}_T - \alpha_{0,T}, u_T^*\rangle + o_p\left(T^{-1/2}\right). \qquad (7.4)
\end{aligned}$$

It follows from (7.3) and (7.4) that

$$\left|\sqrt{T}\frac{f(\widehat{\alpha}_T)-f(\alpha_{0,T})}{\left\|v_T^*\right\|_{sd}}-\sqrt{T}\mu_T\left\{\Delta(Z,\alpha_0)\left[u_T^*\right]\right\}\right|=o_p(1), \tag{7.5}$$

which establishes the first result of the theorem. The second result follows immediately from (7.5) and Assumption 3.4. □

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:

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

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

$$J_{2,T}=\sum_{t=d_T+1}^{T-1}\left(1-\frac{t}{T}\right)\rho_T^*(t).$$

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

$$|J_{1,T}|\le\frac{d_T C_T}{Var\{\Delta(Z,\alpha_0)[v_T^*]\}}=o(1). \tag{7.6}$$

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

$$\left|\frac{||v_T^*||_{sd}^2}{Var\left(\Delta(Z,\alpha_0)[v_T^*]\right)}-1\right|\le 2[|J_{1,T}|+|J_{2,T}|]=o(1), \tag{7.7}$$

which establishes the first claim. This, Assumption 4.1.(i) and Theorem 3.1 together imply the asymptotic normality result in (4.3). □

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:

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

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),

$$\begin{aligned}&\left|E\left\{\Delta(Z_1,\alpha_0)[v_T^*]\Delta(Z_t,\alpha_0)[v_T^*]\right\}\right|\\&=\left|\int_{z_1\in\mathcal{Z}}\int_{z_t\in\mathcal{Z}}\Delta(z_1,\alpha_0)\left[v_T^*\right]\Delta(z_t,\alpha_0)\left[v_T^*\right]\frac{f_{Z_1,Z_t}\left(z_1,z_t\right)}{f_Z\left(z_1\right)f_Z\left(z_t\right)}f_Z\left(z_1\right)f_Z\left(z_t\right)dz_1dz_t\right|\\&\le C\left(\int_{z_1\in\mathcal{Z}}\left|\Delta(z_1,\alpha_0)\left[v_T^*\right]\right|f_Z\left(z_1\right)dz_1\right)^2=C\left\|\Delta(Z,\alpha_0)[v_T^*]\right\|_1^2,\end{aligned}$$

which implies that $C_T \leq C \|\Delta(Z,\alpha_0)[v_T^*]\|_1^2$. This and Assumption 4.3.(ii) imply the existence of a growing $d_T \to \infty$ such that $d_T C_T / \|\Delta(Z,\alpha_0)[v_T^*]\|_2^2 \to 0$, 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:

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

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

$$\begin{aligned}
&\sup_{\alpha\in\mathcal{B}_T}\sup_{v_1,v_2\in\mathcal{V}_T}\frac{\left|T^{-1}\sum_{t=1}^T r(Z_t,\alpha)[v_1,v_2] - E\{r(Z_t,\alpha_0)[v_1,v_2]\}\right|}{\|v_1\|\,\|v_2\|} \\
&\leq \sup_{\alpha\in\mathcal{B}_T}\sup_{v_1,v_2\in\mathcal{W}_T}\left|T^{-1}\sum_{t=1}^T r(Z_t,\alpha)[v_1,v_2] - E\{r(Z_t,\alpha)[v_1,v_2]\}\right| \\
&+ \sup_{\alpha\in\mathcal{B}_T}\sup_{v_1,v_2\in\mathcal{W}_T}|E\{r(Z,\alpha)[v_1,v_2] - r(Z,\alpha_0)[v_1,v_2]\}| = O_p(\epsilon_T^*).
\end{aligned} \tag{7.8}$$

Let $\alpha = \widehat{\alpha}_T$, $v_1 = \widehat{v}_T^*$ and $v_2 = v$. Then it follows from (7.8), the definitions of $\langle\cdot,\cdot\rangle$ and $\langle\cdot,\cdot\rangle_T$ that

$$\frac{\left|T^{-1}\sum_{t=1}^T r(Z_t,\widehat{\alpha}_T)[\widehat{v}_T^*,v] - E\{r(Z_t,\alpha_0)[\widehat{v}_T^*,v]\}\right|}{\left\|\widehat{v}_T^*\right\|\,\|v\|} = \left|\frac{\langle\widehat{v}_T^*,v\rangle_T - \langle\widehat{v}_T^*,v\rangle}{\left\|\widehat{v}_T^*\right\|\,\|v\|}\right| = O_p(\epsilon_T^*). \tag{7.9}$$

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

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

$$= \sup_{v\in\mathcal{V}_T}\left|\frac{\langle \widehat{v}_T^* - v_T^*, v\rangle}{\|v\|}\right| + O_p(\epsilon_T^*\,\|\widehat{v}_T^*\|). \tag{7.10}$$

This implies that

$$\sup_{v\in\mathcal{V}_T}\left|\frac{\langle \widehat{v}_T^* - v_T^*, v\rangle}{\|v\|}\right| = O_p(\epsilon_T^*\,\|\widehat{v}_T^*\|). \tag{7.11}$$

Letting $v = \widehat{v}_T^* - v_T^*$ in (7.11), we get

$$\frac{||\widehat{v}_T^* - v_T^*||}{\|v_T^*\|} = O_p\left(\epsilon_T^*\frac{||\widehat{v}_T^*||}{\|v_T^*\|}\right). \tag{7.12}$$

It follows from this result that

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

$$= O_p\left(\epsilon_T^*\left|\frac{||\widehat{v}_T^*||}{\|v_T^*\|} - 1\right|\right) + O_p\left(\epsilon_T^*\right) \tag{7.13}$$

from which we deduce that

$$\left|\frac{||\widehat{v}_T^*||}{\|v_T^*\|} - 1\right| = O_p(\epsilon_T^*). \tag{7.14}$$

Combining the results in (7.12), (7.13), and (7.14), we get $\frac{||\widehat{v}_T^* - v_T^*||}{\|v_T^*\|} = O_p(\epsilon_T^*)$ as desired. □

Proof of Theorem 5.1. **Part (i)** For $m = 1, 2, \ldots, M$, we write $\widehat{\Lambda}_m$ as

$$\begin{aligned}
\widehat{\Lambda}_m &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_m(\frac{t}{T})\left\{\Delta(Z_t,\widehat{\alpha}_T)[\widehat{v}_T^*] - E(\Delta(Z_t,\widehat{\alpha}_T)[\widehat{v}_T^*]) - \Delta(Z_t,\alpha_0)[\widehat{v}_T^*] + E(\Delta(Z_t,\alpha_0)[\widehat{v}_T^*])\right\}\\
&+ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_m(\frac{t}{T})\left\{E(\Delta(Z_t,\widehat{\alpha}_T)[\widehat{v}_T^*]) - E(\Delta(Z_t,\alpha_0)[\widehat{v}_T^*]) - E(r(Z_t,\alpha_0)\left[\widehat{v}_T^*,\widehat{\alpha}_T-\alpha_0\right])\right\}\\
&+ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_m(\frac{t}{T})E(r(Z_t,\alpha_0)\left[\widehat{v}_T^*,\widehat{\alpha}_T-\alpha_0\right]) + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_m(\frac{t}{T})\Delta(Z_t,\alpha_0)[\widehat{v}_T^*]\\
&\equiv \hat{I}_{m,1} + \hat{I}_{m,2} + \hat{I}_{m,3} + \hat{I}_{m,4}.
\end{aligned}$$

Using Assumption 5.2.(i)-(ii), we have $\hat{I}_{m,1} = o_p(\|\widehat{v}_T^*\|)$ and $\hat{I}_{m,2} = O_p\left(\sqrt{T}\epsilon_T^*\xi_T\|\widehat{v}_T^*\|\right)$. So

$$\begin{aligned}\widehat{\Lambda}_m &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_m(\frac{t}{T})\Delta(Z_t,\alpha_0)[v_T^*] + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_m(\frac{t}{T})\Delta(Z_t,\alpha_0)[\widehat{v}_T^* - v_T^*] \\ &- \left[\frac{1}{T}\sum_{t=1}^{T}\phi_m(\frac{t}{T})\right]\left[\sqrt{T}\langle v_T^*, \widehat{\alpha}_T - \alpha_0\rangle + \sqrt{T}\langle \widehat{v}_T^* - v_T^*, \widehat{\alpha}_T - \alpha_0\rangle\right] \\ &+ o_p(\|\widehat{v}_T^*\|) + O_p\left(\sqrt{T}\epsilon_T^*\xi_T\|\widehat{v}_T^*\|\right).\end{aligned}\tag{7.15}$$

Under Assumptions 3.2 and 3.3, we can invoke equation (7.2) in the proof of Theorem 3.1 to deduce that

$$\sqrt{T}\|v_T^*\|_{sd}^{-1}\langle v_T^*, \widehat{\alpha}_T - \alpha_0\rangle = \frac{1}{\sqrt{T}}\|v_T^*\|_{sd}^{-1}\sum_{t=1}^{T}\Delta(Z_t,\alpha_0)[v_T^*] + o_p(1).\tag{7.16}$$

Using Lemma 5.1 and the Hölder inequality, we get

$$\left|\sqrt{T}\langle \widehat{v}_T^* - v_T^*, \widehat{\alpha}_T - \alpha_0\rangle\right| \leq \sqrt{T}\|\widehat{v}_T^* - v_T^*\|\|\widehat{\alpha}_T - \alpha_0\| = O_p(\sqrt{T}\|v_T^*\|\epsilon_T^*\xi_T).\tag{7.17}$$

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

$$\begin{aligned}&\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_m(\frac{t}{T})\Delta(Z_t,\alpha_0)[\widehat{v}_T^* - v_T^*]\right| \\ &\leq \|\widehat{v}_T^* - v_T^*\|\sup_{v\in\mathcal{W}_T}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_m(\frac{t}{T})\Delta(Z_t,\alpha_0)[v]\right| = O_p(\|v_T^*\|\epsilon_T^*).\end{aligned}\tag{7.18}$$

Now, using Lemma 5.1, (7.15)-(7.18), 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

$$\begin{aligned}&\|v_T^*\|_{sd}^{-1}\widehat{\Lambda}_m \\ &= \frac{1}{\sqrt{T}}\|v_T^*\|_{sd}^{-1}\sum_{t=1}^{T}\left[\phi_m\left(\frac{t}{T}\right) - \frac{1}{T}\sum_{t=1}^{T}\phi_m\left(\frac{t}{T}\right)\right]\Delta(Z_t,\alpha_0)[v_T^*] + o_p(1) \\ &\overset{a}{\sim} \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\phi_m\left(\frac{t}{T}\right) - \frac{1}{T}\sum_{s=1}^{T}\phi_m\left(\frac{s}{T}\right)\right]e_t \equiv \zeta_m\end{aligned}\tag{7.19}$$

Since $\{\phi_m(\cdot), m = 0,1,\ldots,M\}$ is a set of orthonormal functions and $\phi_0(\cdot) = 1$, we have $\zeta_m \overset{a}{\sim} iid\ N(0,1)$ for $m = 1,\ldots,M$, and hence $\|v_T^*\|_{sd}^{-1}\widehat{\Lambda}_m \overset{a}{\sim} iid\ N(0,1)$ for $m = 1,\ldots,M$.

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

$$\|v_T^*\|_{sd}^{-1}\|\widehat{v}_T^*\|_{sd,T}^2\|v_T^*\|_{sd}^{-1} = \frac{1}{M}\sum_{m=1}^{M}\left(\|v_T^*\|_{sd}^{-1}\widehat{\Lambda}_m\right)^2 \overset{a}{\sim} \frac{1}{M}\sum_{m=1}^{M}\zeta_m^2.\tag{7.20}$$

which, combined with Theorem 3.1, further implies that

$$\begin{aligned} t_T &= \frac{\sqrt{T}\,[f(\widehat{\alpha}_T)-f(\alpha_0)]}{\|v_T^*\|_{sd}} \Bigg/ \frac{\|\widehat{v}_T^*\|_{sd,T}}{\|v_T^*\|_{sd}} \\ &= \frac{\sqrt{T}\,[f(\widehat{\alpha}_T)-f(\alpha_0)]}{\|v_T^*\|_{sd}} \Bigg/ \sqrt{M^{-1}\sum_{m=1}^{M}\left(\|v_T^*\|_{sd}^{-1}\widehat{\Lambda}_m\right)^2} \\ &\overset{a}{\sim} \frac{\zeta_0}{\sqrt{M^{-1}\sum_{m=1}^{M}\zeta_m^2}}. \end{aligned} \tag{7.21}$$

where $\zeta_0 = T^{-1/2}\sum_{t=1}^{T} e_t$. Since both ζ_0 and ζ_m are approximately standard normal and

$$cov\,(\zeta_0,\zeta_m) = T^{-1}\sum_{t=1}^{T}\phi_m\,(t/T) = o\,(1)\,,$$

ζ_0 is asymptotically independent of ζ_m for $m=1,\ldots,M$. This implies that $t_T \overset{a}{\sim} t\,(M)$. □

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

$$\|\mathbf{v}_T^*\|_{sd}^{-1}\,\widehat{\Lambda}_m \overset{a}{\sim} T^{-1/2}\sum_{t=1}^{T}\left[\phi_m\,(t/T) - T^{-1}\sum_{s=1}^{T}\phi_m\,(s/T)\right]\mathbf{e}_t \equiv \zeta_m \tag{7.22}$$

and $\zeta_m \overset{a}{\sim} iid\ N(0,I_q)$. It then follows that

$$\|\mathbf{v}_T^*\|_{sd}^{-1}\,\|\widehat{\mathbf{v}}_T^*\|_{sd,T}^2\left(\|\mathbf{v}_T^*\|_{sd}^{-1}\right)' \overset{a}{\sim} M^{-1}\sum_{m=1}^{M}\zeta_m\zeta_m'. \tag{7.23}$$

Using the results in (5.8) and (7.23), we have

$$\begin{aligned} F_T &= T\,[\mathbf{f}(\widehat{\alpha}_T)-\mathbf{f}(\alpha_0)]'\,\|\widehat{\mathbf{v}}_T^*\|_{sd,T}^{-2}\,[\mathbf{f}(\widehat{\alpha}_T)-\mathbf{f}(\alpha_0)]\,/q \\ &\overset{a}{\sim} \left(T^{-1/2}\sum_{t=1}^{T}\mathbf{e}_t\right)'\left\{M^{-1}\sum_{m=1}^{M}\zeta_m\zeta_m'\right\}^{-1}\left(T^{-1/2}\sum_{t=1}^{T}\mathbf{e}_t\right)/q \\ &= \zeta_0'\left\{M^{-1}\sum_{m=1}^{M}\zeta_m\zeta_m'\right\}^{-1}\zeta_0, \end{aligned} \tag{7.24}$$

where $\zeta_0 \equiv T^{-1/2}\sum_{t=1}^{T}\mathbf{e}_t$. Since $\phi_m\,(\cdot)\,,\ m=1,2,\ldots,M$ are orthonormal and integrate to zero, we have

$$F_T \overset{a}{\sim} \xi_0\left(M^{-1}\sum_{m=1}^{M}\xi_m\xi_m'\right)^{-1}\xi_0$$

where $\xi_m \sim iid\ N\,(0,I_q)$ for $m=0,\ldots,M$. 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 $[(M-q+1)\,/M]\,F_T \overset{a}{\sim} F_{q,M-q+1}$ as desired. □

References.

[1] Andrews, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–854.
[2] 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.
[3] Chen, X. and X. Shen (1998): "Sieve Extremum Estimates for Weakly Dependent Data," *Econometrica*, 66, 289–314.
[4] Chen, X., W. Wu and Y. Yi (2009): "Efficient Estimation of Copula-based Semiparametric Markov Models," *Annals of Statistics*, Vol. 37(6B), 4214–4253.
[5] Conley, T.G., L. P. Hansen, W. Liu (1997): "Bootstrapping the Long Run," *Macroeconomic Dynamics*, 1, 1997, 279–311.
[6] Fan, J. and Q. Yao (2003): *Nonlinear Time Series: Nonparametric and Parametric Methods*, New York: Springer.
[7] Gao, J. (2007): *Nonlinear Time Series: Semiparametric and Nonparametric Methods.* London: Chapman & Hall/CRC.
[8] Granger, C.W.J. (2003): "Time Series Concepts for Conditional Distributions," *Oxford Bulletin of Economics and Statistics*, 65, supplement 689–701.
[9] Grenander, U. (1981): *Abstract Inference*, New York: Wiley Series.
[10] Hall, P. and C. C. Heyde (1980): *Martingale Limit Theory and Its Applications*, New York: Academic Press.
[11] Härdle, W., H. Liang and J. Gao (2000): *Partially Linear Models.* Springer.
[12] Hotelling H. (1931): "The Generalization of Student's Ratio," *Annals of Mathematical Statistics*, 2, 360–378.
[13] Pritsker, M. (1998): "Nonparametric Density Estimation and Tests of Continuous Time Interest Rate Models," *Review of Financial Studies*, 11 449–487.
[14] Phillips, P. C. B. (2005): "HAC Estimation by Automated Regression," *Econometric Theory*, 21, 116–142.
[15] Robinson, P. (1983): "Nonparametric Estimators for Time Series," *Journal of Time Series Analysis*, 4, 185–207.
[16] Shen, X. and W. Wong (1994): "Convergence Rate of Sieve Estimates", *The Annals of Statistics,* 22, 580-615.
[17] Sun, Y. (2011a): "Robust Trend Inference with Series Variance Estimator and Testing-optimal Smoothing Parameter," *Journal of Econometrics*, 164(2), 345–366.
[18] Sun, Y. (2011b): "Asymptotic F Test in a GMM Framework," Working paper, Department of Economics, UC San Diego.
[19] 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.

Cowles Foundation for Research in Economics
Yale University
30 Hillhouse Ave., Box 208281
New Haven, CT 06520
E-mail: xiaohong.chen@yale.edu

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

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