SIEVE QUASI LIKELIHOOD RATIO INFERENCE ON
SEMI/NONPARAMETRIC CONDITIONAL MOMENT MODELS

Xiaohong Chen and Demian Pouzo

May 2013

Sieve Quasi Likelihood Ratio Inference on Semi/nonparametric Conditional Moment Models[1]

Xiaohong Chen[2] and Demian Pouzo[3]

First version: March 2009. This version: November 2012.

Abstract

This paper considers inference on functionals of semi/nonparametric conditional moment restrictions with possibly nonsmooth generalized residuals. These models belong to the difficult (nonlinear) ill-posed inverse problems with unknown operators, and include all of the (nonlinear) nonparametric instrumental variables (IV) as special cases. For these models it is generally difficult to verify whether a functional is regular (i.e., root-n estimable) or irregular (i.e., slower than root-n estimable). In this paper we provide computationally simple, unified inference procedures that are asymptotically valid regardless of whether a functional is regular or irregular. We establish the following new results: (1) the asymptotic normality of the plug-in penalized sieve minimum distance (PSMD) estimators of the (possibly iregular) functionals; (2) the consistency of sieve variance estimators of the plug-in PSMD estimators; (3) the asymptotic chi-square distribution of an optimally weighted sieve quasi likelihood ratio (SQLR) statistic; (4) the asymptotic tight distribution of a possibly non-optimally weighted SQLR statistic; (5) the consistency of the nonparametric bootstrap and the weighted bootstrap (possibly non-optimally weighted) SQLR and sieve Wald statistics, which are proved under virtually the same conditions as those for the original-sample statistics. Small simulation studies and an empirical illustration of a nonparametric quantile IV regression are presented.

Keywords: Nonlinear nonparametric instrumental variables; Penalized sieve minimum distance; Irregular functional; Sieve Riesz representer; Sieve quasi likelihood ratio; Asymptotic normality; Bootstrap; Sieve variance estimator.

[1]We thank a co-editor and two referees for very useful comments. Earlier versions, some entitled "On PSMD inference of functionals of nonparametric conditional moment restrictions", were presented in April 2009 at the Banff conference on seminonparametrics, in June 2009 at the Cemmap conference on quantile regression, in July 2009 at the SITE conference on nonparametrics, in September 2009 at the Stats in the Chateau/France, in June 2010 at the Cemmep workshop on recent developments in nonparametric instrumental variable methods, in August 2010 at the Beijing international conference on statistics and society, in May 2012 at London Cemmap Mastersclass on semi-nonparametric models and methods, in September 2012 at Toulouse conference celebrating the 65th birthday of JP Florens, and workshops in UPenn, Purdue/stat, UC Berkeley, Stanford, Seoul National, USC, UCSD, Weierstrass Institute/Berlin, Monash, Arizona Vanderbilt, Yale and Princeton. We thank D. Andrews, P. Bickel, R. Blundell, T. Christensen, J. Fan, M. Jansson, Z. Liao, O. Linton, J. Powell, A. Santos and other participants of these conferences and workshops for helpful comments. Chen acknowledges financial support from National Science Foundation grant SES-0838161. Any errors are the responsibility of the authors.

[2]Cowles Foundation for Research in Economics, Yale University, Box 208281, New Haven, CT 06520, USA. Email: xiaohong.chen@yale.edu.

[3]Department of Economics, UC Berkeley, 530 Evans Hall 3880, Berkeley, CA 94720, USA. Email: dpouzo@econ.berkeley.edu.

1 Introduction

This paper is about inference on functionals of the unknown true parameters $\alpha_0 \equiv (\theta_0', h_0)$ satisfying the semi/nonparametric conditional moment restrictions

$$E[\rho(Y_i, X_i; \theta_0, h_{01}(\cdot), ..., h_{0q}(\cdot))|X_i] = 0 \quad a.s. - X_i, \tag{1.1}$$

where Y_i is a vector of endogenous variables and X_i is a vector of conditioning (or instrumental) variables. The conditional distribution of Y_i given X_i, $F_{Y_i|X_i}$, is not specified beyond that it satisfies (1.1). $\rho(\cdot;\theta_0, h_0)$ is a $d_\rho \times 1-$vector of generalized residual functions whose functional forms are known up to the unknown parameters $\alpha_0 \equiv (\theta_0', h_0) \in \Theta \times \mathcal{H}$, with $\theta_0 \equiv (\theta_{01}, ..., \theta_{0d_\theta})' \in \Theta$ being a $d_\theta \times 1-$vector of finite dimensional parameters and $h_0(\cdot) \equiv (h_{01}(\cdot), ..., h_{0q}(\cdot)) \in \mathcal{H}$ being a $1 \times d_q-$vector valued function. The arguments of each unknown function $h_\ell(\cdot)$ may differ across $\ell = 1, ..., q$, may depend on θ, $h_{\ell'}(\cdot)$, $\ell' \neq \ell$, X_i and Y_i. The residual function $\rho(\cdot;\alpha)$ could be nonlinear and pointwise nonsmooth in the parameters $\alpha \equiv (\theta', h) \in \Theta \times \mathcal{H}$.

The general framework (1.1) nests many widely used nonparametric and semiparametric models in economics, finance and statistics. Well known examples include nonparametric mean instrumental variables regressions (NPIV): $E[Y_{1,i} - h_0(Y_{2,i})|X_i] = 0$ (e.g., Hall and Horowitz (2005), Carrasco et al. (2007), Blundell et al. (2007), Darolles et al. (2011), Horowitz (2011)); nonparametric quantile instrumental variables regressions (NPQIV): $E[1\{Y_{1,i} \leq h_0(Y_{2,i})\} - \gamma|X_i] = 0$ (e.g., Chernozhukov and Hansen (2005), Chernozhukov et al. (2007), Horowitz and Lee (2007), Chen and Pouzo (2012a), Gagliardini and Scaillet (2011)); the system of shape-invariant mean or quantile IV Engel curves (e.g., Blundell et al. (2007), Chen and Pouzo (2009)); random coefficient panel data regressions (e.g., Chamberlain (1992), Graham and Powell (2012)); semi-nonparametric asset pricing models (e.g., Hansen and Richard (1987), Gallant and Tauchen (1989), Chen and Ludvigson (2009)); semi-nonparametric static and dynamic game models (e.g., Bajari et al. (2011)); nonparametric optimal endogenous contract models (e.g., Bontempts et al. (2012)). Additional examples of the general model (1.1) can be found in Chamberlain (1992), Newey and Powell (2003), Ai and Chen (2003), Chen and Pouzo (2012a), Chen et al. (2011) and the references therein. In fact, model (1.1) includes all of the (nonlinear) semi-nonparametric IV regressions when the unknown functions $h_0(\cdot)$ depend on the endogenous variables Y_i:

$$E[\rho(Y_{1,i}; \theta_0, h_0(Y_{2,i}))|X_i] = 0, \tag{1.2}$$

which could lead to difficult (nonlinear) nonparametric ill-posed inverse problems with unknown operators.

Let $\{Z_i \equiv (Y_i', X_i')'\}_{i=1}^n$ be a strictly stationary ergodic sample with the distribution of Z_i the same as the distribution of $Z \equiv (Y', X')'$, and $F_{Y_i|X_i} = F_{Y|X}$. Then we can rewrite model (1.1) as

$$E[\rho(Z;\theta_0, h_0(\cdot))|X] = 0 \quad a.s. - X. \tag{1.3}$$

Let $\phi : \Theta \times \mathcal{H} \to \mathbb{R}^{d_\phi}$ be a functional with a fixed finite $d_\phi \geq 1$. Typical functionals include an Euclidean functional ($\phi(\alpha) = \theta$), an evaluation functional $\phi(\alpha) = h(\overline{y}_2)$ (for $\overline{y}_2 \in \text{supp}(Y_2)$), weighted derivative functionals $\phi(h) = \int w(y_2)\nabla h(y_2)dy_2$ or $\int w(y_2)\left[\nabla^2 h(y_2)\right]^2 dy_2$ (for a known positive weight $w(\cdot)$) and many others. We are interested in simple valid inference on any $\phi(\alpha_0)$ of the general model (1.3).

As pointed out in Chamberlain (1992) and Ai and Chen (2009), there is generally no closed form solution for the semiparametric efficiency bound of $\phi(\alpha_0)$ (including θ_0) satisfying the model (1.3), especially so when the residual functions $\rho(\cdot;\theta_0, h_0)$ contain several unknown functions and/or when the unknown functions $h_0(\cdot)$ of endogenous variables enter $\rho(\cdot;\theta_0, h_0)$ nonlinearly. Therefore, it is very difficult for applied researchers to verify whether the semiparametric efficiency bound for $\phi(\alpha_0)$ is singular or not. Since a non-singular efficiency bound is a necessary condition for $\phi(\alpha_0)$ to be root-n estimable, it is highly desirable for applied researchers to be able to conduct simple valid inference on $\phi(\alpha_0)$ regardless of whether it is root-n estimable or not. This is the main goal of our paper.

In this paper, for the general model (1.3) that could be nonlinearly ill-posed and for any $\phi(\alpha_0)$ that may or may not be root-n estimable, we first establish the asymptotic normality of the plug-in penalized sieve minimum distance (PSMD) estimator $\phi(\widehat{\alpha}_n)$ of $\phi(\alpha_0)$. For the model (1.3) with (pointwise) smooth residuals $\rho(Z;\alpha)$ in α_0, we propose a simple sieve variance estimator for possibly slower than root-n estimator $\phi(\widehat{\alpha}_n)$. However, there is no simple variance estimator for $\phi(\widehat{\alpha}_n)$ when $\rho(Z,\alpha)$ is not pointwise smooth in α_0 (without estimating an extra unknown nuisance function or using numerical derivatives). We then consider a PSMD criterion based test of the null hypothesis $\phi(\alpha_0) = \phi_0$. We show that an optimally weighted sieve quasi likelihood ratio (SQLR) statistic is asymptotically chi-square distributed under the null hypothesis. This allows us to construct confidence sets for $\phi(\alpha_0)$ by inverting the optimally weighted SQLR statistic, without the need to compute a variance estimator for $\phi(\widehat{\alpha}_n)$. Nevertheless, in complicated real data analysis applied researchers might like to use simple but possibly not optimally weighed PSMD procedures for estimation of and inference on $\phi(\alpha_0)$. We show that the non-optimally weighted SQLR statistic still has a tight limiting distribution regardless of whether $\phi(\alpha_0)$ is root-n estimable or not. We establish these large sample theories allowing for weakly dependent data. In addition, for i.i.d. data, we establish the consistency of both the nonparametric and weighted bootstrap (possibly non-optimally weighted) SQLR statistics under virtually the same conditions as those used to derive the limiting distribution of the original-sample SQLR statistic.[4] The bootstrap SQLR would then lead to alternative confidence sets construction for $\phi(\alpha_0)$ without the need to compute a variance estimator for $\phi(\widehat{\alpha}_n)$.

To the best of our knowledge, our paper is the first to provide a unified theory about criterion based inference on any $\phi(\alpha_0)$ satisfying the general semi-nonparametric model (1.3) with possibly nonsmooth residuals. Our results allow applied researchers to obtain limiting distribution of the plug-in PSMD estimator $\phi(\widehat{\alpha}_n)$ and to construct confidence sets for any $\phi(\alpha_0)$ regardless of whether it is regular (i.e., root-n estimable) or irregular (i.e., slower than root-n estimable).

Our new results build upon recent literature on identification and estimation of the unknown true parameters $\alpha_0 \equiv (\theta_0', h_0)$ satisfying the general model (1.3). See, e.g., Newey and Powell (2003) and Chen et al. (2011) for identification; Newey and Powell (2003), Chernozhukov et al. (2007), Chen and Pouzo (2012a) and Liao and Jiang (2012) for consistency of their respective estimators; and Chen and Pouzo (2012a) for the rate of convergence of the PSMD estimator of $h_0()$. In particular, under virtually the same conditions as those in Chen and Pouzo (2012a), we show that our nonparametric and weighted bootstrap PSMD estimators of $\alpha_0 \equiv (\theta_0', h_0)$ are consistent and achieve the same convergence rate as that of the original-sample PSMD estimator $\widehat{\alpha}_n$. This result is then used to establish the consistency of the bootstrap SQLR (and the bootstrap sieve Wald) statistics under virtually the same conditions as those used to derive the limiting distributions of the original-sample statistics. As a bonus, our convergence rate of the bootstrap PSMD estimator is also very useful for the consistency of the bootstrap Wald statistic for semiparametric two step GMM estimators of regular functionals when the first step unknown functions are estimated via a PSMD procedure. See Remark 5.1 for details.

There are some published work about estimation of and inference on θ_0 satisfying the general model (1.3) when θ_0 is assumed to be regular. See Ai and Chen (2003), Chen and Pouzo (2009) and Otsu (2011) for the root-n asymptotically normal and efficient estimation of θ_0; Ai and Chen (2003) for consistent variance estimation of the sieve minimum distance (SMD) estimator $\widehat{\theta}_n$ (with smooth residuals); and Chen and Pouzo (2009) for consistent weighted bootstrap approximation of the limiting distribution of $\sqrt{n}(\widehat{\theta}_n - \theta_0)$ for the PSMD estimator $\widehat{\theta}_n$ (with possibly nonsmooth residuals). However, none of these papers allows for irregular θ_0. When specializing our general theory to inference on θ_0 of the model (1.3), we immediately recover the results of Ai and Chen (2003) and Chen and Pouzo (2009) for regular θ_0. Additionally, our results remain valid even when θ_0 might be irregular, and we provide valid bootstrap (possibly non-optimally weighted) SQLR inference.

When specializing our theory to inference on a specific irregular functional, the evaluation func-

[4] At the costs of extra heavy notation and additional pages of proofs, we could also establish the consistency of block bootstrap (possibly non-optimally weighted) SQLR statistic for strictly stationary weakly dependent data. We leave it for future research.

tional $\phi(\alpha) = h(\overline{y}_2)$, of the (nonlinear) semi-nonparametric IV model (1.2), we automatically obtain the pointwise asymptotic normality of the PSMD estimator of $h_0(\overline{y}_2)$ and different ways to construct its confidence set. These results are directly applicable to the NPIV example with $\rho(Y_1; \theta_0, h_0(Y_2)) = Y_1 - h_0(Y_2)$ and to the NPQIV example with $\rho(Y_1; \theta_0, h_0(Y_2)) = 1\{Y_1 \leq h_0(Y_2)\} - \gamma$. Horowitz (2007) and Gagliardini and Scaillet (2011) established the pointwise asymptotic normality of their kernel based function space Tikhonov regularization estimators of $h_0(\overline{y}_2)$ for the NPIV and the NPQIV examples respectively. As demonstrated in Chen and Pouzo (2012a), the PSMD estimators are easier to compute for the general model (1.3) with possibly nonlinear residuals. In this paper we illustrate that it is also much easier to conduct the SQLR inference or a sieve Wald inference on a possibly irregular $\phi(\alpha_0)$ based on its plug-in PSMD estimator. Immediately after the first version of our paper was presented in April 2009 Banff conference on semiparametrics, the authors of Horowitz and Lee (2011) informed us that they were independently and concurrently working on uniform confidence bands for a particular SMD estimator of the NPIV example. Assuming that $h_0()$ belongs to a Sobolev ball with a bounded support and also satisfies some shape restrictions, their proof strategy depends crucially on the closed form solution of their SMD estimator of the NPIV example. We consider PSMD based inference on any possibly irregular $\phi(\alpha_0)$ of the general nonlinear semi-nonparametric model (1.3), in which the parameter space may not be compact and the PSMD estimator may not have a closed form solution.

The rest of the paper is organized as follows. Section 2 presents the plug-in PSMD estimator $\phi(\widehat{\alpha}_n)$ of any functional ϕ evaluated at $\alpha_0 \equiv (\theta_0', h_0)$ satisfying the model (1.3). It also illustrates the asymptotic results that will be established in the subsequent sections through an evaluation functional $\phi(\alpha) = h(\overline{y}_2)$ and a weighted integration functional $\phi(h) = \int w(y_2)h(y_2)dy_2$ of the NPQIV example. Section 3 derives the asymptotic normality of $\phi(\widehat{\alpha}_n)$ regardless of whether $\phi(\alpha_0)$ is regular or not. It also shows that any possibly non-optimally weighted SQLR statistic still has a tight asymptotic distribution. Section 4 provides inference procedures based on asymptotic critical values. Section 5 establishes the consistency of the bootstrap SQLR statistic and the bootstrap sieve Wald statistic for possibly irregular functionals. Section 6 presents simulation studies and an empirical illustration of the SQLR based confidence sets for the NPQIV regression. Section 7 briefly concludes with future research. Appendix A presents additional low level sufficient conditions, useful lemmas, and the consistency of a computationally attractive sieve score bootstrap. Appendices B and C contain supplementary lemmas and all the proofs.

Notation. We use "$\equiv$" to implicitly define a term or introduce a notation. For any column vector A, we let A' denote its transpose and $||A||_e$ its Euclidean norm (i.e., $||A||_e \equiv \sqrt{A'A}$, although sometimes we use $|A| = ||A||_e$ for simplicity). Let $||A||^2_W \equiv A'WA$ for a positive definite weighting matrix W. Let $\lambda_{\max}(W)$ and $\lambda_{\min}(W)$ denote the maximal and minimal eigenvalues of W respectively. All random variables $Z \equiv (Y', X')'$, $Z_i \equiv (Y_i', X_i')'$ are defined on a complete probability space $(\mathcal{Z}, \mathcal{B}_Z, P_Z)$, where P_Z is the joint probability distribution of (Y', X'). We define $(\mathcal{Z}^\infty, \mathcal{B}_Z^\infty, P_{Z^\infty})$ as the probability space of the sequences $(Z_1, Z_2, ...)$. For simplicity we assume that Y and X are continuous random variables. Let f_X (F_X) be the marginal density (cdf) of X, and $f_{Y|X}$ $(F_{Y|X})$ be the conditional density (cdf) of Y given X. We use $E_P[\cdot]$ to denote the expectation with respect to a measure P. Sometimes we use P for P_{Z^∞} and $E[\cdot]$ for $E_{P_Z}[\cdot]$. 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 $L^p(d\mu)$ and $||g||_{L^p(d\mu)}$ are used for simplicity). For any (possibly random) positive sequences $\{a_n\}_{n=1}^\infty$ and $\{b_n\}_{n=1}^\infty$, $a_n = O_P(b_n)$ means that $\lim_{c\to\infty} \limsup_n \Pr(a_n/b_n > c) = 0$; $a_n = o_P(b_n)$ means that for all $\varepsilon > 0$, $\lim_{n\to\infty} \Pr(a_n/b_n > \varepsilon) = 0$; and $a_n \asymp b_n$ means that there exist two constants $0 < c_1 \leq c_2 < \infty$ such that $c_1 a_n \leq b_n \leq c_2 a_n$. Also, we sometimes "wpa1" for an event A_n, to denote that $\Pr(A_n) \to 1$ as $n \to \infty$. We use $\mathcal{A}_n \equiv \mathcal{A}_{k(n)}$ and $\mathcal{H}_n \equiv \mathcal{H}_{k(n)}$ to denote various sieve spaces. To simplify the presentation, we assume that $\dim(\mathcal{A}_{k(n)}) \asymp \dim(\mathcal{H}_{k(n)}) \asymp k(n)$, all of which grow to infinity with the sample size n. We use *const.*, c or C to mean a positive finite constant that is independent of sample size but can take different values at different places. For sequences, $(a_n)_n$, we sometimes use $a_n \nearrow a$ $(a_n \searrow a)$ to denote, that the sequence converges to a and that is increasing (decreasing) sequence.

2 PSMD Estimation and SQLR Inference: An Overview

2.1 The Penalized Sieve Minimum Distance Estimator

Let $\{Z_i \equiv (Y_i', X_i')'\}_{i=1}^n$ be a strictly stationary weakly dependent sample with $F_{Z_i} = F_Z$ and $F_{Y_i|X_i} = F_{Y|X}$. Let $m(X, \alpha) \equiv E[\rho(Y, X; \alpha)|X] = \int \rho(y, X; \alpha) dF_{Y|X}(y)$. Let $\Sigma(X)$ be a positive definite weighting matrix, and

$$Q(\alpha) \equiv E\left[m(X, \alpha)'\Sigma(X)^{-1}m(X, \alpha)\right] \equiv E\left[||m(X, \alpha)||^2_{\Sigma^{-1}}\right]$$

be the population minimum distance (MD) criterion function. Then the semi/nonparametric conditional moment model (1.3) can be equivalently expressed as $m(X, \alpha_0) = 0$ $a.s. - X$, where $\alpha_0 \equiv (\theta_0', h_0) \in \mathcal{A} \equiv \Theta \times \mathcal{H}$, or as

$$\inf_{\alpha \in \mathcal{A}} Q(\alpha) = Q(\alpha_0) = 0.$$

Let $\Sigma_0(X) \equiv Var(\rho(Y, X; \alpha_0)|X)$ be positive definite for almost all X. We call $E\left[||m(X, \alpha)||^2_{\Sigma_0^{-1}}\right]$ the population optimally weighted MD criterion function.

Let $\phi : \mathcal{A} \to \mathbb{R}^{d_\phi}$ be a functional with a fixed finite $d_\phi \geq 1$. We are interested in inference on

$\phi(\alpha_0)$. Let

$$\widehat{Q}_n(\alpha) \equiv \frac{1}{n}\sum_{i=1}^{n} \widehat{m}(X_i,\alpha)'\widehat{\Sigma}(X_i)^{-1}\widehat{m}(X_i,\alpha) \tag{2.1}$$

be a sample estimate of $Q(\alpha)$, where $\widehat{m}(X,\alpha)$ and $\widehat{\Sigma}(X)$ are any consistent estimators of $m(X,\alpha)$ and $\Sigma(X)$ respectively. When $\widehat{\Sigma}(X) = \widehat{\Sigma}_0(X)$ is a consistent estimator of the optimal weighting matrix $\Sigma_0(X)$, we call the corresponding $\widehat{Q}_n(\alpha)$ the sample optimally weighted MD criterion.

We estimate $\phi(\alpha_0)$ by $\phi(\widehat{\alpha}_n)$, where $\widehat{\alpha}_n \equiv (\widehat{\theta}'_n, \widehat{h}_n)$ is an approximate *penalized sieve minimum distance* (PSMD) estimator of $\alpha_0 \equiv (\theta'_0, h_0)$, defined as

$$\widehat{Q}_n(\widehat{\alpha}_n) + \lambda_n Pen(\widehat{h}_n) \leq \inf_{\alpha\in\mathcal{A}_{k(n)}} \left\{\widehat{Q}_n(\alpha) + \lambda_n Pen(h)\right\} + O_{P_{Z^\infty}}(\eta_n), \tag{2.2}$$

where $\{\eta_n\}_{n=1}^{\infty}$ is a sequence of positive real values such that $\eta_n = o(1)$; $\lambda_n Pen(h) \geq 0$ is a penalty term such that $\lambda_n = o(1)$; and $\mathcal{A}_{k(n)} \equiv \Theta \times \mathcal{H}_{k(n)}$ is a finite dimensional sieve for $\mathcal{A} \equiv \Theta \times \mathcal{H}$, more precisely, $\mathcal{H}_{k(n)}$ is a finite dimensional linear sieve for $\mathcal{H}$:

$$\mathcal{H}_{k(n)} = \left\{ h \in \mathcal{H} : h(\cdot) = \sum_{k=1}^{k(n)} \beta_k q_k(\cdot) = \beta' q^{k(n)}(\cdot) \right\}, \tag{2.3}$$

where $\{q_k\}_{k=1}^{\infty}$ is a sequence of known basis functions of a Banach space $(\mathbf{H}, \|\cdot\|_{\mathbf{H}})$ such as wavelets, splines, Fourier series, Hermite polynomial series, etc. And $k(n) \to \infty$ as $n \to \infty$.

For the purely nonparametric conditional moment models $E\left[\rho(Y,X;h_0(\cdot))|X\right] = 0$, Chen and Pouzo (2012a) proposed more general approximate PSMD estimators of h_0 by allowing for possibly infinite dimensional sieves (i.e., $\dim(\mathcal{H}_{k(n)}) = k(n) \leq \infty$). Nevertheless, both the theoretical properties and Monte Carlo simulations in Chen and Pouzo (2012a) recommend the use of the PSMD procedures with finite dimensional sieves.

In this paper we first establish the large sample theories under a high level "local quadratic approximation" (LQA) condition, which allows for weakly dependent data and any consistent nonparametric estimator $\widehat{m}(X,\alpha)$ such as kernel, local linear regression and series least squares (LS) estimators. In Appendix A we provide low level sufficient conditions for this LQA assumption when data is i.i.d. and $\widehat{m}(X,\alpha)$ is a series least squares (LS) estimator of $m(X,\alpha)$:

$$\widehat{m}(X,\alpha) \equiv p^{J_n}(X)'(P'P)^{-}\sum_{i=1}^{n} p^{J_n}(X_i)\rho(Z_i,\alpha), \tag{2.4}$$

where $\{p_j()\}_{j=1}^{\infty}$ is a sequence of known basis functions that can approximate any square integrable functions of X well, $p^{J_n}(X) = (p_1(X), ..., p_{J_n}(X))'$, $P = (p^{J_n}(X_1), ..., p^{J_n}(X_n))'$, and $(P'P)^{-}$ is the generalized inverse of the matrix $P'P$. To simplify the presentation, we let $p^{J_n}(X)$ be a tensor-product linear sieve basis, and J_n be the dimension of $p^{J_n}(X)$ such that $J_n \to \infty$ slowly as $n \to \infty$. See, e.g., Newey (1997), Huang (1998) and Chen (2007) for more details about tensor-product linear sieves.

2.2 Preview of the SQLR Inference

For simplicity we let $\phi : \mathcal{A} \to \mathbb{R}$ be a real-valued functional. Let $\widehat{\phi}_n \equiv \phi(\widehat{\alpha}_n)$ be the *plug-in PSMD estimator* of $\phi(\alpha_0)$. Under some regularity conditions we establish in Theorem 3.1 that a self-normalized centered $\{\phi(\widehat{\alpha}_n) - \phi(\alpha_0)\}$ is asymptotically normal regardless of whether $\phi(\alpha_0)$ is $\sqrt{n}$ estimable or not. Denote

$$\begin{aligned} \widehat{QLR}_n(\phi_0) &\equiv n\left(\inf_{\alpha \in \mathcal{A}_{k(n)}:\phi(\alpha)=\phi_0} \widehat{Q}_n(\alpha) - \widehat{Q}_n(\widehat{\alpha}_n)\right) \\ &= n\left(\inf_{\alpha \in \mathcal{A}_{k(n)}:\phi(\alpha)=\phi_0} \widehat{Q}_n(\alpha) - \inf_{\alpha \in \mathcal{A}_{k(n)}} \widehat{Q}_n(\alpha)\right) + o_{P_{Z^\infty}}(1) \end{aligned} \tag{2.5}$$

as the *sieve quasi likelihood ratio* (SQLR) statistic. It becomes an *optimally weighted SQLR* statistic when $\widehat{Q}_n(\alpha)$ is the optimally weighted MD criterion. Under some regularity conditions, we show in Theorem 3.2 that the $\widehat{QLR}_n(\phi_0)$ is stochastically bounded under the null hypothesis of $\phi(\alpha_0) = \phi_0$, and that the optimally weighted SQLR statistic is asymptotically chi-square distributed regardless of whether $\phi(\alpha_0)$ is $\sqrt{n}$ estimable or not.

We also consider two bootstrap versions of the SQLR statistic: the nonparametric bootstrap SQLR and the Bayesian (or weighted) bootstrap SQLR. Let $\widehat{QLR}_n^B$ denote a bootstrap version of the SQLR statistic:

$$\widehat{QLR}_n^B(\widehat{\phi}_n) \equiv n\left(\inf_{\alpha \in \mathcal{A}_{k(n)}:\phi(\alpha)=\widehat{\phi}_n} \widehat{Q}_n^B(\alpha) - \inf_{\alpha \in \mathcal{A}_{k(n)}} \widehat{Q}_n^B(\alpha)\right), \tag{2.6}$$

where $\widehat{\phi}_n \equiv \phi(\widehat{\alpha}_n)$ and $\widehat{Q}_n^B(\alpha)$ is a bootstrap version of $\widehat{Q}_n(\alpha)$:

$$\widehat{Q}_n^B(\alpha) \equiv \frac{1}{n}\sum_{i=1}^{n} \widehat{m}^B(X_i,\alpha)'\widehat{\Sigma}(X_i)^{-1}\widehat{m}^B(X_i,\alpha), \tag{2.7}$$

where $\widehat{m}^B(x,\alpha)$ is a bootstrap version of $\widehat{m}(x,\alpha)$ (see Section 5 for details). In Theorem 5.2 we establish that under the null hypothesis of $\phi(\alpha_0) = \phi_0$, $\widehat{QLR}_n^B(\widehat{\phi}_n)$ converges to the same limiting distribution as that of $\widehat{QLR}_n(\phi_0)$, even for possibly non-optimally weighted SQLR statistic and possibly irregular functionals.

An illustration via the NPQIV example. As an application of their general theory, Chen and Pouzo (2012a) presented the consistency and the rate of convergence of the PSMD estimator $\widehat{h}_n \in \mathcal{H}_{k(n)}$ of the NPQIV model with i.i.d. data:

$$Y_1 = h_0(Y_2) + U, \quad \Pr(U \leq 0|X) = \gamma. \tag{2.8}$$

In this example we have $\Sigma_0(X) = \gamma(1-\gamma)$. So we could use $\widehat{\Sigma}(X) = \gamma(1-\gamma)$ and $\widehat{Q}_n(\alpha)$ given in (2.1) becomes the optimally weighted MD criterion.

Applying Theorem 3.1 of this paper, we immediately obtain:

$$\sqrt{n}\frac{\phi(\widehat{h}_n) - \phi(h_0)}{||v_n^*||_{sd}} \Rightarrow N(0,1), \quad \text{where}$$

$$\begin{aligned} ||v_n^*||_{sd}^2 &= \frac{\partial \phi(h_0)}{\partial h}[q^{k(n)}(\cdot)]' \left\{R_{k(n)}\right\}^{-1} \frac{\partial \phi(h_0)}{\partial h}[q^{k(n)}(\cdot)], \\ R_{k(n)} &= \frac{1}{\gamma(1-\gamma)} E\left(E[f_{U|Y_2,X}(0) q^{k(n)}(Y_2)|X] E[f_{U|Y_2,X}(0) q^{k(n)}(Y_2)|X]' \right). \end{aligned}$$

For the evaluation functional $\phi(h) = h(\overline{y}_2)$ we have $\frac{\partial \phi(h_0)}{\partial h}[q^{k(n)}(\cdot)] = q^{k(n)}(\overline{y}_2)$. For the weighted integration functional $\phi(h) = \int w(y_2)h(y_2)dy_2$ we have $\frac{\partial \phi(h_0)}{\partial h}[q^{k(n)}(\cdot)] = \int w(y_2)q^{k(n)}(y_2)dy_2$. See Subsection 6.1 for a Monte Carlo study regarding the finite sample performance of the asymptotic normality approximation.

Under very mild condition (see, e.g., Chen and Pouzo (2012a)), the conditional expectation operator $Th = E\left[f_{U|Y_2,X}(0)h(Y_2)|X\right]$ mapping from $h \in \mathcal{H}$ to $L^2(f_X)$ is compact, so is its adjoint operator T^*. Thus we have

$$T^*T\psi_j(Y_2) = \mu_j^2 \psi_j(Y_2), \ \mu_1^2 \geq \mu_2^2 \geq ... \ and \ \mu_j^2 \searrow 0,$$

where $\{\psi_j(\cdot) : j \geq 1\}$ is the eigenfunction sequence of $\mathcal{H}$ and $\left\{\mu_j^2 : j \geq 1\right\}$ is the eigenvalue sequence of the self-adjoint compact operator T^*T. Suppose that $(q^k(\cdot))_k$ is a Riesz basis for $\mathcal{H}$ such that $\mathcal{H}_{k(n)} = clsp\{\psi_j(\cdot) : j = 1, ..., k(n)\}$. Then $R_{k(n)} = \frac{1}{\gamma(1-\gamma)} Diag\left\{\mu_1^2, ..., \mu_{k(n)}^2\right\}$ and

$$||v_n^*||_{sd}^2 = \gamma(1-\gamma)\sum_{j=1}^{k(n)} \mu_j^{-2} \left(\frac{\partial \phi(h_0)}{\partial h}[\psi_j(\cdot)]\right)^2.$$

For the evaluation functional $\phi(h) = h(\overline{y}_2)$ we have

$$||v_n^*||_{sd}^2 = \gamma(1-\gamma)\sum_{j=1}^{k(n)} \mu_j^{-2} [\psi_j(\overline{y}_2)]^2.$$

For the weighted integration functional $\phi(h) = \int w(y_2)h(y_2)dy_2$ we have

$$||v_n^*||_{sd}^2 = \gamma(1-\gamma)\sum_{j=1}^{k(n)} \mu_j^{-2}\left(\int w(y_2)\psi_j(y_2)dy_2\right)^2.$$

Therefore the variance of $\sqrt{n}\left(\phi(\widehat{h}_n) - \phi(h_0)\right)$ is of a complicated form and may diverge to infinity as $k(n) \to \infty$.

Applying Theorem 3.2 (or more precisely Theorem 4.2(1)), we immediately obtain that the optimally weighted SQLR statistic $\widehat{QLR}_n(\phi_0) \Rightarrow \chi_1^2$ under the null of $\phi(h_0) = \phi_0$. Thus we can compute confidence regions for any functional $\phi(h)$, such as the evaluation and the weighted integration functionals, as

$$\left\{r \in \mathbb{R}:\ \widehat{QLR}_n(r) \le c_{\chi_1^2}(\tau)\right\},$$

where $c_{\chi_1^2}(\tau)$ is the $(1-\tau)$-quantile of the χ_1^2 distribution. See Subsection 6.2 for an empirical illustration of this result to the nonparametric quantile IV Engel curve regression using the British Family Survey data set that was first used in Blundell et al. (2007).

Instead of using the asymptotic critical values based on a χ_1^2 distribution, we could also construct a confidence set using bootstrap methods:

$$\left\{r \in \mathbb{R}:\ \widehat{QLR}_n(r) \le \hat{c}_n(\tau)\right\},$$

where $\hat{c}_n(\tau)$ is the bootstrap $(1-\tau)$-quantile of the asymptotic distribution of the SQLR statistic, computed via either the nonparametric bootstrap or the weighted bootstrap as described in Section 5. See Subsection 6.1 for a Monte Carlo study of the finite sample performance of the bootstrap based confidence sets.

2.3 A Brief Discussion on the Convergence Rate

Before we could derive the asymptotic distribution of the SQLR statistic, we need some consistency and convergence rate results that allow us to concentrate on some shrinking neighborhood of the true parameter value α_0 of the semi-nonparametric model (1.3). For the purely nonparametric conditional moment model $E[\rho(Y,X;h_0(\cdot))|X] = 0$, Chen and Pouzo (2012a) established the consistency and the convergence rates of their various PSMD estimators of h_0. Some of their results can be trivially extended to establish the corresponding properties of our PSMD estimator $\widehat{\alpha}_n \equiv (\widehat{\theta}_n', \widehat{h}_n)$ defined in (2.2). For the sake of easy reference and to introduce basic assumptions and notation, we present some sufficient conditions for consistency and the convergence rate here. These conditions are also needed to establish the consistency and the convergence rate of bootstrap PSMD estimators (see Lemma 5.1). We first impose three conditions on identification, sieve spaces, penalty functions and sample criterion function. We equip the parameter space $\mathcal{A} \equiv \Theta \times \mathcal{H} \subseteq \mathbb{R}^{d_\theta} \times \mathbf{H}$ with a (strong) norm $\|\alpha\|_s \equiv \|\theta\|_e + \|h\|_{\mathbf{H}}$. Let $\Pi_n\alpha \equiv (\theta', \Pi_n h) \in \mathcal{A}_{k(n)} \equiv \Theta \times \mathcal{H}_{k(n)}$.

Assumption 2.1 (Identification, sieves, criterion)**.** *(i) $E[\rho(Y, X; \alpha)|X] = 0$ if and only if $\alpha \in (\mathcal{A}, \|\cdot\|_s)$ with $\|\alpha - \alpha_0\|_s = 0$; (ii) For all $k \geq 1$, $\mathcal{A}_k \equiv \Theta \times \mathcal{H}_k$, Θ is a compact subset in $\mathbb{R}^{d_\theta}$, $\{\mathcal{H}_k : k \geq 1\}$ is a non-decreasing sequence of non-empty closed subsets of $(\mathbf{H}, \|\cdot\|_{\mathbf{H}})$ such that $\mathcal{H} \subseteq cl\left(\cup_k \mathcal{H}_k\right)$, and there is $\Pi_n h_0 \in \mathcal{H}_{k(n)}$ with $||\Pi_n h_0 - h_0||_{\mathbf{H}} = o(1)$; (iii) $Q : (\mathcal{A}, \|\cdot\|_s) \to [0, \infty)$ is lower semicontinuous, $Q(\Pi_n\alpha_0) = o(1)$.* [5] *(iv) $\Sigma(x)$ and $\Sigma_0(x)$ are positive definite, and their smallest and largest eigenvalues are finite and positive uniformly in $x \in \mathcal{X}$.*

Assumption 2.2 (Penalty)**.** *(i) $\lambda_n > 0$, $\lambda_n = o(1)$; (ii) $|Pen(\Pi_n h_0) - Pen(h_0)| = O(1)$ with $Pen(h_0) < \infty$; (iii) $Pen : (\mathcal{H}, \|\cdot\|_{\mathbf{H}}) \to [0, \infty)$ is lower semicompact.* [6]

Let $\{\eta_n\}_{n=1}^\infty$ and $\{\bar{\delta}^2_{m,n}\}_{n=1}^\infty$ be sequences of positive real values that decrease to zero as $n \to \infty$. Let $\mathcal{A}^{M_0}_{k(n)} \equiv \Theta \times \mathcal{H}^{M_0}_{k(n)} \equiv \{\alpha = (\theta', h) \in \mathcal{A}_{k(n)} : \lambda_n Pen(h) \leq \lambda_n M_0\}$ for a large but finite M_0 such that $\Pi_n\alpha_0 \in \mathcal{A}^{M_0}_{k(n)}$ and that $\widehat{\alpha}_n \in \mathcal{A}^{M_0}_{k(n)}$ with probability arbitrarily close to one for all large n.

Assumption 2.3 (Sample Criterion)**.** *(i) $\widehat{Q}_n(\Pi_n\alpha_0) \leq c_0 Q(\Pi_n\alpha_0) + O_{P_{Z^\infty}}(\eta_n)$ for some $\eta_n = o(1)$ and a finite constant $c_0 > 0$; (ii) $\widehat{Q}_n(\alpha) \geq cQ(\alpha) - O_{P_{Z^\infty}}(\bar{\delta}^2_{m,n})$ uniformly over $\mathcal{A}^{M_0}_{k(n)}$ for some $\bar{\delta}^2_{m,n} = o(1)$ and a finite constant $c > 0$.*

The following consistency result is a minor modification of Theorem 3.2 of Chen and Pouzo (2012a).

Lemma 2.1. *Let $\widehat{\alpha}_n$ be the PSMD estimator defined in (2.2). If Assumptions 2.1, 2.2 and 2.3 hold, then: $||\widehat{\alpha}_n - \alpha_0||_s = o_{P_{Z^\infty}}(1)$ and $Pen(\widehat{h}_n) = O_{P_{Z^\infty}}(1)$.*

Given the consistency result, we can restrict our attention to a shrinking $|| \cdot ||_s-$neighborhood around α_0. Let

$$\mathcal{A}_{os} \equiv \{\alpha \in \mathcal{A} : ||\alpha - \alpha_0||_s \leq \epsilon,\ \lambda_n Pen(h) \leq \lambda_n M_0\} \text{ and } \mathcal{A}_{osn} \equiv \mathcal{A}_{os} \cap \mathcal{A}_{k(n)}$$

for a positive finite constant M_0 and a small positive ϵ such that $\Pr(\widehat{\alpha}_n \notin \mathcal{A}_{osn}) < \epsilon$ eventually.

For any $\alpha \in \mathcal{A}_{os}$ we define a pathwise derivative

$$\begin{aligned}
\frac{dm(X, \alpha_0)}{d\alpha}[\alpha - \alpha_0] &\equiv \left.\frac{dE[\rho(Z, (1-\tau)\alpha_0 + \tau\alpha)|X]}{d\tau}\right|_{\tau=0} \quad a.s.\ X \\
&= \frac{dE[\rho(Z, \alpha_0)|X]}{d\theta'}(\theta - \theta_0) + \frac{dE[\rho(Z, \alpha_0)|X]}{dh}[h - h_0] \quad a.s.\ X.
\end{aligned}$$

[5] A function f is lower semicontinuous at a point x_0 iff $\lim_{x \to x_0} f(x) \geq f(x_0)$. We say is lower semicontinuous, if it is lower semicontinuous at any point.

[6] A function f is lower semicompact iff for all M, $\{x\colon f(x) \leq M\}$ is compact.

Following Ai and Chen (2003) and Chen and Pouzo (2009), we introduce a pseudo-metric $||\alpha_1 - \alpha_2||$ for any $\alpha_1, \alpha_2 \in \mathcal{A}_{os}$, as

$$||\alpha_1 - \alpha_2||^2 \equiv E\left[\left(\frac{dm(X,\alpha_0)}{d\alpha}[\alpha_1 - \alpha_2]\right)' \Sigma(X)^{-1} \left(\frac{dm(X,\alpha_0)}{d\alpha}[\alpha_1 - \alpha_2]\right)\right]. \tag{2.9}$$

The next assumption is about the local curvature of the population criterion $Q(\alpha)$.

Assumption 2.4 (Local curvature). *(i) $\mathcal{A}_{os}$ and $\mathcal{A}_{osn}$ are convex, $m(\cdot,\alpha)$ is continuously pathwise differentiable with respect to $\alpha \in \mathcal{A}_{os}$, and there is a finite constant $C > 0$ such that $||\alpha - \alpha_0|| \leq C||\alpha - \alpha_0||_s$ for all $\alpha \in \mathcal{A}_{os}$; (ii) There are finite constants $c_1, c_2 > 0$ such that $c_1||\alpha - \alpha_0||^2 \leq Q(\alpha) \leq c_2||\alpha - \alpha_0||^2$ holds for all $\alpha \in \mathcal{A}_{osn}$.*

Recall the definition of the *sieve measure of local ill-posedness*

$$\tau_n \equiv \sup_{\alpha \in \mathcal{A}_{osn}: ||\alpha - \Pi_n\alpha_0|| \neq 0} \frac{||\alpha - \Pi_n\alpha_0||_s}{||\alpha - \Pi_n\alpha_0||}.$$

The problem of estimating α_0 under $||\cdot||_s$ is *locally ill-posed in rate* if and only if $\limsup_{n\to\infty} \tau_n = \infty$. The following general rate result is a minor modification of Theorem 4.1 and Remark 4.1(1) of Chen and Pouzo (2012a), and hence we omit its proof. Let $\{\delta_{m,n}\}_{n=1}^{\infty}$ be a sequence of positive real values that decrease to zero as $n \to \infty$.

Lemma 2.2. *Let $\widehat{\alpha}_n$ be the PSMD estimator defined in (2.2) with $\max\{\lambda_n, \eta_n\} = o(n^{-1})$. Let Assumptions 2.1, 2.2, 2.3 and 2.4 hold, and $\widehat{Q}_n(\alpha) \geq cQ(\alpha) - O_{P_{Z^\infty}}(\delta_{m,n}^2)$ uniformly over $\mathcal{A}_{osn}$ for some finite constant $c > 0$. If $\max\{\delta_{m,n}^2, Q(\Pi_n\alpha_0), \lambda_n, \eta_n\} = \delta_{m,n}^2$ then:*

$$||\widehat{\alpha}_n - \alpha_0|| = O_{P_{Z^\infty}}(\delta_{m,n}) \quad \text{and} \quad ||\widehat{\alpha}_n - \alpha_0||_s = O_{P_{Z^\infty}}(||\alpha_0 - \Pi_n\alpha_0||_s + \tau_n\delta_{m,n}).$$

The above convergence rate result is applicable to strictly stationary weakly dependent data and general nonparametric estimator $\widehat{m}(X,\alpha)$ of $m(X,\alpha)$ as soon as one could compute $\delta_{m,n}^2$, the rate at which $\widehat{Q}_n(\alpha)$ goes to $Q(\alpha)$. See Chen and Pouzo (2012a) and Chen and Pouzo (2009) for low level sufficient conditions in terms of i.i.d. data and the series LS estimator (2.4) of $m(X,\alpha)$. In particular, Lemma C.2 of Chen and Pouzo (2012a) implies that under very mild conditions, we have $\widehat{Q}_n(\alpha) \asymp Q(\alpha) - O_{P_{Z^\infty}}(\delta_{m,n}^2)$ uniformly over $\mathcal{A}_{osn}$, with $\delta_{m,n}^2 \asymp \max\left\{\frac{J_n}{n}, b_{m,J_n}^2\right\}$ and $Q(\Pi_n\alpha_0) = O(b_{m,J_n}^2)$, where b_{m,J_n} is the $L^2(f_X)-$bias order of approximating $m(\cdot,\alpha)$ by the series basis $p^{J_n}(\cdot)$.

Lemma 2.2 implies that $||\widehat{\alpha}_n - \alpha_0|| = O_{P_{Z^\infty}}(\delta_n)$ and $||\widehat{\alpha}_n - \alpha_0||_s = O_{P_{Z^\infty}}(\delta_{s,n})$, where $\{\delta_n : n \geq 1\}$ and $\{\delta_{s,n} : n \geq 1\}$ are real positive sequences such that $\delta_n \asymp \delta_{m,n} = o(1)$ and $\delta_{s,n} =$

$o(1)$, $\delta_{s,n} \geq \delta_n$. Thus $\widehat{\alpha}_n \in \mathcal{N}_{osn} \subseteq \mathcal{N}_{os}$ wpa1, where

$$\begin{aligned} \mathcal{N}_{os} &\equiv \left\{\alpha \in \mathcal{A}_{os}\colon\ ||\alpha - \alpha_0|| \leq M_n\delta_n,\ ||\alpha - \alpha_0||_s \leq M_n\delta_{s,n},\ \lambda_n Pen(h) \leq \lambda_n M_0\right\}, \\ \mathcal{N}_{osn} &\equiv \mathcal{N}_{os} \cap \mathcal{A}_{k(n)}, \end{aligned}$$

with $M_n \equiv \log(\log(n))$. In the rest of the paper, we can regard $\mathcal{N}_{os}$ as the effective parameter space and $\mathcal{N}_{osn}$ as its sieve space.

3 Local Asymptotic Theory

In this section, we establish the asymptotic normality of the plug-in PSMD estimator $\phi(\widehat{\alpha}_n)$ of a possibly irregular functional $\phi : \mathcal{A} \to \mathbb{R}$ of the general model (1.3) and the limiting distribution of a properly scaled SQLR statistic.

3.1 Riesz representation

We first provide a representation of the functional of interest $\phi : \mathcal{A} \to \mathbb{R}$, which is crucial for all the subsequent asymptotic theories.

Given the definition of the norm $||\cdot||$ (in equation (2.9)) and the local parameter spaces $\mathcal{A}_{os}$ or $\mathcal{N}_{os}$, we can construct a Hilbert space $(\overline{\mathbf{V}}, ||\cdot||)$ with $\overline{\mathbf{V}} \equiv clsp(\mathcal{A}_{os} - \{\alpha_0\})$, where $clsp(\cdot)$ is the closure of the linear span under $||\cdot||$. For any $v_1, v_2 \in \overline{\mathbf{V}}$, we define an inner product induced by the metric $||\cdot||$:

$$\langle v_1, v_2\rangle = E\left[\left(\frac{dm(X,\alpha_0)}{d\alpha}[v_1]\right)' \Sigma(X)^{-1}\left(\frac{dm(X,\alpha_0)}{d\alpha}[v_2]\right)\right],$$

and for any $v \in \overline{\mathbf{V}}$ we call $v = 0$ if and only if $||v|| = 0$ (i.e., functions in $\overline{\mathbf{V}}$ are defined in an equivalent class sense according to the metric $||\cdot||$).

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

$$\frac{d\phi(\alpha_0)}{d\alpha}[v] = \left.\frac{\partial\phi(\alpha_0 + \tau v)}{\partial\tau}\right|_{\tau=0} \quad \text{for any } v \in \overline{\mathbf{V}}.$$

If $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ is *bounded* on the infinite dimensional Hilbert space $(\overline{\mathbf{V}}, ||\cdot||)$, i.e.

$$\sup_{v\in\overline{\mathbf{V}}, v\neq 0} \frac{\left|\frac{d\phi(\alpha_0)}{d\alpha}[v]\right|}{||v||} < \infty,$$

then there is a Riesz representer $v^* \in \overline{\mathbf{V}}$ of the linear functional $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ on $\overline{\mathbf{V}}$ such that

$$\frac{d\phi(\alpha_0)}{d\alpha}[v] = \langle v^*, v\rangle \text{ for all } v \in \overline{\mathbf{V}} \quad \text{and } \|v^*\| \equiv \sup_{v\in\overline{\mathbf{V}}, v\neq 0} \frac{\left|\frac{d\phi(\alpha_0)}{d\alpha}[v]\right|}{\|v\|};$$

In this case we say that $\phi(\cdot)$ is *regular* (at $\alpha = \alpha_0$). If $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ is *unbounded* on the infinite dimensional Hilbert space $(\overline{\mathbf{V}}, ||\cdot||)$, i.e.

$$\sup_{v\in\overline{\mathbf{V}}, v\neq 0} \frac{\left|\frac{d\phi(\alpha_0)}{d\alpha}[v]\right|}{\|v\|} = \infty,$$

then there does not exist a Riesz representer of the linear functional $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ on $\overline{\mathbf{V}}$; In this case we say that $\phi(\cdot)$ is *irregular* (at $\alpha = \alpha_0$).

It is known that $\phi(\cdot)$ being regular (i.e., $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ being bounded on $\overline{\mathbf{V}}$) is a necessary condition for the root-n rate of convergence of $\phi(\widehat{\alpha}_n) - \phi(\alpha_0)$. Unfortunately for complicated semi-nonparametric models (1.3), it is difficult to compute $\sup_{v\in\overline{\mathbf{V}}, v\neq 0}\left\{\left|\frac{d\phi(\alpha_0)}{d\alpha}[v]\right| / \|v\|\right\}$ explicitly; and hence difficult to verify whether a functional $\phi(\cdot)$ is regular or not.

3.1.1 Sieve Riesz representer

Define

$$\alpha_{0,n} \equiv \arg\min_{\alpha\in\mathcal{N}_{osn}} ||\alpha - \alpha_0||. \tag{3.1}$$

Let $\overline{\mathbf{V}}_{k(n)} \equiv clsp\left(\mathcal{N}_{osn} - \{\alpha_{0,n}\}\right)$, where $clsp(.)$ denotes the closed linear span under $\|\cdot\|$. Then $\overline{\mathbf{V}}_{k(n)}$ is a finite dimensional Hilbert space under $\|\cdot\|$. Moreover, $\overline{\mathbf{V}}_{k(n)}$ is dense in $\overline{\mathbf{V}}$ under $\|\cdot\|$. To simplify the presentation, we assume that $\dim(\overline{\mathbf{V}}_{k(n)}) = \dim(\mathcal{A}_{k(n)}) \asymp k(n)$, all of which grow to infinity with n. By definition we have $\langle v_n, \alpha_{0,n} - \alpha_0\rangle = 0$ for all $v_n \in \overline{\mathbf{V}}_{k(n)}$. For any $v_n = \alpha_n - \alpha_{0,n} \in \overline{\mathbf{V}}_{k(n)}$, we let

$$\frac{d\phi(\alpha_0)}{d\alpha}[v_n] = \frac{d\phi(\alpha_0)}{d\alpha}[\alpha_n - \alpha_0] - \frac{d\phi(\alpha_0)}{d\alpha}[\alpha_{0,n} - \alpha_0].$$

So $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ is also a linear functional on $\overline{\mathbf{V}}_{k(n)}$.

Note that $\overline{\mathbf{V}}_{k(n)}$ 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 unique $v_n^* \in \overline{\mathbf{V}}_{k(n)}$ such that

$$\frac{d\phi(\alpha_0)}{d\alpha}[v] = \langle v_n^*, v\rangle \text{ for all } v \in \overline{\mathbf{V}}_{k(n)} \tag{3.2}$$

and

$$\frac{d\phi(\alpha_0)}{d\alpha}[v_n^*] = \|v_n^*\|^2 \equiv \sup_{v\in\overline{\mathbf{V}}_{k(n)}:\|v\|\neq 0} \frac{\left|\frac{d\phi(\alpha_0)}{d\alpha}[v]\right|^2}{\|v\|^2} < \infty. \tag{3.3}$$

We call v_n^* the *sieve Riesz representer* of the functional $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ on $\overline{\mathbf{V}}_{k(n)}$. By definition we have for any non-zero linear functional $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$,

$$0 < \|v_n^*\|^2 = E\left[\left(\frac{dm(X,\alpha_0)}{d\alpha}[v_n^*]\right)' \Sigma(X)^{-1}\left(\frac{dm(X,\alpha_0)}{d\alpha}[v_n^*]\right)\right] < \infty.$$

We emphasize that the sieve Riesz representation (3.2) and (3.3) of the linear functional $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ on $\overline{\mathbf{V}}_{k(n)}$ always exists regardless of whether $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ is bounded on the infinite dimensional Hilbert space $(\overline{\mathbf{V}}, ||\cdot||)$ or not. If $\|v_n^*\| = O(1)$ (in fact $\|v_n^*\| \to \|v^*\| < \infty$ and $\|v^* - v_n^*\| \to 0$ as $k(n) \to \infty$), then $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ is bounded on $(\overline{\mathbf{V}}, ||\cdot||)$ and $\phi(\cdot)$ is regular (at $\alpha = \alpha_0$). If $\|v_n^*\| \to \infty$ as $k(n) \to \infty$, then $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ is unbounded on $(\overline{\mathbf{V}}, ||\cdot||)$ and $\phi(\cdot)$ is irregular (at $\alpha = \alpha_0$).

Moreover, we can always compute the sieve Riesz representer $v_n^* \in \overline{\mathbf{V}}_{k(n)}$ for any functional defined in (3.2) and (3.3) explicitly. Let $\mathcal{A}_{k(n)} = \Theta \times \mathcal{H}_{k(n)}$ where $\mathcal{H}_{k(n)}$ given in (2.3) is a finite dimensional linear sieve. Let $\|\cdot\|$ be the norm defined in (2.9) and $\overline{\mathbf{V}}_{k(n)} = \mathbb{R}^{d_\theta} \times \{v_h(\cdot) = q^{k(n)}(\cdot)'\beta : \beta \in \mathbb{R}^{k(n)}\}$ be dense in the infinite dimensional Hilbert space $(\overline{\mathbf{V}}, \|\cdot\|)$. By definition, the sieve Riesz representer $v_n^* = (v_{\theta,n}^{*\prime}, v_{h,n}^*(\cdot))' = (v_{\theta,n}^{*\prime}, q^{k(n)}(\cdot)'\beta_n^*)' \in \overline{\mathbf{V}}_{k(n)}$ of $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ solves the following optimization problem:

$$\begin{aligned}\frac{d\phi(\alpha_0)}{d\alpha}[v_n^*] = \|v_n^*\|^2 &= \sup_{v=(v_\theta',v_h)'\in\overline{\mathbf{V}}_{k(n)},v\neq 0} \frac{\left|\frac{\partial\phi(\alpha_0)}{\partial\theta'}v_\theta + \frac{\partial\phi(\alpha_0)}{\partial h}[v_h(\cdot)]\right|^2}{E\left[\left(\frac{dm(X,\alpha_0)}{d\alpha}[v]\right)'\Sigma(X)^{-1}\left(\frac{dm(X,\alpha_0)}{d\alpha}[v]\right)\right]} \\ &= \sup_{\gamma=(v_\theta',\beta')'\in\mathbb{R}^{d_\theta+k(n)},\gamma\neq 0} \frac{\gamma' F_{k(n)} F_{k(n)}' \gamma}{\gamma' R_{k(n)}\gamma},\end{aligned} \tag{3.4}$$

where

$$F_{k(n)} \equiv \left(\frac{\partial\phi(\alpha_0)}{\partial\theta'}, \frac{\partial\phi(\alpha_0)}{\partial h}[q^{k(n)}(\cdot)']\right)' \tag{3.5}$$

is a $(d_\theta + k(n)) \times 1$ vector,[7] and

$$\gamma' R_{k(n)}\gamma \equiv E\left[\left(\frac{dm(X,\alpha_0)}{d\alpha}[v]\right)'\Sigma(X)^{-1}\left(\frac{dm(X,\alpha_0)}{d\alpha}[v]\right)\right] \quad \text{for all } v = \left(v_\theta', q^{k(n)}(\cdot)'\beta\right)' \in \overline{\mathbf{V}}_{k(n)}, \tag{3.6}$$

[7]When $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ applies to a vector (matrix), it stands for element-wise (column-wise) operations. We follow the same convention for other operators such as $\frac{dm(X,\alpha_0)}{d\alpha}[\cdot]$ throughout the paper.

with $R_{k(n)}$ being $(d_\theta + k(n)) \times (d_\theta + k(n))$ a positive definite matrix, where

$$R_{k(n)} \equiv \begin{pmatrix} I_{11} & I_{n,12} \\ I'_{n,21} & I_{n,22} \end{pmatrix} \quad \text{with}$$

$I_{11} = E\left[\left(\frac{dm(X,\alpha_0)}{d\theta'}\right)' \Sigma(X)^{-1} \frac{dm(X,\alpha_0)}{d\theta'}\right]$, $I_{n,22} = E\left[\left(\frac{dm(X,\alpha_0)}{dh}[q^{k(n)}(\cdot)']\right)' \Sigma(X)^{-1} \left(\frac{dm(X,\alpha_0)}{dh}[q^{k(n)}(\cdot)']\right)\right]$, and $I_{n,12} = E\left[\left(\frac{dm(X,\alpha_0)}{d\theta'}\right)' \Sigma(X)^{-1} \left(\frac{dm(X,\alpha_0)}{dh}[q^{k(n)}(\cdot)']\right)\right] \in \mathbb{R}^{d_\theta \times k(n)}$.

The sieve Riesz representation (3.2) becomes: for all $v = \left(v'_\theta, q^{k(n)}(\cdot)'\beta\right)' \in \overline{\mathbf{V}}_{k(n)}$,

$$\frac{d\phi(\alpha_0)}{d\alpha}[v] = F'_{k(n)}\gamma = \langle v_n^*, v \rangle = \gamma_n^{*\prime} R_{k(n)}\gamma \quad \text{for all } \gamma = (v'_\theta, \beta')' \in \mathbb{R}^{d_\theta + k(n)}. \tag{3.7}$$

It is obvious that the optimal solution of γ in (3.4) or in (3.7) has a closed-form expression:

$$\gamma_n^* = \left(v_{\theta,n}^{*\prime}, \beta_n^{*\prime}\right)' = R_{k(n)}^{-1} F_{k(n)}. \tag{3.8}$$

The sieve Riesz representer is then given by

$$v_n^* = (v_{\theta,n}^{*\prime}, v_{h,n}^*(\cdot))' = (v_{\theta,n}^{*\prime}, q^{k(n)}(\cdot)'\beta_n^*)' \in \overline{\mathbf{V}}_{k(n)}.$$

Consequently,

$$\|v_n^*\|^2 = \gamma_n^{*\prime} R_{k(n)} \gamma_n^* = F'_{k(n)} R_{k(n)}^{-1} F_{k(n)}, \tag{3.9}$$

and hence $\phi(\cdot)$ is regular (or irregular) at $\alpha = \alpha_0$ if and only if $\lim_{k(n)\to\infty} F'_{k(n)} R_{k(n)}^{-1} F_{k(n)} < \infty$ (or $= \infty$).

Remark 3.1. *Recall that*

$$\begin{aligned} R_{k(n)}^{-1} &\equiv \begin{pmatrix} I_n^{11} & I_n^{12} \\ I_n^{21} & I_n^{22} \end{pmatrix} = \begin{pmatrix} I_n^{11} & -I_{11}^{-1} I_{n,12} I_n^{22} \\ -I_{n,22}^{-1} I'_{n,21} I_n^{11} & I_n^{22} \end{pmatrix} \quad \textit{with} \\ I_n^{11} &\equiv \left(I_{11} - I_{n,12} I_{n,22}^{-1} I'_{n,21}\right)^{-1}, \; I_n^{22} \equiv \left(I_{n,22} - I'_{n,21} I_{11}^{-1} I_{n,12}\right)^{-1}. \end{aligned}$$

For the Euclidean parameter functional $\phi(\alpha) = \lambda'\theta$, we have $F_{k(n)} = (\lambda', \mathbf{0}'_{k(n)})'$ with $\mathbf{0}'_{k(n)} = [0, ..., 0]_{1\times k(n)}$, and hence $v_n^* = (v_{\theta,n}^{*\prime}, q^{k(n)}(\cdot)'\beta_n^*)' \in \overline{\mathbf{V}}_{k(n)}$ with $v_{\theta,n}^* = I_n^{11}\lambda$, $\beta_n^* = I_n^{21}\lambda = -I_{n,22}^{-1} I'_{n,21} v_{\theta,n}^*$, and

$$\|v_n^*\|^2 = F'_{k(n)} R_{k(n)}^{-1} F_{k(n)} = \lambda' I_n^{11} \lambda \le \lambda_{\max}(I_n^{11}) \times \lambda'\lambda.$$

Thus the functional $\phi(\alpha) = \lambda'\theta$ is regular if $\lim_{k(n)\to\infty} \lambda_{\max}(I_n^{11}) < \infty$. In this case,

$$\lim_{k(n)\to\infty} \|v_n^*\|^2 = \lim_{k(n)\to\infty} \lambda' I_n^{11} \lambda = \lambda' \mathcal{I}_*^{-1} \lambda = \|v^*\|^2 ,$$

where

$$\mathcal{I}_* = \inf_{\mathbf{w}} E\left[\left\| \Sigma(X)^{-\frac{1}{2}} \left(\frac{dm(X,\alpha_0)}{d\theta'} - \frac{dm(X,\alpha_0)}{dh}[\mathbf{w}] \right) \right\|_e^2 \right], \tag{3.10}$$

and $v^* = (v_\theta^{*\prime}, v_h^*(\cdot))' \in \overline{\mathbf{V}}$ where $v_\theta^* \equiv \mathcal{I}_*^{-1}\lambda$, $v_h^* \equiv -\mathbf{w}^* \times v_\theta^*$, and $\mathbf{w}^*$ solves (3.10). That is, $v^* = (v_\theta^{*\prime}, v_h^*(\cdot))'$ becomes the Riesz representer for $\phi(\alpha) = \lambda'\theta$ previously computed in Ai and Chen (2003) and Chen and Pouzo (2009). Moreover, if $\Sigma(X) = \Sigma_0(X)$ and data is i.i.d., then $\mathcal{I}_*$ becomes semiparametric efficiency bound for θ_0 satisfying the model (1.3).

3.1.2 Local characterization of $\phi(\alpha)$

As it will become clear later (see Theorem 3.1), the convergence rate of $\phi(\widehat{\alpha}_n) - \phi(\alpha_0)$ depends on the order of $\|v_n^*\|$, and

$$\|v_n^*\|_{sd}^2 \equiv Var\left(n^{-1/2} \sum_{i=1}^{n} \left(\frac{dm(X_i,\alpha_0)}{d\alpha}[v_n^*] \right)' \Sigma(X_i)^{-1} \rho(Z_i,\alpha_0) \right), \tag{3.11}$$

which could go to infinity if $\|v_n^*\| \to \infty$ as $k(n) \to \infty$. Denote

$$S_{n,i}^* \equiv \left(\frac{dm(X_i,\alpha_0)}{d\alpha}[v_n^*] \right)' \Sigma(X_i)^{-1} \rho(Z_i,\alpha_0) \tag{3.12}$$

as the sieve score associated with the i-th observation. Let $\Sigma_0(X) \equiv Var(\rho(Y,X;\alpha_0)|X)$. Then

$$Var\left(S_{n,1}^*\right) = E\left[\left(\frac{dm(X,\alpha_0)}{d\alpha}[v_n^*] \right)' \Sigma(X)^{-1} \Sigma_0(X) \Sigma(X)^{-1} \left(\frac{dm(X,\alpha_0)}{d\alpha}[v_n^*] \right) \right]. \tag{3.13}$$

Let $\rho_n^*(i) \equiv E\left(S_{n,1}^* S_{n,i+1}^*\right) / Var\left(S_{n,1}^*\right)$ for all $i \geq 1$. Then

$$\|v_n^*\|_{sd}^2 = Var\left(S_{n,1}^*\right) \times \left[1 + 2 \sum_{i=1}^{n-1} \left(1 - \frac{i}{n}\right) \rho_n^*(i) \right].$$

The triangular array sieve score process $\left\{S_{n,i}^*\right\}_{i=1}^n$ is said to be weakly dependent if $\sum_{i=1}^{n-1} \left(1 - \frac{i}{n}\right) \rho_n^*(i) = O(1)$. If $\left\{S_{n,i}^*\right\}_{i=1}^n$ is a martingale difference array, then $\rho_n^*(i) = 0$ for all $i \geq 1$ and $\|v_n^*\|_{sd}^2 = Var\left(S_{n,1}^*\right)$. In general we have $\|v_n^*\|_{sd}^2 = O\left\{Var\left(S_{n,1}^*\right)\right\}$ for strictly stationary weakly dependent data.

Let

$$u_n^* \equiv \frac{v_n^*}{\|v_n^*\|_{sd}} \tag{3.14}$$

be the "scaled sieve Riesz representer". Let $\mathcal{T}_n \equiv \{t \in \mathbb{R}\colon |t| \leq 4M_n^2\delta_n\}$, where M_n and δ_n are the ones used in the definition of $\mathcal{N}_{osn}$.

Assumption 3.1 (Local behavior of ϕ)**.** *(i)* $v \mapsto \frac{d\phi(\alpha_0)}{d\alpha}[v]$ *is a bounded linear functional mapping from* $\overline{\mathbf{V}}_{k(n)}$ *to* $\mathbb{R}$*;* $\liminf_{n\to\infty} ||v_n^*|| > 0$ *and* $0 < c \leq \frac{||v_n^*||_{sd}}{||v_n^*||} \leq C < \infty$*;*

$$\text{(ii)} \qquad \sup_{(\alpha,t)\in\mathcal{N}_{osn}\times\mathcal{T}_n} \left| \phi\left(\alpha + tu_n^*\right) - \phi(\alpha_0) - \frac{d\phi(\alpha_0)}{d\alpha}[\alpha + tu_n^* - \alpha_0] \right| = o\left(n^{-1/2}\left\|v_n^*\right\|\right);$$

(iii) $\left|\frac{d\phi(\alpha_0)}{d\alpha}[\alpha_{0,n} - \alpha_0]\right| = o\left(n^{-1/2}\left\|v_n^*\right\|\right)$.

Assumption 3.1(i) requires that $\|v_n^*\|_{sd}$ is proportional to $||v_n^*||$, which is satisfied under mild conditions. To see this, we note that $\|v_n^*\|^2 \asymp Var\left(S_{n,1}^*\right)$ under Assumption 2.1(iv). We also have $\|v_n^*\|_{sd}^2 \asymp Var\left(S_{n,1}^*\right)$ for typical semi-nonparametric models with strictly stationary weakly dependent data. Thus Assumption 3.1(i) is satisfied.

Assumption 3.1(ii) controls the linear approximation error of a possibly nonlinear functional $\phi(\cdot)$, which is automatically satisfied when $\phi(\cdot)$ is a linear functional.

Assumption 3.1(iii) controls the bias part due to the finite dimensional sieve approximation of $\alpha_{0,n}$ to α_0. It is a condition imposed on the growth rate of the sieve dimension $\dim(\mathcal{A}_{k(n)}) \asymp k(n)$, and requires that the sieve approximation error rate is of a smaller order than $n^{-1/2}\left\|v_n^*\right\|$. Given Assumption 3.1(i), Assumption 3.1(iii) requires that the sieve bias term, $\left|\frac{d\phi(\alpha_0)}{d\alpha}[\alpha_{0,n} - \alpha_0]\right|$, is of a smaller order than that of the sieve standard deviation term, $n^{-1/2}\left\|v_n^*\right\|_{sd}$.

Remark 3.2. *When* $\phi(\cdot)$ *is a regular functional, we have* $\|v_n^*\| \to \|v^*\| < \infty$*, and since* $\langle v_n^*, \alpha_{0,n} - \alpha_0\rangle = 0$ *(by definition of* $\alpha_{0,n}$*), we have* $\left|\frac{d\phi(\alpha_0)}{d\alpha}[\alpha_{0,n} - \alpha_0]\right| \leq \|v^* - v_n^*\| \times \|\alpha_{0,n} - \alpha_0\|$*, and hence Assumption 3.1(iii) is satisfied if* $||v^* - v_n^*|| \times ||\alpha_{0,n} - \alpha_0|| = o(n^{-1/2})$*, which is similar to assumption 4.2 in Ai and Chen (2003) and Ai and Chen (2007) and assumption 3.2(iii) in Chen and Pouzo (2009) for regular functionals of the model (1.3).*

3.2 Local quadratic approximation (LQA)

The next assumption is about the local quadratic approximation (LQA) to the sample criterion difference along the scaled sieve Riesz representer direction $u_n^* = v_n^* / \left\|v_n^*\right\|_{sd}$.

For any $t_n \in \mathcal{T}_n$, we let $\widehat{\Lambda}_n(\alpha(t_n), \alpha) \equiv 0.5\{\widehat{Q}_n(\alpha(t_n)) - \widehat{Q}_n(\alpha)\}$ with $\alpha(t_n) \equiv \alpha + t_n u_n^*$. Denote

$$\mathbb{Z}_n \equiv n^{-1}\sum_{i=1}^{n}\left(\frac{dm(X_i, \alpha_0)}{d\alpha}[u_n^*]\right)' \Sigma(X_i)^{-1}\rho(Z_i, \alpha_0) = n^{-1}\sum_{i=1}^{n}\frac{S_{n,i}^*}{\|v_n^*\|_{sd}}. \tag{3.15}$$

Assumption 3.2 (LQA)**.** *(i) For all* $(\alpha, t) \in \mathcal{N}_{osn} \times \mathcal{T}_n$*,* $\alpha(t) \in \mathcal{A}_{k(n)}$*; and with* $r_n(t_n) = \left(\max\{t_n^2, t_n n^{-1/2}, o(n^{-1})\}\right)^{-1}$,

$$\sup_{(\alpha,t_n)\in\mathcal{N}_{osn}\times\mathcal{T}_n} r_n(t_n)\left|\widehat{\Lambda}_n(\alpha(t_n),\alpha) - t_n\left\{\mathbb{Z}_n + \langle u_n^*, \alpha-\alpha_0\rangle\right\} - \frac{B_n}{2}t_n^2\right| = o_{P_{Z^\infty}}(1),$$

where B_n *is a* Z^n *measurable positive random variable with* $B_n = O_{P_{Z^\infty}}(1)$*; (ii)* $\sqrt{n}\mathbb{Z}_n \Rightarrow N(0,1)$*.*

Assumption 3.2(i) implicitly imposes restrictions on the nonparametric estimator $\widehat{m}(x,\alpha)$ of the conditional mean function $m(x,\alpha)$ in a shrinking neighborhood of α_0. A difficulty in the verification of Assumption 3.2(i) is that, due to the non-smooth residual function $\rho(Z,\alpha)$, the estimator $\widehat{m}(x,\alpha)$ (and hence the sample criterion function $\widehat{Q}_n(\alpha)$) could be pointwise non-smooth with respect to α. Fortunately, under Assumption 2.4(i), $\widehat{Q}_n(\alpha)$ could be well approximated by a "smooth" version of it uniformly in $\alpha \in \mathcal{N}_{osn}$. For the regular functional $\phi(\alpha) = \lambda'\theta$ of the model (1.3) with i.i.d. data, Ai and Chen (2003) and Chen and Pouzo (2009) already provide low level sufficient conditions when $\widehat{m}(x,\alpha)$ is a series LS estimator. Their conditions can be modified slightly to allow for irregular functionals as well. In Appendix A we present one set of low level sufficient conditions (Assumption A) for possibly irregular functionals $\phi(\cdot)$ of the model (1.3) with possibly non-smooth residuals. The next lemma formally states the result.

Lemma 3.1. *Let* $\{Z_i\}_{i=1}^n$ *be i.i.d.,* $\widehat{m}$ *be the series LS estimator (2.4) and conditions for Lemma 2.2 hold. If Assumption A in Appendix A holds, then Assumption 3.2(i) holds.*

Assumption 3.2(ii) is a standard one. For i.i.d. data, it is implied by the following Lindeberg condition: For all $\epsilon > 0$,

$$\limsup_{n\to\infty} E\left[\left(\frac{S_{n,i}^*}{\|v_n^*\|_{sd}}\right)^2 1\left\{\left|\frac{S_{n,i}^*}{\epsilon\sqrt{n}\,\|v_n^*\|_{sd}}\right| > 1\right\}\right] = 0, \tag{3.16}$$

which, under Remark 3.2, is automatically satisfied when the functional $\phi(\cdot)$ is regular (i.e., $\|v_n^*\| \to \|v^*\| < \infty$). This is why Assumption 3.2(ii) is not imposed in Ai and Chen (2003) and Chen and Pouzo (2009) in their root-n asymptotically normal estimation of the regular functional $\phi(\alpha) = \lambda'\theta$.

3.3 Asymptotic normality of the plug-in PSMD estimator

We now establish the asymptotic normality of the plug-in PSMD estimator $\phi(\widehat{\alpha}_n)$ of a possibly irregular functional $\phi(\alpha_0)$ of the general model (1.3).

Theorem 3.1. *Let* $\widehat{\alpha}_n$ *be the PSMD estimator (2.2) and conditions for Lemma 2.2 hold. Let Assumptions 3.1(i) and 3.2(i) hold. Then: (1)* $\sqrt{n}\langle u_n^*, \widehat{\alpha}_n - \alpha_0\rangle = -\sqrt{n}\mathbb{Z}_n + o_{P_{Z^\infty}}(1)$*.*

(2) If, in addition, Assumptions 3.1(ii)(iii) and 3.2(ii) hold, then:

$$\sqrt{n}\frac{\phi(\widehat{\alpha}_n) - \phi(\alpha_0)}{||v_n^*||_{sd}} = -\sqrt{n}\mathbb{Z}_n + o_{P_{Z^\infty}}(1) \Rightarrow N(0,1).$$

When the functional $\phi(\cdot)$ is regular at $\alpha = \alpha_0$ (i.e., $\|v_n^*\| = O(1)$), we have $\|v_n^*\|_{sd} \asymp \|v_n^*\| = O(1)$ typically; so $\phi(\widehat{\alpha}_n)$ converges to $\phi(\alpha_0)$ at the parametric rate of $1/\sqrt{n}$. When the functional $\phi(\cdot)$ is irregular at $\alpha = \alpha_0$ (i.e., $\|v_n^*\| \to \infty$), we have $\|v_n^*\|_{sd} \to \infty$ (under Assumption 3.1(i)); so the convergence rate of $\phi(\widehat{\alpha}_n)$ becomes slower than $1/\sqrt{n}$. Regardless of whether the sieve variance $\|v_n^*\|_{sd}^2$ (defined in (3.11)) stays bounded asymptotically (i.e., as $n \to \infty$) or not, it always captures whatever true temporal dependence exists in finite samples.

For any regular functional of the semi-nonparametric model (1.3), Theorem 3.1 implies that

$$\sqrt{n}\left(\phi(\widehat{\alpha}_n) - \phi(\alpha_0)\right) = -n^{-1/2}\sum_{i=1}^{n} S_{n,i}^* + o_{P_{Z^\infty}}(1) \Rightarrow N(0, \sigma_{v^*}^2),$$

with

$$\sigma_{v^*}^2 = \lim_{n\to\infty} \|v_n^*\|_{sd}^2 = \lim_{n\to\infty} Var\left(n^{-1/2}\sum_{i=1}^{n} S_{n,i}^*\right) \in (0,\infty).$$

If the sieve score process $\{S_{n,i}^*\}_{i=1}^n$ (defined in (3.12)) is a martingale difference array (which is the case with i.i.d. data), then $\|v_n^*\|_{sd}^2 = Var\left(S_{n,i}^*\right)$ and

$$\sigma_{v^*}^2 = \lim_{n\to\infty} Var\left(S_{n,1}^*\right) = E\left[\left(\frac{dm(X,\alpha_0)}{d\alpha}[v^*]\right)' \Sigma(X)^{-1}\Sigma_0(X)\Sigma(X)^{-1}\left(\frac{dm(X,\alpha_0)}{d\alpha}[v^*]\right)\right].$$

Thus, our Theorem 3.1 is a natural extension of the asymptotic normality results of Ai and Chen (2003) and Chen and Pouzo (2009) for the specific regular functional $\phi(\alpha_0) = \lambda'\theta_0$ of the model (1.3) with i.i.d. data. See Remark 3.1 for further discussion.

Theorem 3.1 is similar to that of Chen et al. (2012) on the asymptotic normality of their plug-in sieve M estimators for possibly irregular functionals of time series models, except that our model (1.3) allows for nonparametric endogeneity. In the special case of generalized nonlinear least squares models, $\frac{dm(X,\alpha_0)}{d\alpha}[v] = \frac{d\rho(Z,\alpha_0)}{d\alpha}[v]$ for all directions v, our Theorem 3.1 recovers their asymptotic normality result on the plug-in sieve generalized nonlinear least squares estimator: $\sqrt{n}\frac{\phi(\widehat{\alpha}_n)-\phi(\alpha_0)}{||v_n^*||_{sd}} \Rightarrow N(0,1)$ with $\|v_n^*\|_{sd}^2 = Var\left(n^{-1/2}\sum_{i=1}^n \left(\frac{d\rho(Z_i,\alpha_0)}{d\alpha}[v_n^*]\right)' \Sigma(X_i)^{-1}\rho(Z_i,\alpha_0)\right)$.

3.4 Asymptotic distribution of the SQLR

We now characterize the asymptotic behavior of the possibly *non-optimally weighted* SQLR statistic $\widehat{QLR}_n(\phi_0)$ defined in (2.5).

Let $\widehat{\alpha}_n^R \in \{\alpha \in \mathcal{A}_{k(n)} \colon \phi(\alpha) = \phi_0\}$ be the restricted PSMD estimator of $\alpha_0 \equiv (\theta_0', h_0)$, defined as

$$\widehat{Q}_n(\widehat{\alpha}_n^R) + \lambda_n Pen(\widehat{h}_n^R) \leq \inf_{\alpha \in \mathcal{A}_{k(n)}\colon \phi(\alpha)=\phi_0} \left\{\widehat{Q}_n(\alpha) + \lambda_n Pen(h)\right\} + O_{P_{Z^\infty}}(\eta_n). \tag{3.17}$$

Then:

$$\widehat{QLR}_n(\phi_0) = n\left(\widehat{Q}_n(\widehat{\alpha}_n^R) - \widehat{Q}_n(\widehat{\alpha}_n)\right) = n\left(\inf_{\alpha \in \mathcal{A}_{k(n)}\colon \phi(\alpha)=\phi_0} \widehat{Q}_n(\alpha) - \inf_{\alpha \in \mathcal{A}_{k(n)}} \widehat{Q}_n(\alpha)\right) + o_{P_{Z^\infty}}(1).$$

Theorem 3.2. *Let $\widehat{\alpha}_n$ be the PSMD estimator (2.2), $\widehat{\alpha}_n^R$ be the restricted PSMD estimator (3.17) and conditions for Lemma 2.2 hold. Let Assumptions 3.1 and 3.2 with $\left|B_n - ||u_n^*||^2\right| = o_{P_{Z^\infty}}(1)$ hold. Then, under the null hypothesis of $\phi(\alpha_0) = \phi_0$, we have:*

$$||u_n^*||^2 \times \widehat{QLR}_n(\phi_0) = \left(\sqrt{n}\mathbb{Z}_n\right)^2 + o_{P_{Z^\infty}}(1) \Rightarrow \chi_1^2.$$

Compared to Theorem 3.1(2) on the asymptotic normality of $\phi(\widehat{\alpha}_n)$, Theorem 3.2 on the limiting distribution of the SQLR statistic requires an extra condition $\left|B_n - ||u_n^*||^2\right| = o_{P_{Z^\infty}}(1)$, which is also needed even for QLR statistics in parametric extremum estimation and testing problems. Assumption B below provides a simple sufficient condition for $\left|B_n - ||u_n^*||^2\right| = o_{P_{Z^\infty}}(1)$.

Remark 3.3. *If $\phi(\cdot)$ is continuous in $\mathcal{A}_{k(n)}$, then $\{\alpha \in \mathcal{A}_{k(n)} \colon \phi(\alpha) = \phi_0\}$ is a closed set of $\mathcal{A}_{k(n)}$. Given Assumption 3.1, following the same proof of Lemma 2.2 (see, e.g., Chen and Pouzo (2012a) or Chen and Pouzo (2009)) with the sieve $\mathcal{A}_{k(n)}$ replaced by $\{\alpha \in \mathcal{A}_{k(n)} \colon \phi(\alpha) = \phi_0\}$, we obtain $\widehat{\alpha}_n^R \in \mathcal{N}_{osn}$ wpa1. under the null hypothesis of $\phi(\alpha_0) = \phi_0$.*

4 Inference Based on Asymptotic Critical Values

In this section we provide two simple inference procedures for possibly irregular functionals of the general model (1.3). The first one is based on the asymptotic normality Theorem 3.1 with a sieve variance estimator. The second one is based on Theorem 3.2 with the optimally weighted SQLR statistic.

4.1 Sieve estimation of variance of $\phi(\widehat{\alpha}_n)$

In order to apply the asymptotic normality Theorem 3.1, we need an estimator of the sieve variance $\|v_n^*\|_{sd}^2$ defined in (3.11). In this section we propose a simple consistent estimator of $\|v_n^*\|_{sd}^2$ assuming i.i.d. data.

The theoretical sieve Riesz representer v_n^* is not known and has to be estimated. Let $\|\cdot\|_{n,\widehat{W}}$

denote the empirical norm induced by the following empirical inner product

$$\langle v_1, v_2 \rangle_{n,\widehat{W}} \equiv \frac{1}{n}\sum_{i=1}^{n}\left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n)}{d\alpha}[v_1]\right)' \widehat{W}(X_i)\left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n)}{d\alpha}[v_2]\right), \tag{4.1}$$

for any $v_1, v_2 \in \overline{\mathbf{V}}_{k(n)}$, where $W(X)$ is a positive definite weighting matrix for almost all X. Likewise we introduce a theoretical inner product associated with the weighting matrix W as

$$\langle v_1, v_2 \rangle_W \equiv E\left[\left(\frac{dm(X,\alpha_0)}{d\alpha}[v_1]\right)' W(X)\left(\frac{dm(X,\alpha_0)}{d\alpha}[v_2]\right)\right].$$

To simplify notation we also have:

$$\langle v_1, v_2 \rangle_{\Sigma^{-1}} \equiv \langle v_1, v_2 \rangle \quad \text{and} \quad \langle v_1, v_2 \rangle_{\Sigma_0^{-1}} \equiv \langle v_1, v_2 \rangle_0 \quad \text{for all } v_1, v_2 \in \overline{\mathbf{V}}_{k(n)}.$$

We define an empirical sieve Riesz representer $\widehat{v}_n^*$ of the functional $\frac{d\phi(\widehat{\alpha}_n)}{d\alpha}[\cdot]$ with respect to the empirical norm $||\cdot||_{n,\widehat{\Sigma}^{-1}}$ as

$$\frac{d\phi(\widehat{\alpha}_n)}{d\alpha}[\widehat{v}_n^*] = \sup_{v\in\overline{\mathbf{V}}_{k(n)},v\neq 0}\frac{|\frac{d\phi(\widehat{\alpha}_n)}{d\alpha}[v]|^2}{||v||^2_{n,\widehat{\Sigma}^{-1}}} < \infty \tag{4.2}$$

and

$$\frac{d\phi(\widehat{\alpha}_n)}{d\alpha}[v] = \langle \widehat{v}_n^*, v \rangle_{n,\widehat{\Sigma}^{-1}} \quad \text{for any } v \in \overline{\mathbf{V}}_{k(n)}. \tag{4.3}$$

Recall that $\|v_n^*\|_{sd}^2 = Var(S_{n,1}^*)$ when $\{Z_i\}_{i=1}^n$ is i.i.d., we can define a simple sieve variance estimator as

$$||\widehat{v}_n^*||^2_{n,sd} = ||\widehat{v}_n^*||^2_{n,\widehat{\Sigma}^{-1}\widehat{\Sigma}_0\widehat{\Sigma}^{-1}}$$

where $\widehat{\Sigma}_0$ is a consistent estimator of Σ_0, e.g. $\widehat{E}_n[\rho(Z,\widehat{\alpha}_n)\rho(Z,\widehat{\alpha}_n)' \mid \cdot]$, where $\widehat{E}_n$ is some consistent estimator of the conditional mean, such as a series, Kernel or local polynomial based estimator. In the following we denote $\overline{\mathbf{V}}^1_{k(n)} \equiv \{v \in \overline{\mathbf{V}}_{k(n)} \colon ||v|| = 1\}$.

Assumption 4.1. *(i)* $\sup_{\alpha\in\mathcal{N}_{osn}}\sup_{v\in\overline{\mathbf{V}}^1_{k(n)}}\left|\frac{d\phi(\alpha)}{d\alpha}[v] - \frac{d\phi(\alpha_0)}{d\alpha}[v]\right| = o_{P_{Z^\infty}}(1)$*; (ii) for any* $\alpha \in \mathcal{N}_{osn}$*,* $v \mapsto \frac{d\widehat{m}(\cdot,\alpha)}{d\alpha}[v] \in L^2(f_X)$ *is a bounded linear functional measurable with respect to* Z^n*; and*

$$\begin{aligned} \sup_{v_1,v_2\in\overline{\mathbf{V}}^1_{k(n)}}\left|\langle v_1, v_2\rangle_{n,\Sigma^{-1}} - \langle v_1, v_2\rangle_{\Sigma^{-1}}\right| &= o_{P_{Z^\infty}}(1); \\ \sup_{v\in\overline{\mathbf{V}}^1_{k(n)}}\left|\langle v, v\rangle_{n,\Sigma^{-1}\Sigma_0\Sigma^{-1}} - \langle v, v\rangle_{\Sigma^{-1}\Sigma_0\Sigma^{-1}}\right| &= o_{P_{Z^\infty}}(1); \end{aligned}$$

(iii) For $\Gamma(\cdot) \in \{\Sigma(\cdot), \Sigma_0(\cdot)\}$, $\sup_{\mathcal{X}}||\widehat{\Gamma}(x) - \Gamma(x)|| = o_{P_{Z^\infty}}(1)$*;* $\widehat{\Gamma}(x)$ *is positive definite and its*

smallest and largest eigenvalues are finite and positive uniformly in $x \in \mathcal{X}$ wpa1.

Assumption 4.1(i) becomes vacuous if ϕ is linear. Assumption 4.1(ii) implicitly assumes that the residual function $\rho()$ is "smooth" in $\alpha \in \mathcal{N}_{osn}$ (see, e.g., Ai and Chen (2003)) or that $\frac{d\widehat{m}(X,\widehat{\alpha}_n)}{d\alpha}[v]$ can be well approximated by numerical derivatives (see, e.g., Hong et al. (2010)).

Theorem 4.1. *(1) Let $\{Z_i\}_{i=1}^n$ be i.i.d. and conditions for Lemma 2.2 hold. If Assumption 4.1 holds, then:*

$$\left| \frac{||\widehat{v}_n^*||_{n,sd}}{||v_n^*||_{sd}} - 1 \right| = o_{P_{Z^\infty}}(1).$$

(2) If, in addition, all the assumptions of Theorem 3.1(2) hold, then:

$$\sqrt{n}\frac{\phi(\widehat{\alpha}_n) - \phi(\alpha_0)}{||\widehat{v}_n^*||_{n,sd}} = -\sqrt{n}\mathbb{Z}_n + o_{P_{Z^\infty}}(1) \Rightarrow N(0,1).$$

Theorem 4.1 is stated for the semi-nonparametric conditional moment model (1.3) with i.i.d. data. It says that one could treat the linear sieve approximation to unknown function α as if it were parametric and the corresponding parametric standard errors are consistent for the PSMD estimator of the unknown function. This is a natural generalization of the standard error calculation in Newey (1997) for series LS regression. Recently, Chen et al. (2012) proposed sieve robust long run variance estimation for a plug-in sieve M estimator of semi-nonparametric time series models. Their result can be extended to our model (1.3) with strictly stationary weakly dependent data.

Theorem 4.1 (or its time series extension) allows us to construct confidence sets for $\phi(\alpha_0)$ based on a possibly non-optimally weighted plug-in PSMD estimator $\phi(\widehat{\alpha}_n)$. A potential drawback, is that it requires a consistent estimator for $v \mapsto \frac{dm(\cdot,\alpha_0)}{d\alpha}[v]$, which may be hard to compute in practice when the residual function $\rho(Z,\alpha)$ is not pointwise smooth in $\alpha \in \mathcal{N}_{osn}$ such as in the NPQIV (2.8) example.

4.2 Optimally Weighted SQLR

For the specific regular functional $\phi(\alpha) = \lambda'\theta$ of the semi-nonparametric conditional moment model (1.3), Chen and Pouzo (2009) established that the optimally weighted SQLR statistic is asymptotically chi-square distributed. This result is an extension of the earlier results in Murphy and der Vaart (2000), Fan et al. (2001) and Shen and Shi (2005) on semiparametric likelihood ratio test of regular functionals. Here we show that the same result remains valid even for irregular functionals.

In this subsection, to stress the fact that we focus on the optimally weighted PSMD procedure, we use v_n^0 and $||v_n^0||_0$ to denote the corresponding v_n^* and $||v_n^*||$ computed using the optimal weighting matrix $\Sigma = \Sigma_0$. For instance,

$$||v_n^0||_0^2 = E\left[\left(\frac{dm(X,\alpha_0)}{d\alpha}[v_n^0]\right)' \Sigma_0(X)^{-1}\left(\frac{dm(X,\alpha_0)}{d\alpha}[v_n^0]\right)\right].$$

We call the corresponding sieve score, $S_{n,i}^0 \equiv \left(\frac{dm(X_i,\alpha_0)}{d\alpha}[v_n^0]\right)' \Sigma_0(X_i)^{-1}\rho(Z_i,\alpha_0)$, the optimal sieve score. Note that $||v_n^0||_{sd} = ||v_n^0||_0$ if the optimal sieve score process $\left\{S_{n,i}^0\right\}_{i=1}^n$ is a martingale difference array. If $||v_n^0||_{sd} = ||v_n^0||_0$ we call the SQLR statistic the optimally weighted SQLR statistic. Applying Theorem 3.2, we immediately obtain that the optimally weighted SQLR is asymptotically chi-square distributed. This result allows us to compute confidence sets for $\phi(\alpha)$ without the need of a consistent variance estimator for $\phi(\hat{\alpha}_n)$.

By Theorem 3.1(2), $||v_n^0||_{sd}^2 = ||v_n^0||_0^2$ is the asymptotic variance of the optimally weighted PSMD estimator $\phi(\hat{\alpha}_n)$. We could compute a consistent estimator of the variance $||v_n^0||_{sd}^2 = ||v_n^0||_0^2$ by looking at the "slope" of the optimally weighted SQLR. More precisely, let

$$\widehat{||v_n^0||_0^2} \equiv \left(\frac{\widehat{Q}_n(\widetilde{\alpha}_n) - \widehat{Q}_n(\widehat{\alpha}_n)}{\varepsilon_n^2}\right)^{-1} \tag{4.4}$$

where $\tilde{\alpha}_n$ is an approximate minimizer of $\widehat{Q}_n(\alpha)$ over $\{\alpha \in \mathcal{A}_{k(n)}\colon \phi(\alpha) = \phi(\widehat{\alpha}_n) - \varepsilon_n\}$. As discussed in Remark 3.3, one can show that $\tilde{\alpha}_n \in \mathcal{N}_{osn}$ wpa1 easily.

We now formally state these results.

Theorem 4.2. *Let $\widehat{\alpha}_n$ be the optimally weighted PSMD estimator (2.2) with $\Sigma = \Sigma_0$. Let all the conditions of Theorem 3.2 hold with $||v_n^0||_{sd} = ||v_n^0||_0$. Then: (1) under the null hypothesis of $\phi(\alpha_0) = \phi_0$, we have:*

$$\widehat{QLR}_n(\phi_0) = \left(\sqrt{n}\frac{\phi(\widehat{\alpha}_n) - \phi(\alpha_0)}{||v_n^0||_0}\right)^2 + o_{P_{Z^\infty}}(1) = \left(\sqrt{n}\mathbb{Z}_n\right)^2 + o_{P_{Z^\infty}}(1) \Rightarrow \chi_1^2.$$

(2) If $cn^{-1/2}||v_n^0||_0 \le \varepsilon_n \le C\delta_n||v_n^0||_0$ for finite constants $c, C > 0$, and $\tilde{\alpha}_n \in \mathcal{N}_{osn}$ wpa1, then:

$$\frac{\widehat{||v_n^0||_0^2}}{||v_n^0||_0^2} = 1 + o_{P_{Z^\infty}}(1).$$

Theorem 4.2(1) recommends to construct confidence sets for $\phi(\alpha)$ by inverting the optimally weighted SQLR statistic:

$$\left\{r \in \mathbb{R}\colon\ \widehat{QLR}_n(r) \equiv n\left(\inf_{\alpha \in \mathcal{A}_{k(n)}:\phi(\alpha)=r} \widehat{Q}_n(\alpha) - \widehat{Q}_n(\widehat{\alpha}_n)\right) \le c_{\chi_1^2}(\tau)\right\},$$

where $c_{\chi_1^2}(\tau)$ is the $(1-\tau)$-quantile of the χ_1^2 distribution. This result extends that of Chen and Pouzo (2009) to allow for irregular functionals.

When $\widehat{\alpha}_n$ is the optimally weighted PSMD estimator of α_0, Theorem 4.2(2) suggests $\widehat{||v_n^0||_0^2}$ defined in (4.4) as an alternative consistent variance estimator for $\phi(\widehat{\alpha}_n)$. Compared to Theorem 4.1, this alternative variance estimator $\widehat{||v_n^0||_0^2}$ allows for a non-smooth residual function $\rho(Z,\alpha)$, but is only valid for an optimally weighted PSMD estimator. Theorem 4.2(2) extends the result of Murphy and der Vaart (2000) on consistent variance estimation for their profile likelihood estimator of the specific regular functional $\lambda'\theta$ to our semi-nonparametric conditional moment framework (1.3), allowing for possibly irregular functionals.

5 Inference Based on Bootstrap

The inference procedures described in Section 4 are based on the asymptotic critical values. For many parametric models it is known that bootstrap based procedures could approximate finite sample distributions more accurately. In this section we establish the consistency of both the nonparametric and the weighted bootstrap sieve Wald and SQLR statistics under virtually the same conditions as those imposed for the original-sample sieve Wald and SQLR statistics.

Throughout this section, we assume that the original sample $Z^n \equiv \{Z_i\}_{i=1}^n$ is i.i.d. in this section. A bootstrap procedure is described by an array of positive "weights" $\{\omega_{i,n}\}_{i=1}^n$ for each n (we omit the n subscript hereafter), where each bootstrap sample is drawn independently of the original data $\{Z_i\}_{i=1}^n$. Different bootstrap procedures correspond to different choices of the weights $\{\omega_{i,n}\}_{i=1}^n$. For the time being we assume that $\lim_{n\to\infty} Var(\omega_{i,n}) = \sigma_\omega^2 \in (0,\infty)$ for all i.

To be more precise, we introduce some definitions for the new random variables and the enlarged probability spaces. Let $\Omega = \{\omega_{i,n} \colon i = 1,...,n;\ n = 1,...\}$ be the space of weights, defined as a triangle array with elements in $\mathbb{R}_+$, the corresponding σ-algebra and probability are $(\mathcal{B}_\Omega, P_\Omega)$. Let $\mathcal{V}^\infty \equiv \mathcal{Z}^\infty \times \Omega$, $\mathcal{B}^\infty \equiv \mathcal{B}_Z^\infty \times \mathcal{B}_\Omega$ be the σ-algebra, and let P_{V^∞} be the joint probability over $\mathcal{V}^\infty$. Finally, for each n, let $\mathcal{B}^n$ be the σ-algebra generated by $V^n \equiv Z^n \times (\omega_{1,n},...,\omega_{n,n})$.

Let A_n be a random variable that is measurable with respect to $\mathcal{B}^n$. Let $\mathcal{L}_{V^\infty|Z^\infty}(A_n|Z^n)$ be the conditional law of random variable A_n given Z^n. Let $\mathcal{L}(B)$ denote the law of a random variable B. For two real valued random variables, A_n (measurable with respect to $\mathcal{B}^n$) and B (measurable with respect to some σ-algebra $\mathcal{B}_B$), we say

$$\left|\mathcal{L}_{V^\infty|Z^\infty}(A_n|Z^n) - \mathcal{L}(B)\right| = o_{P_{Z^\infty}}(1)$$

if for any $\delta > 0$, there exists a $N(\delta)$ such that

$$P_{Z^\infty}\left(\sup_{f\in BL_1}|E[f(A_n)|Z^n] - E[f(B)]| \leq \delta\right) \geq 1-\delta \quad \text{for all } n \geq N(\delta),$$

which is equivalent to say

$$\sup_{f\in BL_1}|E[f(A_n)|Z^n] - E[f(B)]| = o_{P_{Z^\infty}}(1),$$

where BL_1 denotes the class of uniformly bounded Lipschitz functions $f : \mathbb{R} \to \mathbb{R}$ such that $||f||_{L^\infty} \leq 1$ and $|f(z) - f(z')| \leq |z - z'|$; see Van der Vaart and Wellner (1996) (henceforth, VdV-W) ch. 1.12 for more details.

We say Δ_n is of order $o_{P_{V^\infty|Z^\infty}}(1)$ in P_{Z^∞} probability, and denote it as $\Delta_n = o_{P_{V^\infty|Z^\infty}}(1)$ $wpa1(P_{Z^\infty})$, if for any $\epsilon > 0$,

$$P_{Z^\infty}\left(P_{V^\infty|Z^\infty}(|\Delta_n| > \epsilon \mid Z^n) > \epsilon\right) \to 0, \ as \ n \to \infty.$$

We say Δ_n is of order $O_{P_{V^\infty|Z^\infty}}(1)$ in P_{Z^∞} probability, and denote it as $\Delta_n = O_{P_{V^\infty|Z^\infty}}(1)$ $wpa1(P_{Z^\infty})$, if for any $\epsilon > 0$ there exists a $M \in (0, \infty)$, such that

$$P_{Z^\infty}\left(P_{V^\infty|Z^\infty}(|\Delta_n| > M \mid Z^n) > \epsilon\right) \to 0, \ as \ n \to \infty.$$

In this section we first establish the consistency of various bootstrap statistics under a high level LQA condition for general bootstrap procedures. We then provide low level sufficient conditions for two widely used bootstrap procedures: the weighted bootstrap and the nonparametric bootstrap and, which are described below:

Assumption Boot.1 (Weighted Bootstrap). *Let $(\omega_i)_{i=1}^n$ be a sequence such that $\omega_i \in \mathbb{R}_+$, $\omega_i \sim iidP_\omega$, $E[\omega] = 1$, $Var(\omega) = \sigma_\omega^2$, and $\int_0^\infty \sqrt{P(|\omega - 1| \geq t)}dt < \infty$.*

Assumption Boot.2 (Nonparametric Bootstrap). *Let $(\omega_{in})_{i=1}^n$ be a triangular array of random variables such that $(\omega_{1n}, ..., \omega_{nn}) \sim Multinomial(n; n^{-1}, ..., n^{-1})$.*

Henceforth we omit the n subscript from the weight series. We note that under Assumption Boot.2, $E[\omega_1] = 1$, $Var(\omega_1) = (1 - 1/n) \to 1 \equiv \sigma_\omega^2$ and $Cov(\omega_i, \omega_j) = -n^{-1}$ (for $i \neq j$). Finally, $n^{-1}\max_{1\leq i\leq n}(\omega_i - 1)^2 = o_{P_\omega}(1)$ (see Van der Vaart and Wellner (1996) p. 458). We use these facts in the proofs.

5.1 Consistency and convergence rate of the bootstrap PSMD estimators

Let $V^n \equiv Z^n \times (\omega_{1,n}, ..., \omega_{n,n})$, where $\{\omega_{i,n}\}_{i=1}^n$ is a sample of positive random variables that are independent of the original-sample Z^n, and each $\omega_{i,n}$ acts as a "weight" of Z_i. Let

$$\rho^B(V_i, \alpha) \equiv \omega_{i,n}\rho(Z_i, \alpha),$$

be the bootstrap residual function. Let $\widehat{m}^B(x, \alpha)$ be a bootstrap version of $\widehat{m}(x, \alpha)$, that is, $\widehat{m}^B(x, \alpha)$ is computed in the same way as that of $\widehat{m}(x, \alpha)$ except that we use $\rho^B(V_i, \alpha)$ instead of $\rho(Z_i, \alpha)$. For example, if $\widehat{m}(x, \alpha)$ is a series LS estimator (2.4) of $m(x, \alpha)$, then $\widehat{m}^B(x, \alpha)$ is a bootstrap series LS estimator of $m(x, \alpha)$, defined as:

$$\widehat{m}^B(x, \alpha) \equiv p^{J_n}(x)'(P'P)^{-} \sum_{j=1}^{n} p^{J_n}(X_j)\rho^B(V_j, \alpha). \tag{5.1}$$

Let $\widehat{Q}_n^B(\alpha) \equiv \frac{1}{n}\sum_{i=1}^n \widehat{m}^B(X_i, \alpha)'\widehat{\Sigma}(X_i)^{-1}\widehat{m}^B(X_i, \alpha)$ be a bootstrap version of $\widehat{Q}_n(\alpha)$. We denote the bootstrap PSMD estimator as $\widehat{\alpha}_n^B$, i.e., $\widehat{\alpha}_n^B$ is the approximate minimizer of $\left\{\widehat{Q}_n^B(\alpha) + \lambda_n Pen(h)\right\}$ on $\mathcal{A}_{k(n)}$.

In this subsection we establish the consistency and the convergence rate of the bootstrap PSMD estimator $\widehat{\alpha}_n^B$ under virtually the same conditions as those imposed for the consistency and the convergence rate of the original-sample PSMD estimator $\widehat{\alpha}_n$.

The next assumption is needed to control the difference of the bootstrap criterion function $\widehat{Q}_n^B(\alpha)$ and the original-sample criterion function $\widehat{Q}_n(\alpha)$; it is analogous to Assumption 2.3 for the original sample. Let $\{\overline{\delta}_{m,n}^*\}_{n=1}^\infty$ be a sequence of real valued positive numbers such that $\overline{\delta}_{m,n}^* = o(1)$ and $\overline{\delta}_{m,n}^* \geq \delta_{m,n}$. Let c_0^* and c^* be finite positive constants.

Assumption 5.1 (Bootstrap sample criterion)**.** *(i)* $\widehat{Q}_n^B(\Pi_n\alpha_0) \leq c_0^*\widehat{Q}_n(\Pi_n\alpha_0) + O_{P_{V^\infty|Z^\infty}}(\eta_n)$ *wpa1*(P_{Z^∞})*; (ii)* $\widehat{Q}_n^B(\alpha) \geq c^*\widehat{Q}_n(\alpha) - O_{P_{V^\infty|Z^\infty}}((\overline{\delta}_{m,n}^*)^2)$ *uniformly over* $\mathcal{A}_{k(n)}^{M_0}$ *wpa1*(P_{Z^∞})*.*

Lemma 5.1. *Let Assumptions 2.1 - 2.3 and 5.1 hold. Then: (1)*

$$||\widehat{\alpha}_n^B - \alpha_0||_s = o_{P_{V^\infty|Z^\infty}}(1) \ wpa1(P_{Z^\infty}) \quad and \quad Pen\left(\widehat{h}_n^B\right) = O_{P_{V^\infty|Z^\infty}}(1) \ wpa1(P_{Z^\infty}).$$

(2) In addition, let Assumption 2.4 hold and $\widehat{Q}_n^B(\alpha) \geq c^*\widehat{Q}_n(\alpha) - O_{P_{V^\infty|Z^\infty}}(\delta_{m,n}^2)$ *uniformly over* $\mathcal{A}_{osn}$ *wpa1*(P_{Z^∞})*. If* $\max\{\delta_{m,n}^2, Q(\Pi_n\alpha_0), \lambda_n, \eta_n\} = \delta_{m,n}^2$*, then:*

$$\begin{aligned} ||\widehat{\alpha}_n^B - \alpha_0|| &= O_{P_{V^\infty|Z^\infty}}(\delta_{m,n}) \ wpa1(P_{Z^\infty}); \\ ||\widehat{\alpha}_n^B - \alpha_0||_s &= O_{P_{V^\infty|Z^\infty}}(||\Pi_n\alpha_0 - \alpha_0||_s + \tau_n \times \delta_{m,n}) \ wpa1(P_{Z^\infty}). \end{aligned}$$

Lemma 5.1(2) shows that $\widehat{\alpha}_n^B \in \mathcal{N}_{osn}$ wpa1. Again, when $\widehat{m}^B(x,\alpha)$ is the bootstrap series LS estimator (5.1) of $m(x,\alpha)$, under virtually the same low level sufficient conditions as those in Chen and Pouzo (2012a) and Chen and Pouzo (2009) for their original-sample PSMD estimator $\widehat{\alpha}_n^B$, one can verify Assumption 5.1 and $\widehat{Q}_n^B(\alpha) \geq c^* \widehat{Q}_n(\alpha) - O_{P_{V^\infty|Z^\infty}}(\delta_{m,n}^2)$ uniformly over $\mathcal{A}_{osn}$ wpa1(P_{Z^∞}). This verification is amounts to follow the proof of Lemma C.2 of Chen and Pouzo (2012a) except that the original-sample series LS estimator $\widehat{m}(x,\alpha)$ is replaced by its bootstrap version $\widehat{m}^B(x,\alpha)$.

Remark 5.1. *Theorem B of Chen et al. (2003) establish the consistency of nonparametric bootstrap for a general class of semiparametric two step GMM estimators $\widehat{\theta}_{gmm}$ of root-n estimable Euclidean parameter θ_0:*

$$\left|\mathcal{L}_{V^\infty|Z^\infty}\left(\sqrt{n}\left(\widehat{\theta}_{gmm}^B - \widehat{\theta}_{gmm}\right) \mid Z^n\right) - \mathcal{L}\left(\sqrt{n}\left(\widehat{\theta}_{gmm} - \theta_0\right)\right)\right| = o_{P_{Z^\infty}}(1).$$

Their theorem is proved under a high level assumption that the first step nonparametric bootstrap estimator $\widehat{h}_n^B$ of unknown function h_0 satisfies $||\widehat{h}_n^B - \widehat{h}_n|| = o_{P_{V^\infty|Z^\infty}}\left(n^{-1/4}\right)$ wpa1(P_{Z^∞}). Our Lemmas 2.2 and 5.1 together imply that $||\widehat{h}_n^B - \widehat{h}_n|| = O_{P_{V^\infty|Z^\infty}}(\delta_{m,n})$ wpa1(P_{Z^∞}). Since $\delta_{m,n} \asymp \delta_n = o(n^{-1/4})$ under mild smoothness condition on h_0 (see, e.g., Chen and Pouzo (2012a)), our Lemma 5.1 immediately verifies their convergence rate assumption.

5.2 Bootstrap local quadratic approximation (LQA^B)

For any $t_n \in \mathcal{T}_n$, we let $\widehat{\Lambda}_n^B(\alpha(t_n),\alpha) \equiv 0.5\{\widehat{Q}_n^B(\alpha(t_n)) - \widehat{Q}_n^B(\alpha)\}$ with $\alpha(t_n) \equiv \alpha + t_n u_n^*$. For any sequence of non-negative weights $(b_i)_i$, let

$$\mathbb{Z}_n^b \equiv n^{-1}\sum_{i=1}^n b_i \left(\frac{dm(X_i,\alpha_0)}{d\alpha}[u_n^*]\right)' \Sigma(X_i)^{-1}\rho(Z_i,\alpha_0) = n^{-1}\sum_{i=1}^n b_i \frac{S_{n,i}^*}{\|v_n^*\|_{sd}}.$$

The next assumption is a bootstrap version of the LQA Assumption 3.2.

Assumption 5.2 (LQA^B). *(i) For all $(\alpha,t) \in \mathcal{N}_{osn} \times \mathcal{T}_n$, $\alpha(t) \in \mathcal{A}_{k(n)}$, and with $r_n(t_n) = \left(\max\{t_n^2, t_n n^{-1/2}, o(n^{-1})\}\right)^{-1}$,*

$$\sup_{(\alpha,t_n)\in\mathcal{N}_{osn}\times\mathcal{T}_n} r_n(t_n)\left|\widehat{\Lambda}_n^B(\alpha(t_n),\alpha) - t_n\left\{\mathbb{Z}_n^\omega + \langle u_n^*, \alpha-\alpha_0\rangle\right\} - \frac{B_n^\omega}{2}t_n^2\right| = o_{P_{V^\infty|Z^\infty}}(1) \ wpa1(P_{Z^\infty})$$

where B_n^ω is a V^n measurable positive random variable such that $B_n^\omega = O_{P_{V^\infty|Z^\infty}}(1)$ wpa1(P_{Z^∞});

$$(ii) \quad \left|\mathcal{L}_{V^\infty|Z^\infty}\left(\sqrt{n}\frac{\mathbb{Z}_n^{\omega-1}}{\sigma_\omega} \mid Z^n\right) - \mathcal{L}(\mathbb{Z})\right| = o_{P_{Z^\infty}}(1),$$

where $\mathbb{Z}$ is a standard normal random variable.

Assumption 5.2(i) implicitly imposes restrictions on the bootstrap estimator $\widehat{m}^B(x,\alpha)$ of the conditional mean function $m(x,\alpha)$. Below we provide low level sufficient conditions for Assumption 5.2(i) when $\widehat{m}^B(x,\alpha)$ is a bootstrap series LS estimator.

Denote $g(X,u_n^*) \equiv \{\frac{dm(X,\alpha_0)}{d\alpha}[u_n^*]\}'\Sigma(X)^{-1}$. Then: $E\left[g(X_i,u_n^*)\Sigma(X_i)g(X_i,u_n^*)'\right] = ||u_n^*||^2$ by definition.

Assumption B. *For $\Gamma(\cdot) \in \{\Sigma(\cdot),\Sigma_0(\cdot)\}$,*

$$\left| n^{-1}\sum_{i=1}^{n} g(X_i,u_n^*)\Gamma(X_i)g(X_i,u_n^*)' - E\left[g(X_i,u_n^*)\Gamma(X_i)g(X_i,u_n^*)'\right]\right| = o_{P_{Z^\infty}}(1).$$

Lemma 5.2. *Let $\{Z_i\}_{i=1}^n$ be i.i.d., $\widehat{m}^B(\cdot,\alpha)$ be the bootstrap series LS estimator (5.1), and conditions of Lemmas 3.1 and 5.1 hold. Let either Assumption Boot.1 or Assumption Boot.2 hold. Then:*

(1) Assumption 5.2(i) holds with $B_n^\omega = B_n$.

(2) If Assumption B holds, then $\left|B_n^\omega - ||u_n^*||^2\right| = o_{P_{V^\infty|Z^\infty}}(1)$ wpa1(P_{Z^∞}) and $\left|B_n - ||u_n^*||^2\right| = o_{P_{Z^\infty}}(1)$.

Lemmas 3.1 and 5.2(1) indicate that the low level Assumption A in Appendix A is sufficient for both the original-sample LQA Assumption 3.2(i) and the bootstrap LQA Assumption 5.2(i).

Assumption 5.2(ii) can be easily verified by applying some central limit theorems. For example, if the weights are independent (Assumption Boot.1), we can use Lindeberg-Feller CLT; if the weights are multinomial (Assumption Boot.2) we can apply Hayek CLT (see Van der Vaart and Wellner (1996) p. 458). The next lemma provides some simple sufficient conditions for Assumption 5.2(ii).

Lemma 5.3. *Let $\{Z_i\}_{i=1}^n$ be i.i.d. and either Assumption Boot.1 or Assumption Boot.2 hold. If there is a positive real sequence $(b_n)_n$ such that $b_n = o(\sqrt{n})$ and*

$$\limsup_{n\to\infty} E\left[(g(X,u_n^*)\rho(Z,\alpha_0))^2\, 1\left\{\frac{(g(X,u_n^*)\rho(Z,\alpha_0))^2}{b_n} > 1\right\}\right] = 0. \tag{5.2}$$

Then: Assumptions 5.2(ii) and 3.2(ii) hold.

5.3 Bootstrap sieve Wald statistic

The following result is a bootstrap version of Theorem 4.1(2).

Theorem 5.1. *Let $\widehat{\alpha}_n$ be the PSMD estimator (2.2) and $\widehat{\alpha}_n^B$ the bootstrap PSMD estimator. Let conditions for Lemmas 2.2 and 5.1 hold. If Assumptions 3.1, 3.2(i), 4.1 and 5.2(i) hold, then:*

$$(1) \quad \sqrt{n}\frac{\phi(\widehat{\alpha}_n^B)-\phi(\widehat{\alpha}_n)}{\sigma_\omega ||\widehat{v}_n^*||_{n,sd}} = -\sqrt{n}\frac{\mathbb{Z}_n^{\omega-1}}{\sigma_\omega} + o_{P_{V^\infty|Z^\infty}}(1) \ wpa1(P_{Z^\infty});$$

(2) If further Assumptions 3.2(ii) and 5.2(ii) hold, then

$$\left|\mathcal{L}_{V^\infty|Z^\infty}\left(\sqrt{n}\frac{\phi(\widehat{\alpha}_n^B)-\phi(\widehat{\alpha}_n)}{\sigma_\omega ||\widehat{v}_n^*||_{n,sd}} \mid Z^n\right) - \mathcal{L}\left(\sqrt{n}\frac{\phi(\widehat{\alpha}_n)-\phi(\alpha_0)}{||\widehat{v}_n^*||_{n,sd}}\right)\right| = o_{P_{Z^\infty}}(1).$$

(3) If $\phi(\alpha_0)$ is regular, without imposing Assumption 4.1, we have:

$$\left|\mathcal{L}_{V^\infty|Z^\infty}\left(\sqrt{n}\frac{\phi(\widehat{\alpha}_n^B)-\phi(\widehat{\alpha}_n)}{\sigma_\omega} \mid Z^n\right) - \mathcal{L}\left(\sqrt{n}\left(\phi(\widehat{\alpha}_n)-\phi(\alpha_0)\right)\right)\right| = o_{P_{Z^\infty}}(1).$$

For a regular functional, Theorem 5.1(3) provides one way to construct its confidence sets without the need to compute any variance estimator. Unfortunately for an irregular functional, we need to compute a sieve variance estimator $||\widehat{v}_n^*||_{n,sd}$ to apply Theorem 5.1(2). In Appendix A we establish the consistency of a semiparametric score bootstrap, which does not require to recompute PSMD estimators using the bootstrap sample and hence is computationally simple.

5.4 Bootstrap SQLR statistic

If $\Sigma \neq \Sigma_0$, the SQLR statistic is no longer asymptotically chi-square; Theorem 3.2, however, implies that the SQLR statistic converges weakly to a tight limit. In this subsection we show that the asymptotic distribution of the SQLR can be consistently approximated by those of both the nonparametric and the weighted bootstrap SQLR statistics.

Recall that

$$\widehat{QLR}_n^B(\widehat{\phi}_n) \equiv n\left(\inf_{\{\mathcal{A}_{k(n)}:\ \phi(\alpha)=\widehat{\phi}_n\}} \widehat{Q}_n^B(\alpha) - \widehat{Q}_n^B(\widehat{\alpha}_n^B)\right),$$

with $\widehat{\phi}_n \equiv \phi(\widehat{\alpha}_n)$. Denote $\widehat{\alpha}_n^{R,B}$ as the *restricted* bootstrap PSMD estimator, i.e., the approximate minimizer of $\widehat{Q}_n^B(\alpha) + \lambda_n Pen(h)$ on $\{\alpha \in \mathcal{A}_{k(n)} \colon \phi(\alpha) = \widehat{\phi}_n\}$. By Lemma 5.1(2) and Remark 3.3 we can show that $\widehat{\alpha}_n^{R,B} \in \mathcal{N}_{osn}$ wpa1 under the null hypothesis of $\phi(\alpha_0) = \phi_0$.

Then: $\widehat{QLR}_n^B(\widehat{\phi}_n) = n\left(\widehat{Q}_n^B(\widehat{\alpha}_n^{R,B}) - \widehat{Q}_n^B(\widehat{\alpha}_n^B)\right)$.

Theorem 5.2. *Let $\widehat{\alpha}_n$ be the PSMD estimator (2.2) and $\widehat{\alpha}_n^B$ be the bootstrap PSMD estimator. Let conditions for Lemmas 2.2 and 5.1 hold. Let Assumptions 3.1, 3.2(i) and 5.2(i) hold with*

$\left|B_n^{\omega} - ||u_n^*||^2\right| = o_{P_{V^\infty|Z^\infty}}(1)$ *wpa*$1(P_{Z^\infty})$. *Then: under the null hypothesis of* $\phi(\alpha_0) = \phi_0$,

$$\text{(1)} \quad \frac{\widehat{QLR}_n^B(\widehat{\phi}_n)}{\sigma_\omega^2} = \left(\sqrt{n}\frac{\mathbb{Z}_n^{\omega-1}}{\sigma_\omega ||u_n^*||}\right)^2 + o_{P_{V^\infty|Z^\infty}}(1) \; wpa1(P_{Z^\infty});$$

(2) If further Assumptions 3.2(ii) and 5.2(ii) hold, then

$$\left|\mathcal{L}_{V^\infty|Z^\infty}\left(\frac{\widehat{QLR}_n^B(\widehat{\phi}_n)}{\sigma_\omega^2} \mid Z^n\right) - \mathcal{L}\left(\widehat{QLR}_n(\phi_0)\right)\right| = o_{P_{Z^\infty}}(1).$$

Theorem 5.2 extends the results in Theorems 3.2 and 4.2(1) to the bootstrap SQLR statistic. It allows us to construct valid confidence sets (CS) for $\phi(\alpha_0)$ based on inverting possibly *non*-optimally weighted SQLR statistic without the need to compute a variance estimator. See, e.g., Andrews and Buchinsky (2000) for a thorough discussion about how to construct CS via bootstrap.

For regular functionals of parametric time series models, Andrews (2002) and Camponovo (2012) establish the second order refinement of nonparametric bootstrap Wald and optimally weighted QLR statistics respectively. In this paper, we recommend the use of bootstrap (possibly non-optimally weighted) SQLR to construct CS for $\phi(\alpha_0)$ when it is difficult to compute any consistent variance estimator for $\phi(\widehat{\alpha})$, such as in the cases when the residual function $\rho(Z;\alpha)$ is pointwise non-smooth in α_0.

6 Simulation Studies and An Empirical Illustration

In this section, we present two small simulation studies and an empirical illustration of the PSMD estimation and SQLR based confidence sets for the NPQIV regression $E[1\{Y_1 \le h_0(Y_2)\} - \gamma|X] = 0$. We use the series LS estimator (2.4) of $m(X,h) = E[1\{Y_1 \le h(Y_2)\} - \gamma|X]$ in the computations.

6.1 Simulation Studies

We run Monte Carlo (MC) studies to assess the finite sample performance of our proposed inference procedures via a NPQIV model (2.8): $Y_1 = h_0(Y_2) + U$, $\Pr(U \le 0|X) = \gamma$. We consider two MC designs to ensure that the results are not sensitive to the specific simulation designs.

MC Design 1

Previously, Chen and Pouzo (2012a) and Chen and Pouzo (2009) designed MC studies to respectively investigate the finite sample performance of the PSMD estimator of $h_0(\cdot)$ in a NPQIV model $E[1\{Y_1 \le h_0(Y_2)\} - \gamma|X] = 0$ and the root-n asymptotic normality of the PSMD estimator of θ_0 in a partially linear quantile IV model $E[1\{Y_1 \le h_0(Y_2) + \theta_0' Y_3\} - \gamma|X] = 0$. Their MC designs were drawn from the British Family Expenditure Survey (FES) Engel curve data set that

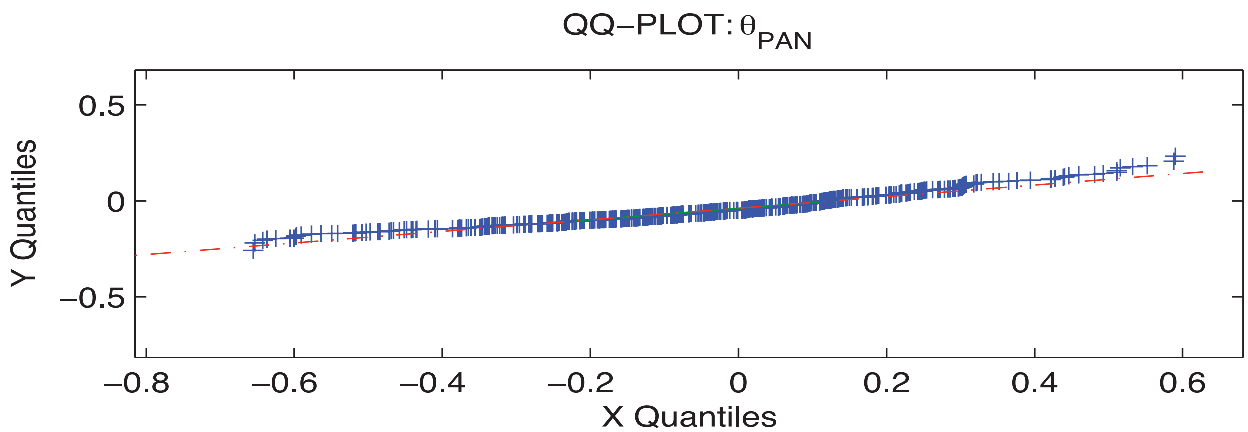

Figure 6.1: MC Design 1: QQ-Plot for $\nabla\widehat{h}(\mu_2)$ (appropriately centered and scaled), $n = 250$.

was first used in Blundell et al. (2007). Their simulation studies indicate remarkable finite sample performances of the PSMD estimator even for a difficult nonlinear, severely ill-posed inverse problem.

In the first MC design, we generate 1000 i.i.d. samples of $n = 250$ and 500 observations from a NPQIV model: $Y_1 = h_0(Y_2) + \sqrt{0.0025}U$, where $U = -\Phi^{-1}\left(\frac{E[h_0(Y_2)|X]-h_0(Y_2)}{25} + \gamma\right) + V$, $V \sim N(0,1)$, and $(Y_2, X) \sim N(\mu, \Sigma)$, where μ_2, μ_X and σ_2^2, σ_X^2 are set to be the sample estimates of the means and variances of Y_2, X from the "no-kids" subsample of British FES Engel curve data set of Blundell et al. (2007), and the correlation (in Σ) between Y_2 and X is set to be $\rho = 0.75$. Finally, $h_0(y_2) = \Phi\left(\frac{y_2-\mu_2}{\sigma_2}\right)$. The parameter of interest is: $\phi(h_0) = \nabla h_0(\mu_2)$.

We present the results for $\gamma = 0.5$. We estimate $h_0(\cdot)$ via the PSMD procedure, using a polynomial spline (P-spline) sieve $\mathcal{H}_{k(n)}$ with $k(n) = 6$, $Pen(h) = ||\nabla^2 h||^2_{L^2}$ with $\lambda_n = 0.0001$, and $p^{J_n}(X)$ is a P-Spline basis with $J_n = 15$. Figure 6.1 presents a QQ-plot for $\phi(\widehat{\alpha}_n) = \nabla\widehat{h}(\mu_2)$ to verify our asymptotic normality result. By inspecting this figure, the asymptotic normal approximation seems to be accurate even for a small sample size of $n = 250$. The QQ-plot corresponds to the larger sample size $n = 500$ is better so we omit it.

Table 6.1 reports the MC bias and standard deviation of the plug-in PSMD estimator $\phi(\widehat{\alpha}_n) = \nabla\widehat{h}(\mu_2)$ for both $n = 250$ and $n = 500$. The bias is an order of magnitude lower, reflecting the need to "undersmooth" since $\nabla h_0(\mu_2)$ is an irregular functional parameter.

	Bias	Std. Dev.
$n = 250$	0.066	0.236
$n = 500$	0.057	0.133

Table 6.1: MC Design 1: MC bias and standard deviation of the PSMD estimator for $\nabla h_0(\mu_2)$.

MC Design 2

The second simulation design is based on the NPIV model MC design of Newey and Powell

NS \ Different PSMD	(I)	(II)	(III)	(IV)
1%	1.1%	0.5%	1.1%	1.3%
5%	4.0%	4.2%	3.6%	5.3%
10%	10.8%	11.0%	8.5%	11.8%

Table 6.2: MC Design 2: Size of the SQLR test of $\phi(h_0) = 0$.

(2003) and Santos (2012), except that we consider the NPQIV model (2.8). Specifically, we generate 450 i.i.d. samples of $n = 750$ observations from the NPQIV model (2.8): $Y_2 = 2\sin(\pi Y_1) + 0.76U$, $U = 2(\Phi(U^*) - \gamma)$, $Y_1 = 2(\Phi(Y_1^*/3) - 0.5)$ and $X = 2(\Phi(X^*/3) - 0.5)$, where

$$\begin{bmatrix} Y_1^* \\ X^* \\ U^* \end{bmatrix} \sim N\left(0, \begin{bmatrix} 1 & 0.8 & 0.5 \\ 0.8 & 1 & 0 \\ 0.5 & 0 & 1 \end{bmatrix}\right),$$

and finally $h_0(Y_1) = 2\sin(\pi Y_1)$. The parameter of interest is: $\phi(h_0) = h_0(0)$.

We estimate $h_0(\cdot)$ via the PSMD procedure, using a polynomial spline (P-spline) sieve $\mathcal{H}_{k(n)}$ with $k(n) \in \{3, 4, 6\}$, $Pen(h) = ||h||_{L^2} + ||\nabla h||_{L^2}$ with $\lambda_n \in \{0.0001, 0.0002, 0.002\}$, and $p^{J_n}(X)$ is a Hermite polynomial basis with $J_n \in \{4, 6, 7\}$. We also considered other bases such as B-splines and results remained essentially the same.

Table 6.1 reports the simulated size of the SQLR test of $H_0\colon \phi(h_0) = 0$ as a function of the nominal size (NS), for different specifications of the tuning parameters. Column (I) corresponds to $k(n) = 4$, $J_n = 6$ and $\lambda_n = 0.0002$; Column (II) corresponds to $k(n) = 3$, $J_n = 4$ and $\lambda_n = 0.0001$; Column (III) corresponds to $k(n) = 6$, $J_n = 7$ and $\lambda_n = 0.0002$; Column (IV) corresponds to $k(n) = 6$, $J_n = 7$ and $\lambda_n = 0.002$. We see that for all cases, the simulated size is close to the NS.

We also compute the rejection probabilities of the null hypothesis as a function of $r \in \{2/\sqrt{n}, 4/\sqrt{n}\}$, where $r\colon \phi(0) = r$; these are respectively 33% and 88% corresponding to Column (I). We note that since our functional $\phi(h) = h(0)$ is estimated at a slower than root-n rate, the deviations considered for r are indeed "small".

We study the finite sample behavior of the nonparametric bootstrap SQLR corresponding to Column (I). We employ 450 nonparametric bootstrap evaluations, and 150 MC repetitions. We reduce the latter from 450 to 150 to save computation time. For nominal sizes of 10%, 5% and 1% we obtained a simulated p-value of 13%, 4% and 2% respectively. We expect that the performance will be much improved if we increase number of bootstrap runs.

6.2 An Empirical Application

We compute SQLR based confidence bands for nonparametric quantile IV Engel curves based on the British FES data set:

$$E[1\{Y_{1,i} \leq h_0(Y_{2,i})\} \mid X_i] = 0.5,$$

where $Y_{1,i}$ is the budget share of the $i-$th household on a particular non-durable goods, say food-in consumption; $Y_{2,i}$ is the log-total expenditure of the household, which is endogenous, and hence we use X_i, the gross earnings of the head of the household, to instrument it. We work with the "no kids" sub-sample of the data set of Blundell et al. (2007), which consists of $n = 628$ observations. See Blundell et al. (2007) for details about the data set.

We estimate $h_0(\cdot)$ via the optimally weighted PSMD procedure with $\widehat{\Sigma} = \Sigma_0 = 0.25$, using a polynomial spline (P-spline) sieve $\mathcal{H}_{k(n)}$ with $k(n) = 4$, $Pen(h) = ||h||_{L^2} + ||\nabla h||_{L^2}$ with $\lambda_n = 0.0005$, and $p^{J_n}(X)$ is a Hermite polynomial basis with $J_n = 6$. We also considered other bases such as P-splines and results remained essentially the same.

We use the fact that the optimally weighted SQLR of testing $\phi(h) = h(y_2)$ (for any fixed y_2) is asymptotically χ_1^2 to construct pointwise confidence bands. That is, for each y_2 in the sample we construct a grid of points for the SQLR test; each of these points where the value of SQLR test corresponding to $h(y_2) = r_i$ for $(r_i)_{i=1}^{30}$. We then, take the smallest interval that included all points r_i that yield a corresponding value of the SQLR test below the 95% percentile of χ_1^2.[8] Figure 6.2 presents the results, where the solid blue line is the point estimate and the red dashed lines are the 95% pointwise confidence bands. We can see that the confidence bands get wider towards the extremes of the sample, but are tight enough to reject the hypothesis that the food-in Engel curve is upward sloping or even constant.

7 Conclusion

In this paper, we provide unified asymptotic theories for estimation and inference on possibly irregular parameters $\phi(\alpha_0)$ of the general semi-nonparametric conditional moment restrictions $E[\rho(Z;\alpha_0)|X] = 0$. Under regularity conditions that allow for weakly dependent data and any consistent nonparametric estimator of the conditional mean function $m(X,\alpha) \equiv E[\rho(Z;\alpha)|X]$, we establish the asymptotic normality of the plug-in PSMD estimator $\phi(\widehat{\alpha}_n)$ of $\phi(\alpha_0)$, as well as the asymptotically tight distribution of a possibly non-optimally weighted SQLR statistic under the

[8]The grid $(r_i)_{i=1}^{n}$ was constructed to have $r_{15} = \widehat{h}_n(y_2)$, for all $i \leq 15$ $r_{i+1} \leq r_i \leq r_{15}$ decreasing in steps of length 0.002 (approx) and for all $i \geq 15$ $r_{i+1} \geq r_i \geq r_{15}$ increasing in steps of length 0.008 (approx); finally, the extremes, r_1 and r_{30}, were chosen so the SQLR test at those points was above the 95% percentile of χ_1^2. We tried different lengths and step sizes and the results remain qualitatively unchanged. For some observations, which only account for less than 4% of the sample, the confidence interval was degenerate at a point; this result is due to numerical approximation issues, and thus were excluded from the reported results.

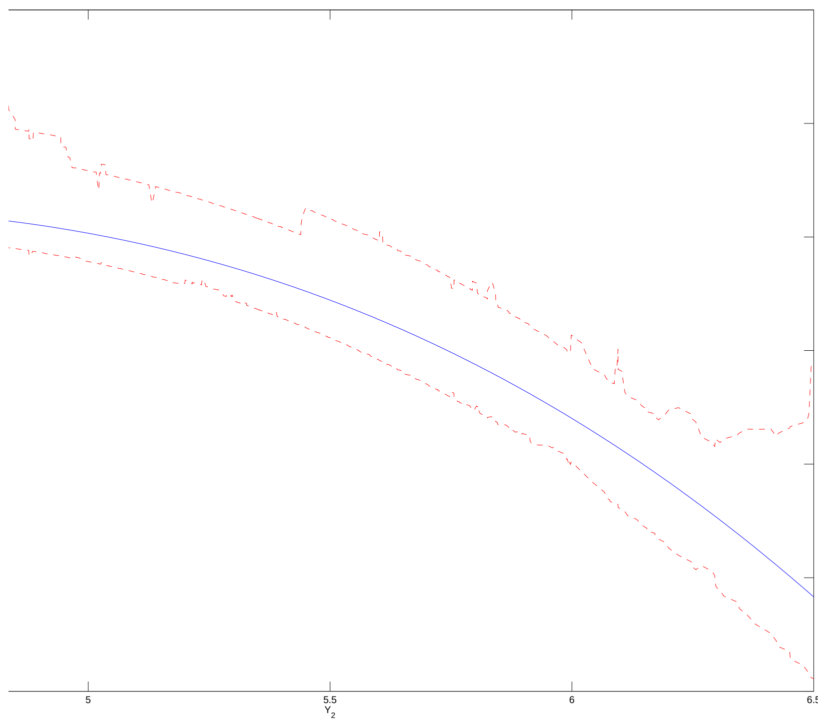

Figure 6.2: PSMD Estimate of the NPQIV food-in Engel curve (blue solid line), with the 95% pointwise confidence bands (red dash lines).

null hypothesis of $\phi(\alpha_0) = \phi_0$. As a simple yet useful by-product, we immediately obtain that an optimally weighted SQLR statistic is asymptotically chi-square distributed under the null hypothesis. For (pointwise) smooth residuals $\rho(Z; \alpha)$ (in α), we propose a simple consistent sieve variance estimator for $\phi(\widehat{\alpha}_n)$. Under i.i.d. data and conditions that are virtually the same as those for the limiting distributions of the original-sample (possibly non-optimally weighted) SQLR statistic, we establish the consistency of both the nonparametric and the weighted bootstrap (possibly non-optimally weighted) SQLR statistics. These results are valid regardless of whether $\phi(\alpha_0)$ is regular or not and whether $\rho(Z; \alpha)$ is pointwise smooth (in α) or not; and they allow applied researchers to construct confidence sets for $\phi(\alpha_0)$ without computing any consistent estimator of the asymptotic variance of $\phi(\widehat{\alpha}_n)$. Monte Carlo studies and an empirical illustration of a nonparametric quantile IV regression demonstrate the good finite sample performance of our SQLR based inference procedures.

In this paper we assume that the semi-nonparametric conditional moment restrictions $E[\rho(Z; \alpha_0)|X] = 0$ uniquely identifies the unknown true parameter value $\alpha_0 \equiv (\theta_0', h_0)$, and conduct inference that is robust to whether or not the semiparametric efficiency bound of $\phi(\alpha_0)$ is singular. Recently, Santos (2012) considered Bierens' type of test of the NPIV model $E[Y_2 - h_0(Y_1)|X] = 0$ without assuming point identification of $h_0(\cdot)$. In Chen et al. (2011) we are currently extending the SQLR inference procedure to allow for partial identification of the general model $E[\rho(Z; \alpha_0)|X] = 0$.

References

Ai, C. and X. Chen (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica 71*, 1795–1843.

Ai, C. and X. Chen (2007). Estimation of possibly misspecified semiparametric conditional moment restriction models with different conditioning variables. *Journal of Econometrics 141*(1), 5–43.

Ai, C. and X. Chen (2009). Semiparametric efficiency bound for models of sequential moment restrictions containing unknown functions. *Cowles Foundation Discussion Paper No. 1731*.

Andrews, D. (2002). Higher-order improvements of a computationally attractive k-step bootstrap for extremum estimators. *Econometrica 70*, 119–162.

Andrews, D. and M. Buchinsky (2000). A three-step method for choosing the number of bootstrap repetition. *Econometrica 68*, 23–51.

Bajari, P., H. Hong, and D. Nekipelov (2011). Game theory and econometrics: A survey of some recent research. *Working paper, Stanford and UC Berkeley*.

Billingsley, P. (1995). *Probability and Measure* (3rd ed.). Wiley.

Blundell, R., X. Chen, and D. Kristensen (2007). Semi-nonparametric iv estimation of shape invariant engel curves. *Econometrica 75*, 1613–1669.

Bontempts, C., D. Martimort, and A. Thomas (2012). Estimating optimal contracts under delegated management, application to the french water industry. *Working paper, Toulouse School of Economics*.

Camponovo, L. (2012). Nonparametric bootstrap for quasi-likelihood ratio tests. *Working paper, University of Lugano*.

Carrasco, M., J. Florens, and E. Renault (2007). Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization. *The Handbook of Econometrics, J.J. Heckman and E.E. Leamer (eds.), North-Holland, Amsterdam 6B.*

Chamberlain, G. (1992). Efficiency bounds for semiparametric regression. *Econometrica 60*(3), 567–592.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. *The Handbook of Econometrics, J.J. Heckman and E.E. Leamer (eds.), North-Holland, Amsterdam 6B.*

Chen, X., V. Chernozhukov, S. Lee, and W. Newey (2011). Local identification of nonparametric and semiparametric models. *Cowles Foundation Discussion Paper No. 1795. Yale University*.

Chen, X., Z. Liao, and X. Sun (2012). Sieve inference on semi-nonparametric time series models. *Cowles Foundation Discussion Paper No 1849. Yale University*.

Chen, X., O. Linton, and I. van Keilegom (2003). Estimation of semiparametric models with the criterion functions is not smooth. *Econometrica 71*, 1591–1608.

Chen, X. and S. Ludvigson (2009). Land of addicts? an empirical investigation of habit-based asset pricing models. *Journal of Applied Econometrics 24*, 1057–1093.

Chen, X. and D. Pouzo (2009). Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals. *Journal of Econometrics 152*, 46–60.

Chen, X. and D. Pouzo (2012a). Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals. *Econometrica 80*, 277–321.

Chen, X. and D. Pouzo (2012b). Supplement to 'estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals'. *Econometrica Supplementary Material*.

Chen, X., D. Pouzo, and E. Tamer (2011). Sieve qlr inference on partially identified semi-nonparametric conditional moment models. *Working paper, Yale, UC Berkeley and Northwestern*.

Chernozhukov, V. and C. Hansen (2005). An iv model of quantile treatment effects. *Econometrica 73*, 245–61.

Chernozhukov, V., G. Imbens, and W. Newey (2007). Instrumental variable estimation of nonseparable models. *Journal of Econometrics 139*, 4–14.

Darolles, S., Y. Fan, J. Florens, and E. Renault (2011). Nonparametric instrumental regression. *Econometrica 79*, 1541–1566.

Fan, J., C. Zhang, and J. Zhang (2001). Generalized likelihood ratio statistics and wilks phenomenon. *The Annals of Statistics 29*, 640–652.

Feller, W. (1970). *An Introduction to Probability Theory and its Applications* (3rd ed.), Volume I. Wiley.

Gagliardini, P. and O. Scaillet (2011). Nonparametric instrumental variable estimation of quantile structural effects. *Forthcoming in Econometrica*.

Gallant, R. and G. Tauchen (1989). Semiparametric estimation of conditional constrained heterogenous processes: Asset pricing applications. *Econometrica 57*, 1091–1120.

Graham, B. and J. Powell (2012). Identification and estimation of "irregular" correlated random coefficient models. *Econometrica, forthcoming*.

Hall, P. and J. Horowitz (2005). Nonparametric methods for inference in the presence of instrumental variables. *The Annals of Statistics 33*, 2904–2929.

Hansen, L. and S. Richard (1987). The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models. *Econometrica 55*, 587–613.

Hong, H., A. Mahajan, and D. Nekipelov (2010). Extremum estimation and numerical derivatives. *Mimeo, Dept. of Economics, UC at Berkeley*.

Horowitz, J. (2007). Asymptotic normality of a nonparametric instrumental variables estimator. *International Economic Review 48*, 1329–1349.

Horowitz, J. (2011). Applied nonparametric instrumental variables estimation. *Econometrica 79*, 347–394.

Horowitz, J. and S. Lee (2007). Nonparametric instrumental variables estimation of quantile regression model. *Econometrica 75*, 1191–1209.

Horowitz, J. and S. Lee (2011). Uniform confidence bands for functions estimated nonparametrically with instrumental variables. *Journal of Econometrics, forthcoming*.

Huang, J. (1998). Projection estimation in multiple regression with application to functional anova models. *The Annals of Statistics 27*, 242–272.

Liao, Y. and W. Jiang (2012). Posterior consistency of nonparametric conditional moment restricted models. *Forthcoming in Annals of Statistics*.

Murphy, S. and A. V. der Vaart (2000). On profile likelihood. *Journal of American Statistical Association 95*, 449–465.

Newey, W. (1997). Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics 79*, 147–168.

Newey, W. and J. Powell (2003). Instrumental variables estimation for nonparametric models. *Econometrica 71*, 1565–1578.

Otsu, T. (2011). Empirical likelihood estimation of conditional moment restriction models with unknown functions. *Econometric Theory 27*, 8–46.

Preastgaard, J. (1991). *General-Weights Bootstrap of the Empirical Process*. Ph. D. thesis, University of Washington.

Santos, A. (2012). Inference in nonparametric instrumental variables with partial identification. *Econometrica 80*, 213–275.

Shen, X. and J. Shi (2005). Sieve likelihood ratio inference on general parameter space. *Science in China 48*, 67–78.

Van der Vaart, A. and J. Wellner (1996). *Weak Convergence and Empirical Processes*. Springer.

A Sufficient Conditions and Additional Results

In Appendix A we first provide some low level sufficient conditions for the high level LQA assumption 3.2(i) and the bootstrap LQA assumption 5.2(i). We then state useful lemmas when the conditional mean function $m(\cdot,\alpha)$ is estimated by series LS estimators. We finally present a consistency theorem for a computationally attractive sieve score bootstrap. See Appendix C for the proofs of all the results in this Appendix.

Throughout Appendix A, we maintain the assumptions that the original-sample $\{Z_i=(Y_i',X_i')'\}_{i=1}^n$ is i.i.d., $\mathcal{A}\equiv\Theta\times\mathcal{H}$, Θ is a compact subset of $\mathbb{R}^{d_\theta}$, $\mathcal{H}\subseteq\mathbf{H}$, $\mathbf{H}$ is a separable Banach space under a norm $\|\cdot\|_{\mathbf{H}}$. $\widehat{m}(x,\alpha)$ is the series LS estimator (2.4) and $\widehat{m}^B(x,\alpha)$ is the bootstrap series LS estimator (5.1) of $m(x,\alpha)$.

A.1 Sufficient conditions for LQA(i) and Bootstrap LQA(i)

Assumption A.1. *(i) $\mathcal{X}$ is a compact connected subset of $\mathbb{R}^{d_x}$ with Lipschitz continuous boundary, and f_X is bounded and bounded away from zero over $\mathcal{X}$; (ii) The smallest and largest eigenvalues of $E[p^{J_n}(X)p^{J_n}(X)']$ are bounded and bounded away from zero for all J_n; (iii) $\sup_{x\in\mathcal{X}}|p_j(x)|\leq const.<\infty$ for all $j=1,...,J_n$; Either $J_n^2=o(n)$ or $J_n\log(J_n)=o(n)$ for $p^{J_n}(X)$ a polynomial spline sieve; (iv) There is $p^{J_n}(X)'\pi$ such that $\sup_x|g(x)-p^{J_n}(x)'\pi|=O(b_{m,J_n})=o(1)$ uniformly in $g\in\{m(\cdot,\alpha):\alpha\in\mathcal{A}_{k(n)}^{M_0}\}$.*

Let $\mathcal{O}_{on}\equiv\{\rho(\cdot,\alpha)-\rho(\cdot,\alpha_0):\alpha\in\mathcal{N}_{osn}\}$. Denote

$$1\leq\sqrt{C_n}\equiv\int_0^1\sqrt{1+\log(N_{[]}(w(M_n\delta_{s,n})^\kappa,\mathcal{O}_{on},||\cdot||_{L^2(f_Z)}))}dw<\infty.$$

Assumption A.2. *(i) There is a sequence $\{\bar{\rho}_n(Z)\}_n$ of measurable functions such that $\sup_{\mathcal{A}_{k(n)}^{M_0}}|\rho(Z,\alpha)|\leq\bar{\rho}_n(Z)$ a.s.-Z and $E[|\bar{\rho}_n(Z)|^2|X]\leq const.<\infty$; (ii) there exist some $\kappa\in(0,1]$ and $K\colon\mathcal{X}\to\mathbb{R}$ measurable with $E[|K(X)|^2]\leq const.$ such that $\forall\delta>0$,*

$$E\left[\sup_{\alpha\in\mathcal{N}_{0sn}:\,||\alpha-\alpha'||_s\leq\delta}\left\|\rho(z,\alpha)-\rho(z,\alpha')\right\|_e^2|X\right]\leq K(X)^2\delta^{2\kappa}\ \forall\alpha'\in\mathcal{N}_{osn}\cup\{\alpha_0\},$$

and $\max\left\{(M_n\delta_n)^2,(M_n\delta_{s,n})^{2\kappa}\right\}=(M_n\delta_{s,n})^{2\kappa}$; (iii) $n\delta_n^2(M_n\delta_{s,n})^\kappa\sqrt{C_n}\max\left\{(M_n\delta_{s,n})^\kappa\sqrt{C_n},M_n\right\}=o(1)$; (iv) $\sup_{\mathcal{X}}||\widehat{\Sigma}(x)-\Sigma(x)||\times(M_n\delta_n)=o_{P_{Z^\infty}}(n^{-1/2})$; $\delta_n\asymp\sqrt{\frac{J_n}{n}}=\max\{\sqrt{\frac{J_n}{n}},b_{m,J_n}\}=o(n^{-1/4})$.

Let $\widetilde{m}(X,\alpha)\equiv p^{J_n}(X)'(P'P)^-\sum_{i=1}^n p^{J_n}(X_i)m(X_i,\alpha)$ be the LS projection of $m(X,\alpha)$ onto $p^{J_n}(X)$, and let $g(X,u_n^*)\equiv\{\frac{dm(X,\alpha_0)}{d\alpha}[u_n^*]\}'\Sigma(X)^{-1}$ and $\widetilde{g}(X,u_n^*)$ be its LS projection onto $p^{J_n}(X)$.

Assumption A.3. *(i) $E_{P_{Z^\infty}}\left[\left\|\frac{d\widetilde{m}(X,\alpha_0)}{d\alpha}[u_n^*]-\frac{dm(X,\alpha_0)}{d\alpha}[u_n^*]\right\|_e^2\right](M_n\delta_n)^2=o(n^{-1})$;*

(ii) $E_{P_{Z^\infty}}\left[\|\widetilde{g}(X,u_n^*)-g(X,u_n^*)\|_e^2\right](M_n\delta_n)^2=o(n^{-1})$; (iii) $\{m(\cdot,\alpha):\alpha\in\mathcal{N}_{osn}\}$ and $\{g(\cdot,u_n^*)m(\cdot,\alpha):\alpha\in\mathcal{N}_{osn}\}$ are $L^2(f_X)$-Donsker; (iv) $E[||g(X,u_n^*)\{m(X,\alpha)-m(X,\alpha_0)\}||_e^2]=o(1)$ for all $\alpha\in\mathcal{A}_{k(n)}$ such that $||\alpha-\alpha_0||_s=o(1)$.

Assumption A.4. *(i) $m(X,\alpha)$ is twice continuously pathwise differentiable in $\alpha\in\mathcal{N}_{os}$, a.s.-$X$;*

$$(ii)\quad E\left[\sup_{\alpha\in\mathcal{N}_{osn}}\left\|\frac{dm(X,\alpha)}{d\alpha}[u_n^*]-\frac{dm(X,\alpha_0)}{d\alpha}[u_n^*]\right\|_e^2\right]\times(M_n\delta_n)^2=o(n^{-1});$$

(iii) $E\left[\sup_{\alpha\in\mathcal{N}_{osn}}\left\|\frac{d^2m(X,\alpha)}{d\alpha^2}[u_n^*,u_n^*]\right\|_e^2\right]\times(M_n\delta_n)^2=o(1)$; *(iv) Uniformly over* $\alpha_1\in\mathcal{N}_{os}$ *and* $\alpha_2\in\mathcal{N}_{osn}$,

$$E\left[g(X,u_n^*)\left(\frac{dm(X,\alpha_1)}{d\alpha}[\alpha_2-\alpha_0]-\frac{dm(X,\alpha_0)}{d\alpha}[\alpha_2-\alpha_0]\right)\right]=o(n^{-1/2}).$$

Assumption A.2 is comparable to that imposed in Chen and Pouzo (2009) for a non-smooth residual function $\rho(Z,\alpha)$. This assumption ensures that the sample criterion function $\widehat{Q}_n$ is well approximated by a "smooth" version of it. Assumptions A.3 and A.4 are similar to those imposed in Ai and Chen (2003), Ai and Chen (2007) and Chen and Pouzo (2009), except that we use the scaled sieve Riesz representer $u_n^*\equiv v_n^*/\left\|v_n^*\right\|_{sd}$. This is because we allow for possibly irregular functionals (i.e., possibly $\|v_n^*\|\to\infty$), while the above mentioned papers only consider regular functionals (i.e., $\|v_n^*\|\to\|v^*\|<\infty$). We refer readers to these papers for detailed discussions and verifications of these assumptions in examples of the general model (1.3).

A.2 Lemmas for Series LS estimation of $m(x,\alpha)$

The next lemma (Lemma A.1) extends Lemma C.3 of Chen and Pouzo (2012a) and Lemma A.1 of Chen and Pouzo (2009) to the bootstrap case. Denote

$$\ell_n(x,\alpha)\equiv\widetilde{m}(x,\alpha)+\widehat{m}(x,\alpha_0)\quad\text{and}\quad\ell_n^B(x,\alpha)\equiv\widetilde{m}(x,\alpha)+\widehat{m}^B(x,\alpha_0).$$

Lemma A.1. *Let $\widehat{m}^B(\cdot,\alpha)$ be the bootstrap series LS estimator (5.1). Let Assumptions 2.1(iv), 2.4, 4.1(iii), A.1, A.2(i)(ii), and Boot.1 or Boot.2 hold. Then: (1) For all $\delta>0$, there is a $M(\delta)>0$ such that for all $M\geq M(\delta)$,*

$$P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\sup_{\alpha\in\mathcal{N}_{osn}}\frac{\tau_n}{n}\sum_{i=1}^n\left\|\widehat{m}^B(X_i,\alpha)-\ell_n^B(X_i,\alpha)\right\|_e^2\geq M\mid Z^n\right)\geq\delta\right)<\delta$$

eventually, with $\tau_n^{-1}\equiv(\delta_n)^2\left(M_n\delta_{s,n}\right)^{2\kappa}C_n$.

(2) For all $\delta>0$, there is a $M(\delta)>0$ such that for all $M\geq M(\delta)$,

$$P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\sup_{\alpha\in\mathcal{N}_{osn}}\frac{\tau_n'}{n}\sum_{i=1}^n\left\|\ell_n^B(X_i,\alpha)\right\|_e^2\geq M\mid Z^n\right)\geq\delta\right)<\delta$$

eventually, with

$$(\tau_n')^{-1}=\max\{\frac{J_n}{n},b_{m,J_n}^2,(M_n\delta_n)^2\}=const.\times(M_n\delta_n)^2.$$

(3) Let Assumption A.2(iii) hold. For all $\delta>0$, there is $N(\delta)$ such that, for all $n\geq N(\delta)$,

$$P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\sup_{\mathcal{N}_{osn}}\frac{s_n}{n}\left|\sum_{i=1}^n\left\|\widehat{m}^B\left(X_i,\alpha\right)\right\|_{\widehat{\Sigma}^{-1}}^2-\sum_{i=1}^n\left\|\ell_n^B\left(X_i,\alpha\right)\right\|_{\widehat{\Sigma}^{-1}}^2\right|\geq\delta\mid Z^n\right)\geq\delta\right)<\delta$$

with

$$s_n^{-1}\leq(\delta_n)^2(M_n\delta_{s,n})^\kappa\sqrt{C_n}\max\left\{(M_n\delta_{s,n})^\kappa\sqrt{C_n},M_n\right\}L_n=o(n^{-1}),$$

where $\{L_n\}_{n=1}^\infty$ is a slowly divergent sequence of positive real numbers (such a choice of L_n exists under assumption A.2(iii)).

Recall that

$$\mathbb{Z}_n^{\omega} = \frac{1}{n}\sum_{i=1}^{n}\omega_{i,n}\left(\frac{dm(X_i,\alpha_0)}{d\alpha}[u_n^*]\right)'\Sigma(X_i)^{-1}\rho(Z_i,\alpha_0) = \frac{1}{n}\sum_{i=1}^{n} g(X_i,u_n^*)\omega_{i,n}\rho(Z_i,\alpha_0).$$

Lemma A.2. *Let all the conditions for Lemma A.1(2) hold. If Assumptions A.2(iv), A.3 and A.4(i)(ii)(iv) hold, then: for all $\delta>0$, there is a $N(\delta)$ such that for all $n\geq N(\delta)$,*

$$P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\sup_{\mathcal{N}_{osn}}\sqrt{n}\left|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{d\widetilde{m}(X_i,\alpha)}{d\alpha}[u_n^*]\right)'(\widehat{\Sigma}(X_i))^{-1}\ell_n^B(X_i,\alpha)-\{\mathbb{Z}_n^{\omega}+\langle u_n^*,\alpha-\alpha_0\rangle\}\right|\geq\delta\mid Z^n\right)\geq\delta\right)$$
$< \delta$.

Lemma A.3. *Let all the conditions for Lemma A.1(2) hold. If Assumption A.4(i)(iii) holds, then: for all $\delta>0$, there is a $N(\delta)$ such that for all $n\geq N(\delta)$,*

$$P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\sup_{\mathcal{N}_{osn}} n^{-1}\sum_{i=1}^{n}\left(\frac{d^2\widetilde{m}(X_i,\alpha)}{d\alpha^2}[u_n^*,u_n^*]\right)'(\widehat{\Sigma}(X_i))^{-1}\ell_n^B(X_i,\alpha)\geq\delta\mid Z^n\right)\geq\delta\right)<\delta.$$

Lemma A.4. *Let Assumptions 2.1(iv), 2.4(i), 4.1(iii), A.1, A.3(i), A.4(ii) hold. Then: (1) For all $\delta>0$ there is a $M(\delta)>0$, such that for all $M\geq M(\delta)$,*

$$P_{Z^\infty}\left(\sup_{\mathcal{N}_{osn}}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{d\widetilde{m}(X_i,\alpha)}{d\alpha}[u_n^*]\right)'\widehat{\Sigma}^{-1}(X_i)\left(\frac{d\widetilde{m}(X_i,\alpha)}{d\alpha}[u_n^*]\right)\geq M\right)<\delta$$

eventually.

(2) If in addition, Assumption B holds, then: For all $\delta>0$, there is a $N(\delta)$ such that for all $n\geq N(\delta)$,

$$P_{Z^\infty}\left(\sup_{\mathcal{N}_{osn}}\left|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{d\widetilde{m}(X_i,\alpha)}{d\alpha}[u_n^*]\right)'\widehat{\Sigma}^{-1}(X_i)\left(\frac{d\widetilde{m}(X_i,\alpha)}{d\alpha}[u_n^*]\right)-||u_n^*||^2\right|\geq\delta\right)<\delta.$$

A.3 Sieve score bootstrap

In the main text we present the consistency of bootstrap SQLR statistic and bootstrap sieve Wald statistic. Here we consider a sieve score bootstrap, which does not require to recompute PSMD estimators of α_0 using the bootstrap sample and hence is computationally attractive.

Recall that $\widehat{\alpha}_n^R$ is the original-sample restricted PSMD estimator (3.17). Let $\widetilde{v}_n^*$ be computed in the same way as that in Subsection 4.1, except that we use $\widehat{\alpha}_n^R$ instead of $\widehat{\alpha}_n$. Denote

$$T_n \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n^R)}{d\alpha}[\widetilde{v}_n^*/||\widetilde{v}_n^*||_{n,sd}]\right)'\widehat{\Sigma}^{-1}(X_i)\widehat{m}(X_i,\widehat{\alpha}_n^R)$$

and

$$T_n^B \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n^R)}{d\alpha}[\widetilde{v}_n^*/||\widetilde{v}_n^*||_{n,sd}]\right)'\widehat{\Sigma}^{-1}(X_i)\{\widehat{m}^B(X_i,\widehat{\alpha}_n^R)-\widehat{m}(X_i,\widehat{\alpha}_n^R)\}.$$

For series LS estimators of $m(x,\alpha)$, we have: $\widehat{m}^B(x,\widehat{\alpha}_n^R)-\widehat{m}(x,\widehat{\alpha}_n^R)=p^{J_n}(x)'(P'P)^{-}\sum_{j=1}^{n}p^{J_n}(X_j)(\omega_j-1)\rho(Z_j,\widehat{\alpha}_n^R)$. This is not crucial, but it simplifies both the proof and the implementation.

Let $\{\epsilon_n\}_{n=1}^{\infty}$ and $\{\zeta_n\}_{n=1}^{\infty}$ be real valued positive sequences such that $\epsilon_n=o(1)$ and $\zeta_n=o(1)$.

Assumption C. *(i)* $\max\{\epsilon_n, n^{-1/4}\}M_n\delta_n = o(n^{-1/2})$

$$\sup_{\mathcal{N}_{osn}} \sup_{u\in\overline{\mathbf{V}}_n:\,||u||=1} n^{-1}\sum_{i=1}^{n}\left\|\frac{d\widehat{m}(X_i,\alpha)}{d\alpha}[u]-\frac{dm(X_i,\alpha)}{d\alpha}[u]\right\|_e^2 = O_{P_{Z^\infty}}(\max\{n^{-1/2},\epsilon_n^2\});$$

(ii) there is a continuous mapping $\Upsilon:$ *such that* $\max\{\Upsilon(\zeta_n), n^{-1/4}\}M_n\delta_n = o(n^{-1/2})$ *and*

$$\sup_{\mathcal{N}_{osn}} \sup_{\overline{\mathbf{V}}_n:\,||u_n^*-u||\le\zeta_n} n^{-1}\sum_{i=1}^{n}\left\|\frac{dm(X_i,\alpha)}{d\alpha}[u_n^*]-\frac{dm(X_i,\alpha)}{d\alpha}[u]\right\|_e^2 = O_{P_{Z^\infty}}(\max\{n^{-1/2},(\Upsilon(\zeta_n))^2\});$$

(iii) $||\widetilde{u}_n^* - u_n^*|| = O_{P_{Z^\infty}}(\zeta_n)$ *where* $\widetilde{u}_n^* \equiv \widetilde{v}_n^*/||\widetilde{v}_n^*||_{sd}$.

Assumption C(i) can be obtained by similar conditions to those imposed in Ai and Chen (2003). Assumption C(ii) can be established by $\left\{\left\|\frac{dm(\cdot,\alpha)}{d\alpha}[u]-\frac{dm(\cdot,\alpha)}{d\alpha}[u_n^*]\right\|_e^2 : \alpha\in\mathcal{N}_{osn};\ u\in\{\overline{\mathbf{V}}_n\colon ||u-u_n^*||\le\zeta_n\}\right\}$ being a Donsker class and $E\left[\left\|\frac{dm(X,\alpha)}{d\alpha}[u_n^*]-\frac{dm(X,\alpha)}{d\alpha}[u]\right\|_e^2\right] = o(1)$ for all $||u_n^*-u||<\zeta_n$. However, it can be obtained by weaker conditions, yielding a $(\Upsilon(\zeta_n))^2$ that is slower than $O(n^{-1/2})$ provided that $\Upsilon(\zeta_n)M_n\delta_n = o(n^{-1/2})$. In the proof we show that $||\widetilde{u}_n^*-u_n^*|| = o_{P_{Z^\infty}}(1)$; faster rates of convergence will relax the conditions needed to show part (ii).

Theorem A.1. *Let* $\widehat{\alpha}_n^R$ *be the restricted PSMD estimator (3.17) and conditions for Lemmas 2.2 and 5.1 hold. Let Assumptions 3.1, A, Boot.1 or Boot.2, 4.1, C, 3.2(ii) and 5.2(ii) hold and that* $n\delta_n^2\left(M_n\delta_{s,n}\right)^{2\kappa}C_n = o(1)$. *Then, under the null hypothesis of* $\phi(\alpha_0)=\phi_0$,

$$|\mathcal{L}_{V^\infty|Z^\infty}(\sigma_\omega^{-1}T_n^B \mid Z^n) - \mathcal{L}(T_n)| = o_{P_{Z^\infty}}(1).$$

Appendices B and C are available upon request.