SIEVE WALD AND QLR INFERENCES ON SEMI/NONPARAMETRIC CONDITIONAL MOMENT MODELS

By

Xiaohong Chen and Demian Pouzo

May 2013
Revised April 2014

COWLES FOUNDATION DISCUSSION PAPER NO. 1897R

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

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

Sieve Wald and QLR Inferences on Semi/nonparametric Conditional Moment Models[1]

Xiaohong Chen[2] and Demian Pouzo[3]

First version: March 2009. This version: January 26, 2014

Abstract

This paper considers inference on functionals of semi/nonparametric conditional moment restrictions with possibly nonsmooth generalized residuals, which include all of the (nonlinear) nonparametric instrumental variables (IV) as special cases. For these models it is often difficult to verify whether a functional is regular (i.e., root-n estimable) or irregular (i.e., slower than root-n estimable). We provide computationally simple, unified inference procedures that are asymptotically valid regardless of whether a functional is regular or not. We establish the following new useful results: (1) the asymptotic normality of a plug-in penalized sieve minimum distance (PSMD) estimator of a (possibly irregular) functional; (2) the consistency of simple sieve variance estimators of the plug-in PSMD estimator, and hence the asymptotic chi-square distribution of the sieve Wald statistic; (3) the asymptotic chi-square distribution of an optimally weighted sieve quasi likelihood ratio (QLR) test under the null hypothesis; (4) the asymptotic tight distribution of a non-optimally weighted sieve QLR statistic under the null; (5) the consistency of generalized residual bootstrap sieve Wald and QLR tests; (6) local power properties of sieve Wald and QLR tests and of their bootstrap versions; (7) Wilks phenomenon of the sieve QLR test of hypothesis with increasing dimension. 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 variance estimators; Sieve Wald; Sieve quasi likelihood ratio; Generalized residual bootstrap; Local power; Wilks phenomenon.

[1]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, and econometric workshops in numerous universities. We thank a co-editor, two referees, D. Andrews, P. Bickel, G. Chamberlain, T. Christensen, M. Jansson, 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: d-pouzo@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

$$E[\rho(Y, X; \theta_0, h_0)|X] = 0 \quad a.s. - X, \tag{1.1}$$

where Y is a vector of endogenous variables and X is a vector of conditioning (or instrumental) variables. The conditional distribution of Y given X, $F_{Y|X}$, 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 $h_0 \equiv (h_{01}(\cdot), ..., h_{0q}(\cdot)) \in \mathcal{H}$ 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 and Y. The residual function $\rho(\cdot; \alpha)$ could be nonlinear and pointwise non-smooth 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 and finance. Well known examples include nonparametric mean instrumental variables regressions (NPIV): $E[Y_1 - h_0(Y_2)|X] = 0$ (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): $E[1\{Y_1 \leq h_0(Y_2)\} - \gamma|X] = 0$ (e.g., Chernozhukov and Hansen (2005), Chernozhukov et al. (2007), Horowitz and Lee (2007), Chen and Pouzo (2012a), Gagliardini and Scaillet (2012)); semi/nonparametric demand models with endogeneity (e.g., Blundell et al. (2007), Chen and Pouzo (2009), Souza-Rodrigues (2012)); semi/nonparametric random coefficient panel data regressions (e.g., Chamberlain (1992), Graham and Powell (2012)); semi/nonparametric spatial models with endogeneity (e.g., Pinkse et al. (2002), Merlo and de Paula (2013)); semi/nonparametric asset pricing models (e.g., Hansen and Richard (1987), Gallant and Tauchen (1989), Chen and Ludvigson (2009), Chen et al. (2013), Penaranda and Sentana (2013)); semi/nonparametric static and dynamic game models (e.g., Bajari et al. (2011)); nonparametric optimal endogenous contract models (e.g., Bontemps and Martimort (2013)). 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. (2013) and the references therein. In fact, model (1.1) includes all of the (nonlinear) semi/nonparametric IV regressions when the unknown functions h_0 depend on the endogenous variables Y:

$$E[\rho(Y_1; \theta_0, h_0(Y_2))|X] = 0 \quad a.s. - X, \tag{1.2}$$

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 random sample from the distribution of $Z \equiv (Y', X')'$ that satisfies the conditional moment restrictions (1.1) with a unique $\alpha_0 \equiv (\theta_0', h_0)$. Let $\phi : \Theta \times \mathcal{H} \to \mathbb{R}^{d_\phi}$ be a functional with a finite $d_\phi \geq 1$. Typical functionals include an Euclidean functional $\phi(\alpha) = \theta$, a (point) evaluation functional $\phi(\alpha) = h(\overline{y}_2)$ (for $\overline{y}_2 \in \text{supp}(Y_2)$), a weighted derivative functional $\phi(h) = \int w(y_2) \nabla h(y_2) dy_2$ or a quadratic functional $\int w(y_2) |h(y_2)|^2 dy_2$ (for a known positive weight $w(\cdot)$) and many others. We are interested in computationally simple, valid inferences on any $\phi(\alpha_0)$ of the general model (1.1) with i.i.d. data.[4]

Although some functionals of the model (1.1), such as the (point) evaluation functional, are known *a priori* to be estimated at slower than root-n rates, others, such as the weighted derivative functional, are far less clear without a stare at their semiparametric efficiency bound expressions. This is because a non-singular semiparametric efficiency bound is a necessary condition for $\phi(\alpha_0)$ to be root-n estimable. Unfortunately, as pointed out in Chamberlain (1992) and Ai and Chen (2012), there is generally no closed form solution for the semiparametric efficiency bound of $\phi(\alpha_0)$ (including θ_0) of model (1.1), especially so when $\rho(\cdot; \theta_0, h_0)$ contains several unknown functions and/or when the unknown functions h_0 of endogenous variables enter $\rho(\cdot; \theta_0, h_0)$ nonlinearly. It is thus difficult to verify whether the semiparametric efficiency bound for $\phi(\alpha_0)$ is singular or not. Therefore, it is highly desirable for applied researchers to be able to conduct simple valid inferences 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 (1.1) 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 (1.1) with (pointwise) smooth residuals $\rho(Z; \alpha)$ in α_0, we propose two simple sieve variance estimators for possibly slower than root-n estimator $\phi(\widehat{\alpha}_n)$, which immediately leads to the asymptotic chi-square distribution of the sieve Wald statistic. 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 under the null regardless of whether $\phi(\alpha_0)$ is root-n estimable or not. In addition, we establish the consistency of the generalized residual bootstrap (possibly

[4]See our Cowles Foundation Discussion Paper No. 1897 for general theory allowing for weakly dependent data.

non-optimally weighted) SQLR and sieve Wald tests under virtually the same conditions as those used to derive the limiting distributions of the original-sample statistics. 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 ease notation burden, we present the above listed theoretical results for a scalar-valued functional in the main text. In Appendix A we present the asymptotic properties of sieve Wald and SQLR for functionals of increasing dimension (i.e., $d_\phi = dim(\phi)$ could grow with sample size n), and establish the Wilks phenomenon of the SQLR test on hypothesis with increasing dimension. We also provide the local power properties of sieve Wald and SQLR tests as well as their bootstrap versions in Appendix A. Regardless of whether a possibly nonlinear functional $\phi(\alpha_0)$ is root-n estimable or not, we show that the optimally weighted SQLR is more powerful than the non-optimally weighed SQLR, and that the SQLR and the sieve Wald using the same weighting matrix have the same local power in terms of first order asymptotic theory.

To the best of our knowledge, our paper is the first to provide a unified theory about sieve Wald and SQLR inferences on any $\phi(\alpha_0)$ satisfying the general semi/nonparametric model (1.1) with possibly non-smooth residuals.[5] 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 paper is also the first to provide local power properties of sieve Wald and SQLR tests of general nonlinear hypotheses for semi/nonparametric model (1.1).

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 (1.1). See, e.g., Newey and Powell (2003) and Chen et al. (2013) for identification; Newey and Powell (2003), Chernozhukov et al. (2007), Chen and Pouzo (2012a) and Liao and Jiang (2011) for consistency of their respective estimators; and Chen and Pouzo (2012a) for the rate of convergence of the PSMD estimator of the nonparametric h_0. In particular, under virtually the same conditions as those in Chen and Pouzo (2012a), we show that our generalized residual bootstrap PSMD estimator of α_0 is consistent and achieves the same convergence rate as that of the original-sample PSMD estimator $\widehat{\alpha}_n \equiv (\widehat{\theta}_n', \widehat{h}_n)$. This result is then used to establish the consistency of the bootstrap sieve Wald (and the bootstrap SQLR) 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 (1.1) when θ_0 is assumed to be regular. See Ai and Chen (2003), Chen and Pouzo (2009)

[5] We also provide asymptotic properties of sieve score and bootstrap sieve score statistics in online Appendix D.

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 non-smooth residuals). However, none of these papers allows for irregular θ_0. When specializing our general theory to inference on a regular θ_0 of the model (1.1), we not only recover the results of Ai and Chen (2003) and Chen and Pouzo (2009), but also provide local power properties of sieve Wald and SQLR as well as valid bootstrap (possibly non-optimally weighted) SQLR inference. Moreover, our results remain valid even when θ_0 might be irregular.[6]

When specializing our theory to inference on a particular irregular functional, the evaluation functional $\phi(\alpha) = h(\overline{y}_2)$, 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 (2012) 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 (1.1) 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 (2012) informed us that they were independently and concurrently working on confidence bands for a particular SMD estimator of the NPIV example.

The rest of the paper is organized as follows. Section 2 presents the plug-in PSMD estimator $\phi(\widehat{\alpha}_n)$ of a functional ϕ evaluated at $\alpha_0 \equiv (\theta_0', h_0)$ satisfying the model (1.1). It also provides an overview of the main asymptotic results that will be established in the subsequent sections, and illustrates the applications through a point evaluation functional $\phi(\alpha) = h(\overline{y}_2)$, a weighted derivative functional $\phi(h) = \int w(y_2) \nabla h(y_2) dy_2$, and a quadratic functional $\phi(\alpha) = \int w(y_2) \left| h(y_2) \right|^2 dy_2$ of the NPIV and NPQIV examples. Section 3 establishes the asymptotic normality of $\phi(\widehat{\alpha}_n)$, and the tight asymptotic null distribution of a possibly non-optimally weighted SQLR statistic. It also verifies the key regularity conditions for the asymptotic properties via the three functionals of the NPIV and NPQIV examples presented in Section 2. Section 4 provides sieve Wald and SQLR inference procedures based on asymptotic critical values. Section 5 establishes the consistency of the

[6] It is known that θ_0 could have singular semiparametric efficiency bound and could not be root-n estimable; see Chamberlain (2010), Kahn and Tamer (2010), Graham and Powell (2012) and the references therein. Following Kahn and Tamer (2010) and Graham and Powell (2012) we call such a θ_0 irregular. Some applied papers on complicated semi/nonparametric models simply assume that θ_0 is root-n estimable.

bootstrap sieve Wald and the bootstrap SQLR statistics for possibly irregular functionals. Section 6 presents simulation studies and an empirical illustration. Section 7 briefly concludes. Appendix A consists of several subsections, presenting (1) low level sufficient conditions when the conditional mean function $m(\cdot, \alpha) = E[\rho(Y, X; \alpha)|X = \cdot]$ is estimated via a series least squares (LS) procedure; (2) additional useful lemmas; (3) the consistency of additional bootstrap sieve Wald tests; (4) the local power properties of sieve Wald and SQLR tests; and (5) asymptotic properties of sieve Wald and SQLR for functionals of increasing dimension. Online Appendices B and C contain the proofs of the results stated in the main text and in Appendix A respectively. Online Appendix D provides computationally attractive sieve score test and sieve score bootstrap.

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 $(\mathcal{Z}^\infty, \mathcal{B}^\infty_Z, P_{Z^\infty})$ 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 $E_{P_{Z^\infty}}[\cdot]$. 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 $\{a_n\}^\infty_{n=1}$ and $\{b_n\}^\infty_{n=1}$, $a_n = O_P(b_n)$ 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 use "wpa1-$P_{Z^\infty}$" (or simply wpa1) for an event A_n, to denote that $P_{Z^\infty}(A_n) \to 1$ as $n \to \infty$. We use $\mathcal{A}_n \equiv \mathcal{A}_{k(n)}$ and $\mathcal{H}_n \equiv \mathcal{H}_{k(n)}$ for 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. For any mapping $\digamma : \mathbf{H}_1 \to \mathbf{H}_2$ between two generic Banach spaces, $\frac{d\digamma(\alpha_0)}{d\alpha}[v] \equiv \left.\frac{d\digamma(\alpha_0+\tau v)}{d\tau}\right|_{\tau=0}$ is the pathwise (or Gateaux) derivative at α_0 in the direction $v \in \mathbf{H}_1$.

2 PSMD Estimation and Inferences: An Overview

2.1 The Penalized Sieve Minimum Distance Estimator

Let $m(X,\alpha) \equiv E\left[\rho(Y,X;\alpha)|X\right] = \int \rho(y,X;\alpha)dF_{Y|X}(y)$ be a $d_\rho \times 1$ vector valued conditional mean function, $\Sigma(X)$ be a $d_\rho \times d_\rho$ 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 (1.1) 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. In this paper as well as in most applications $\Sigma(X)$ is chosen to be either I_{d_ρ} (identity) or $\Sigma_0(X)$ for almost all X. We call $Q^0(\alpha) \equiv E\left[||m(X,\alpha)||^2_{\Sigma_0^{-1}}\right]$ the population optimally weighted MD criterion function.

Let $\phi : \mathcal{A} \to \mathbb{R}^{d_\phi}$ be a functional with a 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

$$\widehat{Q}_n(\widehat{\alpha}_n) + \lambda_n Pen(\widehat{h}_n) \leq \inf_{\alpha \in \mathcal{A}_{k(n)}} \left\{\widehat{Q}_n(\alpha) + \lambda_n Pen(h)\right\} + o_{P_{Z^\infty}}(n^{-1}), \tag{2.2}$$

where $\lambda_n Pen(h) \geq 0$ 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 $E[\rho(Y, X; h_0)|X] = 0$, Chen and Pouzo (2012a) proposed more general approximate PSMD estimators of h_0 by allowing for possibly infinite dimensional sieves (i.e., $\dim(\mathcal{H}_{k(n)}) = k(n) \le \infty$). Nevertheless, both the theoretical properties and Monte Carlo simulations in Chen and Pouzo (2012a) recommend the use of the PSMD procedures with slowly growing finite-dimensional linear sieves with small penalty (i.e., $k(n)$ grows with n slowly but λ_n goes to zero fast say $\lambda_n = o(n^{-1})$, so the main smoothing parameter is the sieve dimension $k(n)$). This class of PSMD estimators include the original SMD estimators of Newey and Powell (2003) and Ai and Chen (2003) as special cases, and has been used in recent empirical estimation of semiparametric structural models in microeconomics and asset pricing with endogeneity. See, e.g., Blundell et al. (2007), Horowitz (2011), Chen and Pouzo (2009), Bajari et al. (2011), Souza-Rodrigues (2012), Pinkse et al. (2002), Merlo and de Paula (2013), Bontemps and Martimort (2013), Chen and Ludvigson (2009), Chen et al. (2013), Penaranda and Sentana (2013) and others.

In this paper we shall develop inferential theory for $\phi(\alpha_0)$ based on the PSMD procedures with slowly growing finite-dimensional sieves $\mathcal{A}_{k(n)} = \Theta \times \mathcal{H}_{k(n)}$. We first establish the large sample theories under a high level "local quadratic approximation" (LQA) condition, which allows for any consistent nonparametric estimator $\widehat{m}(x, \alpha)$ that is linear in $\rho(Z, \alpha)$:

$$\widehat{m}(x, \alpha) \equiv \sum_{i=1}^{n} \rho(Z_i, \alpha) A_n(X_i, x) \tag{2.4}$$

where $A_n(X_i, x)$ is a known measurable function of $\{X_j\}_{j=1}^n$, whose expression varies according to different nonparametric procedures such as kernel, local linear regression, series and nearest neighbors. In Appendix A we provide lower level sufficient conditions for this LQA assumption when $\widehat{m}(x, \alpha)$ is the series least squares (LS) estimator (2.5):

$$\widehat{m}(x, \alpha) = \left(\sum_{i=1}^{n} \rho(Z_i, \alpha) p^{J_n}(X_i)' \right) (P'P)^{-} p^{J_n}(x), \tag{2.5}$$

which is a linear nonparametric estimator (2.4) with $A_n(X_i, x) = p^{J_n}(X_i)'(P'P)^{-} p^{J_n}(x)$, 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 $J_n \ge d_\theta + k(n) \to \infty$ slowly as $n \to \infty$.[7]

[7]See, e.g., Ai and Chen (2003), Blundell et al. (2007) and Chen and Pouzo (2009) for details about implementation of the PSMD procedures using a series LS estimator (2.5).

2.2 Preview of the Main Results for 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 $\phi(\alpha_0)$.

Sieve t (or Wald) statistic. Regardless of whether $\phi(\alpha_0)$ is $\sqrt{n}$ estimable or not, under some regularity conditions we establish in Theorem 3.1 that $\frac{\sqrt{n}\{\phi(\widehat{\alpha}_n)-\phi(\alpha_0)\}}{||v_n^*||_{sd}}$ is asymptotically standard normal, and the sieve variance $||v_n^*||_{sd}^2$ has a *closed form* expression:

$$||v_n^*||_{sd}^2 = \left(\frac{d\phi(\alpha_0)}{d\alpha}[\overline{q}^{k(n)}(\cdot)]\right)' D_n^{-1}\mho_n D_n^{-1}\left(\frac{d\phi(\alpha_0)}{d\alpha}[\overline{q}^{k(n)}(\cdot)]\right), \tag{2.6}$$

where $\overline{q}^{k(n)}(\cdot) \equiv \left(\mathbf{1}'_{d_\theta}, q^{k(n)}(\cdot)'\right)'$ is a $(d_\theta + k(n)) \times 1$ vector with $\mathbf{1}_{d_\theta}$ a $d_\theta \times 1$ vector of 1's,

$$\frac{d\phi(\alpha_0)}{d\alpha}[\overline{q}^{k(n)}(\cdot)] \equiv \left(\frac{d\phi(\alpha_0)}{d\theta'}, \frac{d\phi(\alpha_0)}{dh}[q^{k(n)}(\cdot)']\right)' \tag{2.7}$$

is a $(d_\theta + k(n)) \times 1$ vector,[8] $\frac{dm(X,\alpha_0)}{d\alpha}[\overline{q}^{k(n)}(\cdot)']$ is a $d_\rho \times (d_\theta + k(n))$ matrix, and

$$D_n = E\left[\left(\frac{dm(X,\alpha_0)}{d\alpha}[\overline{q}^{k(n)}(\cdot)']\right)' \Sigma(X)^{-1}\left(\frac{dm(X,\alpha_0)}{d\alpha}[\overline{q}^{k(n)}(\cdot)']\right)\right], \tag{2.8}$$

$$\mho_n = E\left[\left(\frac{dm(X,\alpha_0)}{d\alpha}[\overline{q}^{k(n)}(\cdot)']\right)' \Sigma(X)^{-1}\rho(Z,\alpha_0)\rho(Z,\alpha_0)'\Sigma(X)^{-1}\left(\frac{dm(X,\alpha_0)}{d\alpha}[\overline{q}^{k(n)}(\cdot)']\right)\right]. \tag{2.9}$$

The closed form expression of $||v_n^*||_{sd}^2$ immediately leads to simple consistent plug-in sieve variance estimators; one of which is

$$||\widehat{v}_n^*||_{n,sd}^2 = \widehat{V}_1 = \left(\frac{d\phi(\widehat{\alpha}_n)}{d\alpha}[\overline{q}^{k(n)}(\cdot)]\right)' \widehat{D}_n^{-1}\widehat{\mho}_n\widehat{D}_n^{-1}\left(\frac{d\phi(\widehat{\alpha}_n)}{d\alpha}[\overline{q}^{k(n)}(\cdot)]\right), \tag{2.10}$$

where

$$\widehat{D}_n = \frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n)}{d\alpha}[\overline{q}^{k(n)}(\cdot)']\right)' \widehat{\Sigma}(X_i)^{-1}\left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n)}{d\alpha}[\overline{q}^{k(n)}(\cdot)']\right)\right], \tag{2.11}$$

$$\widehat{\mho}_n = \frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n)}{d\alpha}[\overline{q}^{k(n)}(\cdot)']\right)' \widehat{\Sigma}(X_i)^{-1}\rho(Z_i,\widehat{\alpha}_n)\rho(Z_i,\widehat{\alpha}_n)'\widehat{\Sigma}(X_i)^{-1}\left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n)}{d\alpha}[\overline{q}^{k(n)}(\cdot)']\right)\right]. \tag{2.12}$$

(See Subsection 4.1 for other consistent sieve variance estimators.) Theorem 4.1 then presents the

[8] 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.

asymptotic normality of the sieve (Student's) t statistic:[9]

$$\widehat{W}_n \equiv \sqrt{n}\frac{\phi(\widehat{\alpha}_n)-\phi(\alpha_0)}{||\widehat{v}_n^*||_{n,sd}} \Rightarrow N(0,1).$$

Sieve QLR statistic. In addition to the sieve t (or sieve Wald) statistic, we could also use sieve quasi likelihood ratio for constructing confidence set of $\phi(\alpha_0)$ and for hypothesis testing of $H_0 : \phi(\alpha_0) = \phi_0$ against $H_1 : \phi(\alpha_0) \neq \phi_0$. Denote

$$\widehat{QLR}_n(\phi_0) \equiv n\left(\inf_{\alpha\in\mathcal{A}_{k(n)}:\phi(\alpha)=\phi_0} \widehat{Q}_n(\alpha) - \widehat{Q}_n(\widehat{\alpha}_n)\right) \tag{2.13}$$

as the *sieve quasi likelihood ratio* (SQLR) statistic. It becomes an *optimally weighted SQLR* statistic, $\widehat{QLR}^0_n(\phi_0)$, when $\widehat{Q}_n(\alpha)$ is the optimally weighted MD criterion. Regardless of whether $\phi(\alpha_0)$ is $\sqrt{n}$ estimable or not, Theorems 4.2 and 3.3 show that $\widehat{QLR}^0_n(\phi_0)$ is asymptotically chi-square distributed under the null H_0, and diverges to infinity under the fixed alternatives H_1. Theorem A.2 in Appendix A states that $\widehat{QLR}^0_n(\phi_0)$ is asymptotically noncentral chi-square distributed under local alternatives. One could compute $100(1-\tau)\%$ confidence set for $\phi(\alpha_0)$ as

$$\left\{r \in \mathbb{R}\colon\ \widehat{QLR}^0_n(r) \leq c_{\chi^2_1}(1-\tau)\right\},$$

where $c_{\chi^2_1}(1-\tau)$ is the $(1-\tau)$-th quantile of the χ^2_1 distribution.

Bootstrap sieve QLR statistic. Regardless of whether $\phi(\alpha_0)$ is $\sqrt{n}$ estimable or not, Theorems 3.2 and 3.3 establish that the possibly non-optimally weighted SQLR statistic $\widehat{QLR}_n(\phi_0)$ is stochastically bounded under the null H_0 and diverges to infinity under the fixed alternatives H_1. We then consider a bootstrap version of the SQLR statistic. Let $\widehat{QLR}^B_n$ denote a bootstrap SQLR statistic:

$$\widehat{QLR}^B_n(\widehat{\phi}_n) \equiv n\left(\inf_{\alpha\in\mathcal{A}_{k(n)}:\phi(\alpha)=\widehat{\phi}_n} \widehat{Q}^B_n(\alpha) - \inf_{\alpha\in\mathcal{A}_{k(n)}} \widehat{Q}^B_n(\alpha)\right), \tag{2.14}$$

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

$$\widehat{Q}^B_n(\alpha) \equiv \frac{1}{n}\sum_{i=1}^{n} \widehat{m}^B(X_i,\alpha)'\widehat{\Sigma}(X_i)^{-1}\widehat{m}^B(X_i,\alpha), \tag{2.15}$$

where $\widehat{m}^B(x,\alpha)$ is a bootstrap version of $\widehat{m}(x,\alpha)$, which is computed in the same way as that of $\widehat{m}(x,\alpha)$ except that we use $\omega_{i,n}\rho(Z_i,\alpha)$ instead of $\rho(Z_i,\alpha)$. Here $\{\omega_{i,n} \geq 0\}_{i=1}^n$ is bootstrap weights that has mean 1 and is independent of the original data $\{Z_i\}_{i=1}^n$. Typical weights include an i.i.d.

[9]See Theorems 5.2, A.1 and A.5 for properties of bootstrap sieve t statistics.

weight $\{\omega_i \geq 0\}_{i=1}^n$ with $E[\omega_i] = 1$, $E[|\omega_i - 1|^2] = 1$ and $E[|\omega_i - 1|^{2+\epsilon}] < \infty$ for some $\epsilon > 0$, or a multinomial weight (i.e., $(\omega_{1,n}, ..., \omega_{n,n}) \sim Multinomial(n; n^{-1}, ..., n^{-1})$). For example, if $\widehat{m}(x, \alpha)$ is a series LS estimator (2.5) of $m(x, \alpha)$, then $\widehat{m}^B(x, \alpha)$ is a bootstrap series LS estimator of $m(x, \alpha)$, defined as:

$$\widehat{m}^B(x, \alpha) \equiv \left(\sum_{i=1}^n \omega_{i,n} \rho(Z_i, \alpha) p^{J_n}(X_i)' \right) (P'P)^- p^{J_n}(x). \tag{2.16}$$

We sometimes call our bootstrap procedure "*generalized residual bootstrap*" since it is based on randomly perturbing the generalized residual function $\rho(Z, \alpha)$; see Section 5 for details. Theorems 5.3 and A.3 establish that under the null H_0, the fixed alternatives H_1 or the local alternatives,[10] the conditional distribution of $\widehat{QLR}_n^B(\widehat{\phi}_n)$ (given the data) always converges to the asymptotic null distribution of $\widehat{QLR}_n(\phi_0)$. Let $\widehat{c}_n(a)$ be the $a - th$ quantile of the distribution of $\widehat{QLR}_n^B(\widehat{\phi}_n)$ (conditional on the data $\{Z_i\}_{i=1}^n$). Then for any $\tau \in (0, 1)$, we have $\lim_{n\to\infty} \Pr\{\widehat{QLR}_n(\phi_0) > \widehat{c}_n(1 - \tau)\} = \tau$ under the null H_0, $\lim_{n\to\infty} \Pr\{\widehat{QLR}_n(\phi_0) > \widehat{c}_n(1 - \tau)\} = 1$ under the fixed alternatives H_1, and $\lim_{n\to\infty} \Pr\{\widehat{QLR}_n(\phi_0) > \widehat{c}_n(1 - \tau)\} > \tau$ under the local alternatives. We could also construct a $100(1 - \tau)\%$ confidence set using the bootstrap critical values:

$$\left\{ r \in \mathbb{R} : \ \widehat{QLR}_n(r) \leq \widehat{c}_n(1 - \tau) \right\}. \tag{2.17}$$

The bootstrap consistency holds for possibly non-optimally weighted SQLR statistic and possibly irregular functionals, without the need to compute standard errors.

Which method to use? When sieve Wald and SQLR tests are computed using the same weighting matrix $\widehat{\Sigma}$, there is no local power difference in terms of first order asymptotic theories; see Appendix A. As will be demonstrated in simulation Section 6, while SQLR and bootstrap SQLR tests are useful for models (1.1) with (pointwise) non-smooth $\rho(Z; \alpha)$, sieve Wald (or t) statistic is computationally attractive for models with smooth $\rho(Z; \alpha)$. Empirical researchers could apply either inference method depending on whether the residual function $\rho(Z; \alpha)$ in their specific application is pointwise differentiable with respect to α or not.

2.2.1 Applications to NPIV and NPQIV models

An illustration via the NPIV model. Blundell et al. (2007) and Chen and Reiß (2011) established the convergence rate of the identity weighted (i.e., $\widehat{\Sigma} = \Sigma = 1$) PSMD estimator $\widehat{h}_n \in \mathcal{H}_{k(n)}$ of the NPIV model:

$$Y_1 = h_0(Y_2) + U, \quad E(U|X) = 0. \tag{2.18}$$

[10] See Section A.4 for definition of the local alternatives and the behaviors of $\widehat{QLR}_n(\phi_0)$ and $\widehat{QLR}_n^B(\widehat{\phi}_n)$ under the local alternatives.

By Theorem 3.1 $\sqrt{n}\frac{\phi(\widehat{h}_n)-\phi(h_0)}{||v_n^*||_{sd}} \Rightarrow N(0,1)$ with $||v_n^*||_{sd}^2 = \frac{d\phi(h_0)}{dh}[q^{k(n)}(\cdot)]' D_n^{-1}\mho_n D_n^{-1}\frac{d\phi(h_0)}{dh}[q^{k(n)}(\cdot)]$,

$$D_n = E\left(E[q^{k(n)}(Y_2)|X]E[q^{k(n)}(Y_2)|X]'\right) \text{ and } \mho_n = E\left(E[q^{k(n)}(Y_2)|X]U^2 E[q^{k(n)}(Y_2)|X]'\right). \tag{2.19}$$

For a functional $\phi(h) = h(\overline{y}_2)$, or $= \int w(y)\nabla h(y)dy$ or $= \int w(y)\,|h(y)|^2\,dy$, we have $\frac{d\phi(h_0)}{dh}[q^{k(n)}(\cdot)] = q^{k(n)}(\overline{y}_2)$, or $= \int w(y)\nabla q^{k(n)}(y)dy$ or $= 2\int h_0(y)w(y)q^{k(n)}(y)dy$.

If $0 < \inf_x \Sigma_0(x) \le \sup_x \Sigma_0(x) < \infty$ then $||v_n^*||_{sd}^2 \asymp \frac{d\phi(h_0)}{dh}[q^{k(n)}(\cdot)]' D_n^{-1}\frac{d\phi(h_0)}{dh}[q^{k(n)}(\cdot)]$. Without endogeneity (say $Y_2 = X$) the model becomes the nonparametric LS regression

$$E[Y_1 = h_0(Y_2) + U, \quad E(U|Y_2) = 0,$$

and the variance satisfies $||v_n^*||_{sd,ex}^2 \asymp \frac{d\phi(h_0)}{dh}[q^{k(n)}(\cdot)]' D_{n,ex}^{-1}\frac{d\phi(h_0)}{dh}[q^{k(n)}(\cdot)]$, $D_{n,ex} = E[\{q^{k(n)}(Y_2)\}\{q^{k(n)}(Y_2)\}']$. Since the conditional expectation $E[q^{k(n)}(Y_2)|X]$ is a contraction, $D_n \le D_{n,ex}$ and $||v_n^*||_{sd}^2 \ge const.||v_n^*||_{sd,ex}^2$. Under mild conditions (see, e.g., Newey and Powell (2003), Blundell et al. (2007), Darolles et al. (2011), Horowitz (2011)), the minimal eigenvalue of D_n, $\lambda_{\min}(D_n)$, goes to zero while $\lambda_{\min}(D_{n,ex})$ stays strictly positive as $k(n) \to \infty$. In fact, $D_{n,ex} = I_{k(n)}$ and $\lambda_{\min}(D_{n,ex}) = 1$ if $\{q_j\}_{j=1}^\infty$ is an orthonormal basis of $L^2(f_{Y_2})$, while $\lambda_{\min}(D_n) \asymp \exp(-k(n))$ if the conditional density of Y_2 given X is normal. Therefore, while $\lim_{k(n)\to\infty}||v_n^*||_{sd,ex}^2 = \infty$ always implies $\lim_{k(n)\to\infty}||v_n^*||_{sd}^2 = \infty$, it is possible that $\lim_{k(n)\to\infty}||v_n^*||_{sd,ex}^2 < \infty$ but $\lim_{k(n)\to\infty}||v_n^*||_{sd}^2 = \infty$. For example, the point evaluation functional $\phi(h) = h(\overline{y}_2)$ is known to be irregular for the nonparametric LS regression and hence for the NPIV (2.18) as well. After mild conditions on the weight function $w()$ and the smoothness of h_0, the weighted derivative functional ($\phi(h) = \int w(y)\nabla h(y)dy$) and the quadratic functional ($\phi(h) = \int w(y)\,|h(y)|^2\,dy$) of the nonparametric LS regression are typically regular, but they could be regular or irregular for the NPIV (2.18). See Subsection 3.5 for details.

It is in general difficult to figure out if the sieve variance $||v_n^*||_{sd}^2$ of the functional $\phi(h)$ (at h_0) goes to infinity or not. Nevertheless, this paper shows that the sieve variance $||v_n^*||_{sd}^2$ has a closed form expression and can be consistently estimated by a plug-in sieve variance estimator $||\widehat{v}_n^*||_{n,sd}^2$. By Theorem 4.1 we obtain $\sqrt{n}\frac{\phi(\widehat{h}_n)-\phi(h_0)}{||\widehat{v}_n^*||_{n,sd}} \Rightarrow N(0,1)$.

When the conditional mean function $m(x,h)$ is estimated by the series LS estimator (2.5) as in Newey and Powell (2003), Ai and Chen (2003) and Blundell et al. (2007), with $\widehat{U}_i = Y_{1i} - \widehat{h}_n(Y_{2i})$, the sieve variance estimator $||\widehat{v}_n^*||_{n,sd}^2$ given in (2.10) has a more explicit expression:

$$||\widehat{v}_n^*||_{n,sd}^2 = \widehat{V}_1 = \left(\frac{d\phi(\widehat{h}_n)}{dh}[q^{k(n)}(\cdot)]\right)' \widehat{D}_n^{-1}\widehat{\mho}_n\widehat{D}_n^{-1}\left(\frac{d\phi(\widehat{h}_n)}{dh}[q^{k(n)}(\cdot)]\right), \quad \text{where}$$

$$\widehat{D}_n = \frac{1}{n}\widehat{C}_n(P'P)^-(\widehat{C}_n)', \quad \widehat{C}_n \equiv \sum_{j=1}^{n} q^{k(n)}(Y_{2j})p^{J_n}(X_j)',$$

$$\widehat{\mho}_n = \frac{1}{n}\widehat{C}_n(P'P)^- \left(\sum_{i=1}^{n} p^{J_n}(X_i)\widehat{U}_i^2 p^{J_n}(X_i)'\right)(P'P)^-(\widehat{C}_n)'. \tag{2.20}$$

Interestingly, this sieve variance estimator becomes the one computed via the two stage least squares (2SLS) as if the NPIV model (2.18) were a parametric IV regression: $Y_1 = q^{k(n)}(Y_{2j})'\beta_{0n} + U$, $E[q^{k(n)}(Y_2)U] \neq 0$, $E[p^{J_n}(X)U] = 0$ and $E[p^{J_n}(X)q^{k(n)}(Y_2)']$ has a column rank $k(n) \leq J_n$. See Subsection 6.1 for simulation studies of finite sample performances of this sieve variance estimator $\widehat{V}_1$ for both a linear and a nonlinear functional $\phi(h)$.

An illustration via the NPQIV model. 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:

$$Y_1 = h_0(Y_2) + U, \quad \Pr(U \leq 0|X) = \gamma. \tag{2.21}$$

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.
By Theorem 3.1 $\sqrt{n}\frac{\phi(\widehat{h}_n)-\phi(h_0)}{||v_n^*||_{sd}} \Rightarrow N(0,1)$ with $||v_n^*||_{sd}^2 = \left(\frac{d\phi(h_0)}{dh}[q^{k(n)}(\cdot)]\right)' D_n^{-1}\left(\frac{d\phi(h_0)}{dh}[q^{k(n)}(\cdot)]\right)$ and

$$D_n = \frac{1}{\gamma(1-\gamma)}E\left(E[f_{U|Y_2,X}(0)q^{k(n)}(Y_2)|X]E[f_{U|Y_2,X}(0)q^{k(n)}(Y_2)|X]'\right). \tag{2.22}$$

Without endogeneity (say $Y_2 = X$), the model becomes the nonparametric quantile regression

$$Y_1 = h_0(Y_2) + U, \quad \Pr(U \leq 0|Y_2) = \gamma,$$

and the sieve variance becomes $||v_n^*||_{sd,ex}^2 = \left(\frac{d\phi(h_0)}{dh}[q^{k(n)}(\cdot)]\right)' D_{n,ex}^{-1}\left(\frac{d\phi(h_0)}{dh}[q^{k(n)}(\cdot)]\right)$ with $D_{n,ex} = \frac{1}{\gamma(1-\gamma)}E\left[\{f_{U|Y_2}(0)\}^2\{q^{k(n)}(Y_2)\}\{q^{k(n)}(Y_2)\}'\right]$. Again $D_n \leq D_{n,ex}$ and $||v_n^*||_{sd}^2 \geq ||v_n^*||_{sd,ex}^2$. Under mild conditions (see, e.g., Chen and Pouzo (2012a), Chen et al. (2013)), $\lambda_{\min}(D_n) \to 0$ while $\lambda_{\min}(D_{n,ex})$ stays strictly positive as $k(n) \to \infty$. All of the above discussions for a functional $\phi(h)$ of the NPIV (2.18) now apply to the functional of the NPQIV (2.21). In particular, a functional $\phi(h)$ could be regular for the nonparametric quantile regression ($\lim_{k(n)\to\infty}||v_n^*||_{sd,ex}^2 < \infty$) but irregular for the NPQIV (2.21) ($\lim_{k(n)\to\infty}||v_n^*||_{sd}^2 = \infty$). See Subsection 3.5 for details.

Applying Theorem 4.2, we immediately obtain that the optimally weighted SQLR statistic $\widehat{QLR}_n^0(\phi_0) \Rightarrow \chi_1^2$ under the null of $\phi(h_0) = \phi_0$. Thus we can compute confidence set for a functional $\phi(h)$, such as an evaluation or a weighted derivative functional, as $\left\{r \in \mathbb{R}\colon \widehat{QLR}_n^0(r) \leq c_{\chi_1^2}(\tau)\right\}$. See Subsection 6.2 for an empirical illustration of this result to the NPQIV Engel curve regression using the British Family Survey data set that was first used in Blundell et al. (2007). Instead of using the critical values based on a χ_1^2 distribution, we could also construct a confidence set using the bootstrap critical values as in (2.17).

2.3 A Brief Discussion on the Convergence Rate

Before we could derive the asymptotic distributions of the sieve Wald and sieve QLR statistics for inference on $\phi(\alpha_0) = \phi_0$, 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 (1.1). 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. 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 $\mathcal{H} \subseteq cl(\cup_k \mathcal{H}_k)$, 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,*[11] *$Q(\Pi_n \alpha_0) = o(1)$; (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.*[12]

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

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

[11] A function Q is lower semicontinuous at a point $\alpha_o \in \mathcal{A}$ iff $\lim_{\|\alpha - \alpha_o\|_s \to 0} Q(\alpha) \geq Q(\alpha_o)$; is lower semicontinuous if it is lower semicontinuous at any point in $\mathcal{A}$.

[12] A function Pen is lower semicompact iff for all M, $\{h \in \mathcal{H} : Pen(h) \leq M\}$ is a compact subset in $(\mathcal{H}, \|\cdot\|_{\mathbf{H}})$.

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, 2.3 and $Q(\Pi_n\alpha_0) + o(n^{-1}) = O(\lambda_n)$ 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 convex, open $||\cdot||_s-$neighborhood around α_0, denoted as $\mathcal{A}_{os}$ such that

$$\mathcal{A}_{os} \subset \{\alpha \in \mathcal{A} : ||\alpha - \alpha_0||_s \leq M_0,\ \lambda_n Pen(h) \leq \lambda_n M_0\}$$

for a positive finite constant M_0. For any $\alpha \in \mathcal{A}_{os}$ we define a pathwise derivative as

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

Following Ai and Chen (2003) and Chen and Pouzo (2009), we introduce two pseudo-metrics $||\cdot||$ and $||\cdot||_0$ on $\mathcal{A}_{os}$ as: for any $\alpha_1,\, \alpha_2 \in \mathcal{A}_{os}$,

$$||\alpha_1 - \alpha_2||^2 \equiv E\left[\left(\frac{dm(X,\alpha_0)}{d\alpha}[\alpha_1 - \alpha_2]\right)' \Sigma(X)^{-1} \left(\frac{dm(X,\alpha_0)}{d\alpha}[\alpha_1 - \alpha_2]\right)\right]; \tag{2.23}$$

$$||\alpha_1 - \alpha_2||_0^2 \equiv E\left[\left(\frac{dm(X,\alpha_0)}{d\alpha}[\alpha_1 - \alpha_2]\right)' \Sigma_0(X)^{-1} \left(\frac{dm(X,\alpha_0)}{d\alpha}[\alpha_1 - \alpha_2]\right)\right]. \tag{2.24}$$

It is clear that, under Assumption 2.1(iv), these two pseudo-metrics are equivalent, i.e., $||\cdot|| \asymp ||\cdot||_0$ on $\mathcal{A}_{os}$. This is the reason why we impose the strong sufficient condition, Assumption 2.1(iv), throughout the paper.

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

Assumption 2.4 (Local curvature)**.** *There exists an open $||\cdot||_s-$neighborhood of α_0, $\mathcal{A}_{os}$, such that* [13] *(i) $\mathcal{A}_{os}$ is 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 $\alpha \in \mathcal{A}_{os}$.*

Let $\mathcal{A}_{osn} = \mathcal{A}_{os} \cap \mathcal{A}_{k(n)}$. Recall the definition of the *sieve measure of local ill-posedness*

$$\tau_n \equiv \sup_{\alpha \in \mathcal{A}_{osn} : ||\alpha - \Pi_n\alpha_0|| \neq 0} \frac{||\alpha - \Pi_n\alpha_0||_s}{||\alpha - \Pi_n\alpha_0||}. \tag{2.25}$$

[13] Given the consistency result, the PSMD estimator will belong to any $||\cdot||_s-$neighborhood around α_0 eventually, so the restriction to an open neighborhood is warranted.

The problem of estimating α_0 under $||\cdot||_s$ is *locally ill-posed in rate* if and only if $\limsup_{n\to\infty} \tau_n = \infty$. We say the problem is *mildly ill-posed* if $\tau_n = O([k(n)]^a)$, and *severely ill-posed* if $\tau_n = O(\exp\{\frac{a}{2}k(n)\})$ for some finite $a > 0$. 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 conditions for Lemma 2.1 hold. Let Assumption 2.4 hold, and $\widehat{Q}_n(\alpha) \geq cQ(\alpha) - O_{P_{Z^\infty}}(\delta_{m,n}^2)$ uniformly over $\mathcal{A}_{osn}$. If $\max\{\delta_{m,n}^2, Q(\Pi_n\alpha_0), \lambda_n, o(n^{-1})\} = \delta_{m,n}^2$ then:*

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

The above convergence rate result is applicable to any 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 the series LS estimator (2.5) of $m(X,\alpha)$. In particular, Lemma C.2 of Chen and Pouzo (2012a) shows that $\widehat{Q}_n(\alpha) \asymp Q(\alpha) - O_{P_{Z^\infty}}(\delta_{m,n}^2)$ uniformly over $\mathcal{A}_{osn}$.

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 $\delta_{s,n} = ||h_0 - \Pi_n h_0||_s + \tau_n\delta_n = o(1)$. Thus $\widehat{\alpha}_n \in \mathcal{N}_{osn} \subseteq \mathcal{N}_{os}$ wpa1-P_{Z^∞}, where

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

Remark 2.1. *To simplify presentation, in the rest of the paper we impose: (1) all the conditions for Lemma 2.2; and (2) $\lambda_n \times \sup_{h,\widetilde{h}\in\mathcal{N}_{os}} \left|Pen(h) - Pen(\widetilde{h})\right| = o(n^{-1})$ or $\lambda_n = o(n^{-1})$.*

Under Remark 2.1(1) we can regard $\mathcal{N}_{os}$ as the effective parameter space and $\mathcal{N}_{osn}$ as its sieve space in the rest of the paper. Under Remark 2.1(1) and (2), we can ignore penalty effect in the first order local asymptotic analysis.

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 (1.1) and the limiting distribution of a properly scaled SQLR statistic. See Appendix A for the case of a vector-valued functional $\phi : \mathcal{A} \to \mathbb{R}^{d_\phi}$ (where d_ϕ could grow slowly with n).

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 (2.23)) 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 let $\frac{d\phi(\alpha_0)}{d\alpha}[v]$ be the pathwise (directional) derivative of the functional $\phi(\cdot)$ at α_0 in the direction of $v = \alpha - \alpha_0 \in \overline{\mathbf{V}}$:

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

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 $(\overline{\mathbf{V}}, ||\cdot||)$ such that

$$\frac{d\phi(\alpha_0)}{d\alpha}[v] = \langle v^*, v \rangle \text{ for all } v \in \overline{\mathbf{V}} \quad \text{and} \quad \|v^*\| \equiv \sup_{v \in \overline{\mathbf{V}}, v \neq 0} \frac{\left|\frac{d\phi(\alpha_0)}{d\alpha}[v]\right|}{\|v\|} < \infty. \tag{3.1}$$

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 any Riesz representer of the linear functional $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ on $(\overline{\mathbf{V}}, ||\cdot||)$.

The above definitions seem to depend on the weighting matrix Σ, but, under Assumption 2.1(iv), we have $||\cdot|| \asymp ||\cdot||_0$, (i.e., the norm $||\cdot||$ (using Σ) is equivalent to the norm $||\cdot||_0$ (using Σ_0) defined in (2.24)), and the Hilbert space $\overline{\mathbf{V}}$ under $||\cdot||$ is the same as that under $||\cdot||_0$. Therefore, under Assumption 2.1(iv), $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ is bounded on $(\overline{\mathbf{V}}, ||\cdot||)$ iff $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ is *bounded* on $(\overline{\mathbf{V}}, ||\cdot||_0)$, i.e.,

$$\sup_{v\in\overline{\mathbf{V}}, v\neq 0} \frac{\left|\frac{d\phi(\alpha_0)}{d\alpha}[v]\right|}{\|v\|_0} < \infty,$$

which corresponds to *non-singular semiparametric efficiency bound*, and in this case we say that $\phi(\cdot)$ is *regular* (at $\alpha = \alpha_0$).[14] Likewise, $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ is unbounded on $(\overline{\mathbf{V}}, ||\cdot||)$ iff $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ is *unbounded* on $(\overline{\mathbf{V}}, ||\cdot||_0)$ i.e., $\sup_{v\in\overline{\mathbf{V}}, v\neq 0}\left\{\left|\frac{d\phi(\alpha_0)}{d\alpha}[v]\right| / \|v\|_0\right\} = \infty$, in this case we say that $\phi(\cdot)$ is *irregular* (at $\alpha = \alpha_0$).

It is known that non-singular semiparametric efficiency bound (i.e., $\phi(\cdot)$ being regular or $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ being bounded on $(\overline{\mathbf{V}}, ||\cdot||_0)$) 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 (1.1), it is difficult to compute $\sup_{v\in\overline{\mathbf{V}}, v\neq 0}\left\{\left|\frac{d\phi(\alpha_0)}{d\alpha}[v]\right| / \|v\|_0\right\}$ explicitly; and hence difficult to verify its root-n estimableness.

3.1.1 Sieve Riesz representer and sieve variance

Let $\alpha_{0,n} \in \mathcal{A}_{osn}$ be such that

$$||\alpha_{0,n} - \alpha_0|| \equiv \min_{\alpha\in\mathcal{A}_{osn}} ||\alpha - \alpha_0||. \tag{3.2}$$

Let $\overline{\mathbf{V}}_{k(n)} \equiv clsp(\mathcal{A}_{osn} - \{\alpha_{0,n}\})$, 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 $v_n^* \in \overline{\mathbf{V}}_{k(n)}$ such that

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

[14] Following the proof in appendix E of Ai and Chen (2003), it is easy to see the equivalence between $\sup_{v\in\overline{\mathbf{V}}, v\neq 0}\left\{\left|\frac{d\phi(\alpha_0)}{d\alpha}[v]\right| / \|v\|_0\right\}$ being finite and the semiparametric efficiency bound being non-singular.

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, for any non-zero linear functional $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$, we have:

$$0 < \|v_n^*\|^2 = E\left[\left(\frac{dm(X,\alpha_0)}{d\alpha}[v_n^*]\right)' \Sigma(X)^{-1}\left(\frac{dm(X,\alpha_0)}{d\alpha}[v_n^*]\right)\right] \text{ is non-decreasing in } k(n).$$

We emphasize that the sieve Riesz representer v_n^* of a 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. Moreover, $v_n^* \in \overline{\mathbf{V}}_{k(n)}$ and its norm $\|v_n^*\|$ can be computed in closed form (see Subsection 3.3.1). The next lemma allows us to verify whether or not $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ is bounded on $(\overline{\mathbf{V}}, ||\cdot||)$ (i.e., $\phi(\cdot)$ is regular at $\alpha = \alpha_0$) by checking whether or not $\lim_{k(n)\to\infty} \|v_n^*\| < \infty$.

Lemma 3.1. *Let $\{\overline{\mathbf{V}}_k\}_{k=1}^{\infty}$ be an increasing sequence of finite dimensional Hilbert spaces that is dense in $(\overline{\mathbf{V}}, \|\cdot\|)$, and $v_n^* \in \overline{\mathbf{V}}_{k(n)}$ be defined in (3.3). (1) $\lim_{k(n)\to\infty} \|v_n^*\| < \infty$ iff $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ is bounded on $(\overline{\mathbf{V}}, ||\cdot||)$; and in this case (3.1) holds, $v_n^* = \arg\min_{v\in\overline{\mathbf{V}}_{k(n)}} \|v^* - v\|$ and $\|v^* - v_n^*\| \to 0$, $\|v_n^*\| \to \|v^*\| < \infty$ as $k(n) \to \infty$. (2) $\lim_{k(n)\to\infty} \|v_n^*\| = \infty$ iff $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ is unbounded on $(\overline{\mathbf{V}}, ||\cdot||)$.*

Sieve score and sieve variance. For each sieve dimension $k(n)$, we call

$$S_{n,i}^* \equiv \left(\frac{dm(X_i,\alpha_0)}{d\alpha}[v_n^*]\right)' \Sigma(X_i)^{-1}\rho(Z_i,\alpha_0) \tag{3.4}$$

the *sieve score* associated with the i-th observation, and $\|v_n^*\|_{sd}^2 \equiv Var\left(S_{n,i}^*\right)$ as the *sieve variance*. Recall that $\Sigma_0(X) \equiv Var(\rho(Z;\alpha_0)|X)$ a.s.-X. Then

$$\|v_n^*\|_{sd}^2 = E[S_{n,i}^* S_{n,i}^{*\prime}] = E\left[\left(\frac{dm(X,\alpha_0)}{d\alpha}[v_n^*]\right)' \Sigma(X)^{-1}\Sigma_0(X)\Sigma(X)^{-1}\left(\frac{dm(X,\alpha_0)}{d\alpha}[v_n^*]\right)\right]. \tag{3.5}$$

(See Subsection 3.3.1 for closed form expressions of $\|v_n^*\|_{sd}^2$.) Under Assumption 2.1(iv), we have $\|v_n^*\|_{sd}^2 \asymp \|v_n^*\|^2$, and hence $\|v_n^*\|_{sd}^2 \to \infty$ iff $\|v_n^*\|^2 \to \infty$ (iff $\phi(\cdot)$ is irregular at $\alpha = \alpha_0$). Moreover, if $\phi(\cdot)$ is regular at $\alpha = \alpha_0$ then we can define

$$S_i^* \equiv \left(\frac{dm(X_i,\alpha_0)}{d\alpha}[v^*]\right)' \Sigma(X_i)^{-1}\rho(Z_i,\alpha_0)$$

as the *score* associated with the i-th observation, and $\|v^*\|_{sd}^2 \equiv Var(S_i^*)$ as the *asymptotic variance*. By Lemma 3.1(1) for a regular functional we have: $\|v^*\|_{sd}^2 \asymp \|v^*\| < \infty$ and $Var\left(S_i^* - S_{n,i}^*\right) \asymp \|v^* - v_n^*\|^2 \to 0$ as $k(n) \to \infty$. See Remark 3.2 for further discussion.

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

For all $k(n)$, let

$$u_n^* \equiv \frac{v_n^*}{\|v_n^*\|_{sd}} \tag{3.6}$$

be the "scaled sieve Riesz representer". Since $\|v_n^*\|_{sd}^2 \asymp \|v_n^*\|^2$ (under Assumption 2.1(iv)), we have: $\|u_n^*\| \asymp 1$ and $\|u_n^*\|_s \leq c\tau_n$ for τ_n defined in (2.25) and a finite constant $c > 0$.

Let $\mathcal{T}_n \equiv \{t \in \mathbb{R} \colon |t| \leq 4M_n^2\delta_n\}$ with M_n and δ_n given 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 non-zero linear functional mapping from $\overline{\mathbf{V}}$ to $\mathbb{R}$; $\{\overline{\mathbf{V}}_k\}_{k=1}^{\infty}$ is an increasing sequence of finite dimensional Hilbert spaces that is dense in $(\overline{\mathbf{V}}, \|\cdot\|)$; and $\frac{\|v_n^*\|}{\sqrt{n}} = o(1)$;*

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

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

Since $\|v_n^*\|_{sd}^2 \asymp \|v_n^*\|^2$ (under Assumption 2.1(iv)), we could rewrite Assumption 3.1 using $\|v_n^*\|_{sd}$ instead $\|v_n^*\|$. As it will become clear in Theorem 3.1 that $\frac{\|v_n^*\|_{sd}^2}{n}$ is the variance of $\phi(\widehat{\alpha}_n) - \phi(\alpha_0)$, Assumption 3.1(i) puts a restriction on how fast the sieve dimension $k(n)$ could grow with the sample size n.

Assumption 3.1(ii) controls the nonlinearity bias of $\phi\left(\cdot\right)$ (i.e., the linear approximation error of a possibly nonlinear functional $\phi\left(\cdot\right)$). It is implied by the following condition:

Assumption 3.1(ii)': *there are finite non-negative constants $C \geq 0, \omega_1, \omega_2 \geq 0$ such that for all $(\alpha, t) \in \mathcal{N}_{osn} \times \mathcal{T}_n$,*

$$\left|\phi(\alpha + tu_n^*) - \phi(\alpha_0) - \frac{d\phi(\alpha_0)}{d\alpha}[\alpha + tu_n^* - \alpha_0]\right| \leq C \times (||\alpha - \alpha_0 + tu_n^*||^{\omega_1} \times ||\alpha - \alpha_0 + tu_n^*||_s^{\omega_2}), \quad \textit{and}$$

$$C \times \frac{\sqrt{n} \times (\delta_n(1 + M_n^2))^{\omega_1} \times (\delta_{s,n} + M_n^2\delta_n||u_n^*||_s)^{\omega_2}}{||v_n^*||} = o\left(1\right).$$

Assumption 3.1(ii) (or (ii)') is automatically satisfied when $\phi\left(\cdot\right)$ is a linear functional, such as the Euclidean parameter functional, the evaluation functional, the weighted integration functional; the weighted derivative functional and others. For a nonlinear functional $\phi\left(\cdot\right)$ (such as the quadratic functional), it can be verified using the smoothness of $\phi\left(\cdot\right)$ and the convergence rates in both $|| \cdot ||$ and $|| \cdot ||_s$ metrics (the definition of $\mathcal{N}_{osn}$).

Assumption 3.1(iii) controls the linear 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 $k(n)$.

When $\phi(\cdot)$ is an irregular functional, we have $\|v_n^*\| \nearrow \infty$. 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, $n^{-1/2}\|v_n^*\|_{sd}$. This is a standard condition imposed for the asymptotic normality of any plug-in nonparametric estimator of an irregular functional (such as a point evaluation functional of a nonparametric mean regression).

Remark 3.1. *When $\phi(\cdot)$ is a regular functional (i.e., $\|v_n^*\| \nearrow \|v^*\| < \infty$), 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 Assumption 3.1(iii) is satisfied if*

$$||v^* - v_n^*|| \times ||\alpha_{0,n} - \alpha_0|| = o(n^{-1/2}). \tag{3.7}$$

This is similar to assumption 4.2 in Ai and Chen (2003) and assumption 3.2(iii) in Chen and Pouzo (2009) for the regular Euclidean parameter θ satisfying the model (1.1). As pointed out by Chen and Pouzo (2009), under Lemma 3.1(1), Condition (3.7) could be satisfied when $\dim(\mathcal{A}_{k(n)}) \asymp k(n)$ is chosen to obtain optimal nonparametric convergence rate in $||\cdot||_s$ norm. But this nice feature only applies to regular functionals.

Assumption 3.1 can be verified for typical functionals in semi/nonparametric econometrics. See Subsection 3.5 for the verification via several functionals of the NPIV and NPQIV models.

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 $u_n^* = v_n^* / \|v_n^*\|_{sd}$.

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

$$\mathbb{Z}_n \equiv n^{-1} \sum_{i=1}^{n} \left(\frac{dm(X_i, \alpha_0)}{d\alpha}[u_n^*] \right)' \Sigma(X_i)^{-1} \rho(Z_i, \alpha_0) = n^{-1} \sum_{i=1}^{n} \frac{S_{n,i}^*}{\|v_n^*\|_{sd}}. \tag{3.8}$$

Assumption 3.2 (LQA). *(i) For all n, $(\alpha, t) \in \mathcal{N}_{osn} \times \mathcal{T}_n$ and $\alpha(t) \in \mathcal{A}_{k(n)}$; 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 $(B_n)_n$ is such that, for each n, B_n is Z^n measurable positive random variable and $B_n = O_{P_{Z^\infty}}(1)$; (ii) $\sqrt{n}\mathbb{Z}_n \Rightarrow N(0,1)$.

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

$$\limsup_{n\to\infty} E\left[\left(\frac{S_{n,i}^*}{\|v_n^*\|_{sd}}\right)^2 1\left\{\left|\frac{S_{n,i}^*}{\epsilon\sqrt{n}\,\|v_n^*\|_{sd}}\right| > 1\right\}\right] = 0, \tag{3.9}$$

which, under Lemma 3.1(1) and Assumption 2.1(iv), is satisfied when the functional $\phi(\cdot)$ is regular ($\|v_n^*\|_{sd} \asymp \|v_n^*\| \to \|v^*\| < \infty$). 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 $\phi(\alpha) = \lambda'\theta$.

Assumption 3.2(i) implicitly imposes restrictions on the nonparametric estimator $\widehat{m}(x, \alpha)$ of the conditional mean function $m(x, \alpha) = E[\rho(Z, \alpha)|X = x]$ in a shrinking neighborhood of α_0, so that the criterion difference could be well approximated by a quadratic form. It is trivially satisfied when $\widehat{m}(x, \alpha)$ is linear in α, such as the series LS estimator (2.5) when $\rho(Z, \alpha)$ is linear in α. There are two potential difficulties in verification of this assumption for nonlinear conditional moment models with nonparametric endogeneity. First, due to the non-smooth residual function $\rho(Z, \alpha)$, the estimator $\widehat{m}(x, \alpha)$ (and hence the sample criterion $\widehat{Q}_n(\alpha)$) could be pointwise non-smooth with respect to α. Second, due to the slow convergence rates in the strong norm $||\cdot||_s$ present in nonlinear nonparametric ill-posed inverse problems, it could be challenging to control the remainder of a quadratic approximation. In Appendix A we present one set of relatively low level sufficient conditions (Assumptions A.1 - A.4) to tackle both issues. More precisely, when $\widehat{m}(x, \alpha)$ is a series LS estimator of $m(x, \alpha)$, we show that, under these conditions, $\widehat{Q}_n(\alpha)$ can be well approximated by a "smooth" version of it uniformly in $\alpha \in \mathcal{N}_{osn}$, and that the remainder term of a quadratic approximation is of the right order. The next lemma formally states the result.

Lemma 3.2. *Let $\widehat{m}$ be the series LS estimator (2.5) and conditions for Remark 2.1 hold. If Assumptions A.1 - A.4 in Appendix A hold, then Assumption 3.2(i) holds.*

We note that Assumptions A.1 - A.4 in Appendix A are comparable to the ones imposed in Chen and Pouzo (2009) for the root-n asymptotic normality of the PSMD estimator $\widehat{\theta}_n$ when the Euclidean parameter functional $\phi(\alpha) = \lambda'\theta$ is assumed to be regular. These conditions are already verified in Chen and Pouzo (2009) for a non-trivial, partially linear quantile IV regression model $E[1\{Y_1 \leq h_0(Y_2) + Y_3'\theta_0\} - \gamma|X] = 0$. See Subsection 3.5 for verification of these conditions for irregular functionals of NPIV and NPQIV models.

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 (1.1). Recall that $u_n^* \equiv v_n^* / \|v_n^*\|_{sd}$.

Theorem 3.1. *Let $\widehat{\alpha}_n$ be the PSMD estimator (2.2) and conditions for Remark 2.1 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 $\alpha = \alpha_0$, we have $\|v_n^*\|_{sd} \asymp \|v_n^*\| = O(1)$ and $\phi(\widehat{\alpha}_n)$ converges to $\phi(\alpha_0)$ at the parametric rate of $1/\sqrt{n}$. When the functional $\phi(\cdot)$ is irregular at $\alpha = \alpha_0$, we have $\|v_n^*\|_{sd} \asymp \|v_n^*\| \to \infty$; so the convergence rate of $\phi(\widehat{\alpha}_n)$ becomes slower than $1/\sqrt{n}$.

For any regular functional of the semi/nonparametric model (1.1), Theorem 3.1 implies that

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

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

Thus, 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 (1.1). See Remark 3.2 for further discussion.

3.3.1 Closed form expressions of sieve Riesz representer and sieve variance

To apply Theorem 3.1, one needs to know the sieve Riesz representer v_n^* defined in (3.3) and the sieve variance $\|v_n^*\|_{sd}^2$ given in (3.5). It turns out that both can be computed in closed form.

Lemma 3.3. *Let* $\overline{\mathbf{V}}_{k(n)} = \mathbb{R}^{d_\theta} \times \{v_h(\cdot) = \psi^{k(n)}(\cdot)'\beta : \beta \in \mathbb{R}^{k(n)}\} = \{v(\cdot) = \overline{\psi}^{k(n)}(\cdot)'\gamma : \gamma \in \mathbb{R}^{d_\theta + k(n)}\}$ *be dense in the infinite dimensional Hilbert space* $(\overline{\mathbf{V}}, \|\cdot\|)$ *with the norm* $\|\cdot\|$ *defined in (2.23). Then: the sieve Riesz representer* $v_n^* = (v_{\theta,n}^{*\prime}, v_{h,n}^*(\cdot))' \in \overline{\mathbf{V}}_{k(n)}$ *of* $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ *has a closed form expression:*

$$v_n^* = (v_{\theta,n}^{*\prime}, \psi^{k(n)}(\cdot)'\beta_n^*)' = \overline{\psi}^{k(n)}(\cdot)'\gamma_n^*, \text{ and } \gamma_n^* = D_n^{-1}\digamma_n \tag{3.10}$$

with $D_n = E\left[\left(\frac{dm(X,\alpha_0)}{d\alpha}[\overline{\psi}^{k(n)}(\cdot)']\right)'\Sigma(X)^{-1}\left(\frac{dm(X,\alpha_0)}{d\alpha}[\overline{\psi}^{k(n)}(\cdot)']\right)\right]$ *and* $\digamma_n = \frac{d\phi(\alpha_0)}{d\alpha}[\overline{\psi}^{k(n)}(\cdot)]$. *Thus*

$$\|v_n^*\|^2 = \gamma_n^{*\prime}D_n\gamma_n^* = \digamma_n' D_n^{-1}\digamma_n. \tag{3.11}$$

Let $\mathcal{A}_{k(n)} = \Theta \times \mathcal{H}_{k(n)}$ *with* $\mathcal{H}_{k(n)}$ *given in (2.3). Then* $\overline{\mathbf{V}}_{k(n)} = clsp\left(\mathcal{A}_{k(n)} - \{\alpha_{0,n}\}\right)$ *and one could let* $\overline{\psi}^{k(n)}(\cdot) = \overline{q}^{k(n)}(\cdot)$.

Lemmas 3.1 and 3.3 imply that $\phi(\cdot)$ is *regular (or irregular) at* $\alpha = \alpha_0$ *iff* $\lim_{k(n)\to\infty}\left(\digamma_n' D_n^{-1}\digamma_n\right) < \infty$ *(or* $= \infty$*).*

For a semi/nonparametric conditional moment model with $\alpha_0 = (\theta_0', h_0)$, it is convenient to rewrite D_n and its inverse in Lemma 3.3 as

$$D_n \equiv \begin{pmatrix} I_{11} & I_{n,12} \\ I_{n,12}' & I_{n,22} \end{pmatrix} \quad \text{and} \quad D_n^{-1} = \begin{pmatrix} I_n^{11} & -I_{11}^{-1} I_{n,12} I_n^{22} \\ -I_{n,22}^{-1} I_{n,12}' I_n^{11} & I_n^{22} \end{pmatrix},$$

$I_{11} = E\left[\left(\frac{dm(X,\alpha_0)}{d\theta'}\right)' \Sigma(X)^{-1} \frac{dm(X,\alpha_0)}{d\theta'}\right]$, $I_{n,22} = E\left[\left(\frac{dm(X,\alpha_0)}{dh}[\psi^{k(n)}(\cdot)']\right)' \Sigma(X)^{-1} \left(\frac{dm(X,\alpha_0)}{dh}[\psi^{k(n)}(\cdot)']\right)\right]$, $I_{n,12} = E\left[\left(\frac{dm(X,\alpha_0)}{d\theta'}\right)' \Sigma(X)^{-1} \left(\frac{dm(X,\alpha_0)}{dh}[\psi^{k(n)}(\cdot)']\right)\right]$, $I_n^{11} = \left(I_{11} - I_{n,12} I_{n,22}^{-1} I_{n,21}'\right)^{-1}$ and $I_n^{22} = \left(I_{n,22} - I_{n,21}' I_{11}^{-1} I_{n,12}\right)^{-1}$.

Remark 3.2. *For the Euclidean parameter functional $\phi(\alpha) = \lambda'\theta$, we have $F_n = (\lambda', \mathbf{0}_{k(n)}')'$ with $\mathbf{0}_{k(n)}' = [0, ..., 0]_{1\times k(n)}$, and hence $v_n^* = (v_{\theta,n}^{*\prime}, \psi^{k(n)}(\cdot)'\beta_n^*)' \in \overline{\mathbf{V}}_{k(n)}$ with $v_{\theta,n}^* = I_n^{11}\lambda$, $\beta_n^* = -I_{n,22}^{-1} I_{n,21}' v_{\theta,n}^*$, and $\|v_n^*\|^2 = \lambda' I_n^{11}\lambda$. Thus the functional $\phi(\alpha) = \lambda'\theta$ is regular iff $\lim_{k(n)\to\infty} \lambda' I_n^{11}\lambda < \infty$; 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

$$\mathcal{I}_* = \inf_{\mathbf{w}} E\left[\left\|\Sigma(X)^{-\frac{1}{2}}\left(\frac{dm(X,\alpha_0)}{d\theta'} - \frac{dm(X,\alpha_0)}{dh}[\mathbf{w}]\right)\right\|_e^2\right], \tag{3.12}$$

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 (3.12). 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 $\Sigma(X) = \Sigma_0(X)$, then $\mathcal{I}_*$ becomes the semiparametric efficiency bound for θ_0 that was derived in Chamberlain (1992) and Ai and Chen (2003) for the model (1.1). Lemma 3.1 implies that one could check whether θ_0 has non-singular efficiency bound or not by checking if $\lim_{k(n)\to\infty} \lambda' I_n^{11}\lambda < \infty$ or not.

By Lemma 3.3, the sieve variance (3.5) also has closed form expressions:

$$||v_n^*||_{sd}^2 = F_n' D_n^{-1} \mho_n D_n^{-1} F_n = F_n' D_n^{-1} \Omega_n D_n^{-1} F_n, \tag{3.13}$$

$$\begin{aligned}
\mho_n &\equiv E\left[\left(\frac{dm(X,\alpha_0)}{d\alpha}[\overline{\psi}^{k(n)}(\cdot)']\right)' \Sigma(X)^{-1}\rho(Z,\alpha_0)\rho(Z,\alpha_0)'\Sigma(X)^{-1}\left(\frac{dm(X,\alpha_0)}{d\alpha}[\overline{\psi}^{k(n)}(\cdot)']\right)\right], \\
\Omega_n &\equiv E\left[\left(\frac{dm(X,\alpha_0)}{d\alpha}[\overline{\psi}^{k(n)}(\cdot)']\right)' \Sigma(X)^{-1}\Sigma_0(X)\Sigma(X)^{-1}\left(\frac{dm(X,\alpha_0)}{d\alpha}[\overline{\psi}^{k(n)}(\cdot)']\right)\right] = \mho_n,
\end{aligned}$$

which coincides with the sieve variance expression given in (2.6) when $\psi^{k(n)}(\cdot) = q^{k(n)}(\cdot)$ sieve is used.

According to Lemma 3.3 we could use different finite dimensional linear sieve basis $\psi^{k(n)}$ to compute sieve Riesz representer $v_n^* = (v_{\theta,n}^{*\prime}, v_{h,n}^*(\cdot))' \in \overline{\mathbf{V}}_{k(n)}$, $\|v_n^*\|^2$ and $||v_n^*||_{sd}^2$. Most typical choices include orthonormal bases and the original sieve basis $q^{k(n)}$ (used to approximate unknown function h_0). It is typically easier to characterize the speed of $\|v_n^*\|^2 = \digamma_n' D_n^{-1} \digamma_n$ as a function of $k(n)$ when an orthonormal basis is used, while there is a nice interpretation in terms of sieve variance estimation when the original sieve basis $q^{k(n)}$ is used. See Subsections 2.2, 3.5 and 4.1 for related discussions.

3.4 Asymptotic properties of the SQLR

We now characterize the asymptotic behavior of the possibly *non-optimally weighted* SQLR statistic $\widehat{QLR}_n(\phi_0)$ defined in (2.13).

Let $\mathcal{A}_{k(n)}^R \equiv \{\alpha \in \mathcal{A}_{k(n)} : \phi(\alpha) = \phi_0\}$ be the restricted sieve space, and $\widehat{\alpha}_n^R \in \mathcal{A}_{k(n)}^R$ be a restricted approximate PSMD estimator, defined as

$$\widehat{Q}_n(\widehat{\alpha}_n^R) + \lambda_n Pen(\widehat{h}_n^R) \leq \inf_{\alpha \in \mathcal{A}_{k(n)}^R} \left\{ \widehat{Q}_n(\alpha) + \lambda_n Pen(h) \right\} + o_{P_{Z^\infty}}(n^{-1}). \tag{3.14}$$

Then:

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

Theorem 3.2. *Let Conditions for Remark 2.1, Assumptions 3.1 and 3.2 with* $\left|B_n - ||u_n^*||^2\right| = o_{P_{Z^\infty}}(1)$ *hold. If* $\widehat{\alpha}_n^R \in \mathcal{N}_{osn}$ *wpa1-*P_{Z^∞}*, then: under the null* $H_0 : \phi(\alpha_0) = \phi_0$*,*

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

See Theorem A.2 in Appendix A for the asymptotic behavior under local alternatives.

Compared to Theorem 3.1(2) on the asymptotic normality of $\phi(\widehat{\alpha}_n)$, Theorem 3.2 on the asymptotic null distribution of the SQLR statistic requires two extra conditions: $\left|B_n - ||u_n^*||^2\right| = o_{P_{Z^\infty}}(1)$ and $\widehat{\alpha}_n^R \in \mathcal{N}_{osn}$ wpa1-P_{Z^∞}. Both conditions are also needed even for QLR statistics in parametric extremum estimation and testing problems. Lemma 5.2(2) in Section 5 provides a simple sufficient condition (Assumption B) for $\left|B_n - ||u_n^*||^2\right| = o_{P_{Z^\infty}}(1)$. Proposition B.1 in Appendix B establishes $\widehat{\alpha}_n^R \in \mathcal{N}_{osn}$ wpa1-P_{Z^∞} under the null $H_0 : \phi(\alpha_0) = \phi_0$ and other conditions virtually the same as those for Lemma 2.2 (i.e., $\widehat{\alpha}_n \in \mathcal{N}_{osn}$ wpa1-P_{Z^∞}).

Next, we consider the asymptotic behavior of $\widehat{QLR}_n(\phi_0)$ under the fixed alternatives H_1 : $\phi(\alpha_0) \neq \phi_0$. Let $\mathcal{A}^{R,M} \equiv \{\alpha \in \mathcal{A} : \phi(\alpha) = \phi_0, Pen(h) \leq M\}$ be a restricted parameter space (where $M < \infty$ is such that $Pen(h_0) < M$). Then $\alpha_0 \in \mathcal{A}^{R,M}$ iff the null $H_0 : \phi(\alpha_0) = \phi_0$ holds.

Theorem 3.3. *Let Assumptions 2.1, 2.2 and 2.3 hold. Suppose that* $\sup_{h\in\mathcal{H}} Pen(h) < \infty$ *and* $\mathcal{A}^{R,M}$ *is non-empty, compact (in* $||\cdot||_s$*). Then: under* $H_1 : \phi(\alpha_0) \neq \phi_0$*, there is a constant* $C > 0$ *such that*

$$\liminf_{n\to\infty} \frac{\widehat{QLR}_n(\phi_0)}{n} \geq C > 0 \quad \textit{in probability.}$$

3.5 Verification of Assumptions 3.1 and 3.2(i)

In this subsection, we illustrate the verification of the two key regularity conditions, Assumption 3.1 and Assumption 3.2(i), via some functionals $\phi(h)$ of the (nonlinear) nonparametric IV regressions:

$$E[\rho(Y_1; h_0(Y_2))|X] = 0 \quad a.s. - X, \tag{3.15}$$

where the scalar valued residual function $\rho()$ could be nonlinear and pointwise non-smooth in h. This model includes the NPIV and NPQIV as special cases. To be concrete, we consider a PSMD estimator $\widehat{h} \in \mathcal{H}_n$ of h_0 with $\widehat{\Sigma} = \Sigma = 1$, and $\widehat{m}(\cdot, h)$ being the series LS estimator (2.5) of $m(\cdot, h) = E[\rho(Y_1; h(Y_2))|X = \cdot]$ with $J_n = ck(n)$ for a finite constant $c \geq 1$. We assume that $h_0 \in \mathcal{H} = \Lambda_c^\varsigma([-1, 1])$ with smoothness $\zeta > 1/2$ (a Hölder ball with support $[-1, 1]$, see, e.g., Chen et al. (2003)).[15] By definition, $\mathcal{H} \subset L^2(f_{Y_2})$ and we let $||\cdot||_s = ||\cdot||_{L^2(f_{Y_2})}$. We assume that $\mathcal{H}_n = clsp\{q_1, ..., q_{k(n)}\}$ with $\{q_k\}_{k=1}^\infty$ being a Riesz basis of $(\mathcal{H}, ||\cdot||_s)$. The convergence rates of $\widehat{h}$ to h_0 in both $||\cdot||$ and $||\cdot||_s = ||\cdot||_{L^2(f_{Y_2})}$ metrics have already been established in Chen and Pouzo (2012a), and hence will not be repeated here.

We use $\mathcal{H}_{os}$ and $\mathcal{H}_{osn}$ for $\mathcal{A}_{os}$ and $\mathcal{A}_{osn}$ defined in Subsection 2.3 (since there is no θ here). Denote $T \equiv \frac{dm(\cdot,h_0)}{dh} : \mathcal{H}_{os} \subset L^2(f_{Y_2}) \to L^2(f_X)$, i.e., for any $h \in \mathcal{H}_{os} \subset L^2(f_{Y_2})$,

$$Th \equiv \left. \frac{dE[\rho(Y_1; h_0(Y_2) + \tau h(Y_2))|X = \cdot]}{d\tau} \right|_{\tau=0}.$$

Let T^* be the adjoint of T. Then for all $h \in \mathcal{H}_{os}$, we have $||h||^2 \equiv ||Th||^2_{L^2(f_X)} = ||(T^*T)^{1/2}h||^2_{L^2(f_{Y_2})}$. Under mild conditions as stated in Chen and Pouzo (2012a), T and T^* are compact. Then T has a singular value decomposition $\{\mu_k; \psi_k, \phi_{0k}\}_{k=1}^\infty$, where $\{\mu_k > 0\}_{k=1}^\infty$ is the sequence of singular values in non-increasing order ($\mu_k \geq \mu_{k+1} \geq ...$) with $\liminf_{k\to\infty} \mu_k = 0$, $\{\psi_k \in L^2(f_{Y_2})\}_{k=1}^\infty$ and $\{\phi_{0k} \in L^2(f_X)\}_{k=1}^\infty$ are sequences of eigenfunctions of the operators $(T^*T)^{1/2}$ and $(TT^*)^{1/2}$:

$$T\psi_k = \mu_k\phi_{0k}, \quad (T^*T)^{1/2}\psi_k = \mu_k\psi_k \quad \text{and} \quad (TT^*)^{1/2}\phi_{0k} = \mu_k\phi_{0k} \quad \text{for all } k.$$

[15] This Hölder ball condition and several other conditions assumed in this subsection are for illustration only, and can be replaced by weaker sufficient conditions.

Since $\{q_k\}_{k=1}^{\infty}$ is a Riesz basis of $(\mathcal{H}, ||\cdot||_s)$ we could also have $\mathcal{H}_n = clsp\{\psi_1, ..., \psi_{k(n)}\}$. The sieve measure of local ill-posedness now becomes $\tau_n = \mu_{k(n)}^{-1}$ (see, e.g., Blundell et al. (2007) and Chen and Pouzo (2012a)), and hence $\|u_n^*\|_s \le c\mu_{k(n)}^{-1}$ for a finite constant $c > 0$. Also, $\Pi_n h_0 \equiv \arg\min_{h\in\mathcal{H}_n} ||h - h_0||_s = \sum_{k=1}^{k(n)} \langle h_0, \psi_k\rangle_s \psi_k$ is the LS projection of h_0 onto the sieve space $\mathcal{H}_n$ under the strong norm $||\cdot||_s = ||\cdot||_{L^2(f_{Y_2})}$. Recall that $h_{0,n} \equiv \arg\min_{h\in\mathcal{H}_{osn}} ||h - h_0||^2 \equiv \arg\min_{h\in\mathcal{H}_{osn}} ||T[h - h_0]||^2_{L^2(f_X)}$. We have:

$$h_{0,n} = \arg\min_{\{a_k\}} \left[\sum_{k=1}^{k(n)} \left(\langle h_0, \psi_k\rangle_s - a_k\right)^2 \mu_k^2 + \sum_{k=k(n)+1}^{\infty} \langle h_0, \psi_k\rangle_s^2 \mu_k^2 \right] = \sum_{k=1}^{k(n)} \langle h_0, \psi_k\rangle_s \psi_k = \Pi_n h_0. \quad (3.16)$$

The next remark specializes Theorem 3.1 to a general functional $\phi(h)$ of the model (3.15).

Remark 3.3. *Let $\widehat{m}$ be the series LS estimator (2.5) for the model (3.15) with $\widehat{\Sigma} = \Sigma = 1$, and conditions for Remark 2.1 hold with $\delta_n = O\left(\sqrt{\frac{k(n)}{n}}\right) = o(n^{-1/4})$ and $\delta_{s,n} = O\left(\{k(n)\}^{-\varsigma} + \mu_{k(n)}^{-1}\sqrt{\frac{k(n)}{n}}\right) = o(1)$. Let Assumption 3.1, equation (3.9) and Assumptions A.1 - A.4 hold. Then:*

$$\sqrt{n}\frac{\phi(\widehat{h}_n) - \phi(h_0)}{||v_n^*||_{sd}} \Rightarrow N(0,1) \text{ with } ||v_n^*||_{sd}^2 = \left(\frac{d\phi(h_0)}{dh}[q^{k(n)}(\cdot)]\right)' D_n^{-1} \mho_n D_n^{-1} \left(\frac{d\phi(h_0)}{dh}[q^{k(n)}(\cdot)]\right), \quad (3.17)$$

$$D_n = E\left[\left(T[q^{k(n)}(\cdot)']\right)' \left(T[q^{k(n)}(\cdot)']\right)\right] \text{ and } \mho_n = E\left[\left(T[q^{k(n)}(\cdot)']\right)' \rho(Z, h_0)^2 \left(T[q^{k(n)}(\cdot)']\right)\right].$$

Remark 3.3 includes the NPIV and NPQIV examples in Subsection 2.2 as special cases. In particular, the sieve variance expression (3.17) reproduces the one for the NPIV model (2.18) with $T[q^{k(n)}(\cdot)'] = E[q^{k(n)}(Y_2)'|X]$, and the one for the NPQIV model (2.21) with $T[q^{k(n)}(\cdot)'] = E[f_{U|Y_2,X}(0)q^{k(n)}(Y_2)'|X]$.

By the result in Chen and Pouzo (2012a), the sieve dimension k_n^* satisfying $\{k_n^*\}^{-\varsigma} \asymp \mu_{k_n^*}^{-1} \times \sqrt{\frac{k_n^*}{n}}$ leads to the nonparametric optimal convergence rate of $||\widehat{h} - h_0||_s = O_{P_{Z^\infty}}(\delta_{s,n}^*) = o(1)$ in strong norm, where $\delta_{s,n}^* \asymp \{k_n^*\}^{-\varsigma}$. In particular, $k_n^* \asymp n^{\frac{1}{2(\varsigma+a)+1}}$ and $\delta_{s,n}^* = n^{-\frac{\varsigma}{2(\varsigma+a)+1}}$ for the *mildly ill-posed case* $\mu_k \asymp k^{-a}$ for a finite $a > 0$; and $\delta_{s,n}^* = \{\ln n\}^{-\varsigma}$ for the *severely ill-posed case* $\mu_k \asymp \exp\{-0.5ak\}$ for a finite $a > 0$. However this paper aims at simple valid inferences on functional $\phi(h_0)$. As will be illustrated in the next subsection, although the nonparametric optimal choice k_n^* is compatible with the sufficient conditions for the asymptotic normality of $\sqrt{n}(\phi(\widehat{h}) - \phi(h_0))$ for a regular linear functional $\phi(h_0)$ (see Remark 3.1), it is typically ruled out by Assumption 3.1(iii) for irregular functionals.

3.5.1 Verification of Assumption 3.1

Let $b_j \equiv \frac{d\phi(h_0)}{dh}[\psi_j(\cdot)]$ for all j. By Lemma 3.3 $D_n = E\left[\left(T[q^{k(n)}(\cdot)']\right)'\left(T[q^{k(n)}(\cdot)']\right)\right] = Diag\left\{\mu_1^2, ..., \mu_{k(n)}^2\right\}$ and

$$||v_n^*||^2 = \left(\frac{d\phi(h_0)}{dh}[q^{k(n)}(\cdot)]\right)' D_n^{-1}\left(\frac{d\phi(h_0)}{dh}[q^{k(n)}(\cdot)]\right) = \sum_{j=1}^{k(n)} \mu_j^{-2} b_j^2. \tag{3.18}$$

By Lemma 3.1, $\phi(h)$ of the model (3.15) is regular (at $h = h_0$) iff $\sum_{j=1}^{\infty} \mu_j^{-2} b_j^2 < \infty$, and is irregular (at $h = h_0$) iff $\sum_{j=1}^{\infty} \mu_j^{-2} b_j^2 = \infty$.

For the same functional $\phi(h)$ of a model (3.19) without endogeneity:

$$E[\rho(Y_1; h_0(Y_2))|Y_2] = 0 \quad a.s. - Y_2, \tag{3.19}$$

we have $D_n \asymp I_{k(n)}$ and $||v_n^*||^2 \asymp \sum_{j=1}^{k(n)} b_j^2$. Thus, $\phi(h)$ of the model (3.19) is regular (or irregular) iff $\sum_{j=1}^{\infty} b_j^2 < \infty$ (or $= \infty$).

Since $\mu_{k(n)} \to 0$ as $k(n) \to \infty$, if a functional $\phi(h)$ is irregular for the model (3.19) without endogeneity, then it is irregular for the model (3.15). But, even if a functional $\phi(h)$ is regular for the model (3.19) without endogeneity, it could still be irregular for the model (3.15) with endogeneity.

Linear functionals of the model (3.15) For a linear functional $\phi(h)$ of the model (3.15), given relation (3.16), Assumption 3.1 is satisfied provided that the sieve dimension $k(n)$ satisfies (3.20):

$$\frac{||v_n^*||}{\sqrt{n}} = o(1) \quad \text{and} \quad \sqrt{n}\frac{\left|\frac{d\phi(h_0)}{dh}[\Pi_n h_0 - h_0]\right|}{||v_n^*||} = o(1). \tag{3.20}$$

When $\phi(h)$ of the model (3.15) is regular, Remark 3.1 implies that (3.20) is satisfied provided

$$\sum_{j=1}^{\infty} \mu_j^{-2} b_j^2 < \infty \quad \text{and} \quad n \times \sum_{j=k(n)+1}^{\infty} \mu_j^{-2} b_j^2 \times ||\Pi_n h_0 - h_0||^2 = o(1). \tag{3.21}$$

We shall illustrate below that both these sufficient conditions allow for severely ill-posed problems.

Example 1 (evaluation functional). For $\phi(h) = h(\overline{y}_2)$, we have: $||v_n^*||^2 = \sum_{j=1}^{k(n)} \mu_j^{-2}[\psi_j(\overline{y}_2)]^2$,

$$\left|\frac{d\phi(h_0)}{dh}[\Pi_n h_0 - h_0]\right| = |(\Pi_n h_0)(\overline{y}_2) - h_0(\overline{y}_2)| \leq ||\Pi_n h_0 - h_0||_\infty \leq const.\{k(n)\}^{-\varsigma}.$$

To provide concrete sufficient condition for (3.20), we assume $||v_n^*||^2 \asymp E\left(\sum_{j=1}^{k(n)} \mu_j^{-2}[\psi_j(Y_2)]^2\right) = \sum_{k=1}^{k(n)} \mu_k^{-2}$. Since $\lim_{k(n)\to\infty} ||v_n^*||^2 = \infty$, the evaluation functional is irregular. Condition (3.20) is satisfied provided that

$$\frac{||v_n^*||^2}{n} = \frac{\sum_{k=1}^{k(n)} \mu_k^{-2}}{n} = o(1) \quad \text{and} \quad \frac{\{k(n)\}^{-2\varsigma}}{\frac{1}{n}||v_n^*||^2} = \frac{\{k(n)\}^{-2\varsigma}}{\frac{1}{n}\sum_{k=1}^{k(n)} \mu_k^{-2}} = o(1). \tag{3.22}$$

Condition (3.22) allows for both mildly and severely ill-posed cases.

(a) *Mildly ill-posed*: $\mu_k \asymp k^{-a}$ for a finite $a > 0$. Then $||v_n^*||^2 \asymp \{k(n)\}^{2a+1}$. Condition (3.22) is satisfied by a wide range of sieve dimensions, such as $k(n) \asymp n^{\frac{1}{2(\varsigma+a)+1}}(\ln\ln n)^{\varpi}$ or $n^{\frac{1}{2(\varsigma+a)+1}}(\ln n)^{\varpi}$ for any finite $\varpi > 0$, or $k(n) \asymp n^{\epsilon}$ for any $\epsilon \in (\frac{1}{2(\varsigma+a)+1}, \frac{1}{2a+1})$. Note that any $k(n)$ satisfying Condition (3.22) also ensures $\delta_{s,n} = o(1)$. However, it does require $k(n)/k_n^* \to \infty$, where $k_n^* \asymp n^{\frac{1}{2(\varsigma+a)+1}}$ is the choice for the nonparametric optimal convergence rate in strong norm.

(b) *Severely ill-posed*: $\mu_k \asymp \exp\{-0.5ak\}$ for a finite $a > 0$. Then $||v_n^*||^2 \asymp \exp\{ak(n)\}$. Condition (3.22) is satisfied with $k(n) \asymp a^{-1}[\ln n - \varpi \ln(\ln n)]$ for $0 < \varpi < 2\varsigma$. In addition we need $\varpi > 1$ (and hence $\varsigma > 1/2$) to ensure $\delta_{s,n} = O\left(\{k(n)\}^{-\varsigma} + \mu_{k(n)}^{-1}\sqrt{\frac{k(n)}{n}}\right) = o(1)$.

Example 2 (weighted derivative functional). For $\phi(h) = \int w(y)\nabla h(y)dy$, where $w(y)$ is a weight satisfying the integration by part formula: $\phi(h) = \int w(y)\nabla h(y)dy = -\int h(y)\nabla w(y)dy$, we have: $||v_n^*||^2 = \sum_{j=1}^{k(n)} \mu_j^{-2} b_j^2$ with $b_j = \int \psi_j(y)\nabla w(y)dy$ for all j, and

$$\left|\frac{d\phi(h_0)}{dh}[\Pi_n h_0 - h_0]\right| = \left|\int [\Pi_n h_0(y) - h_0(y)]\nabla w(y)dy\right| \le C \times ||\Pi_n h_0 - h_0||_{L^2(f_{Y_2})} \le const.\{k(n)\}^{-\varsigma}$$

provided that $E\left(\left[\frac{\nabla w(Y_2)}{f_{Y_2}(Y_2)}\right]^2\right) = \sum_{j=1}^{\infty} b_j^2 = C < \infty$. That is, the weighted derivative is assumed to be regular for the model (3.19) without endogeneity.

(i) When the weighted derivative is regular (i.e., $\sum_{j=1}^{\infty} \mu_j^{-2} b_j^2 < \infty$) for the model (3.15), Condition (3.21) is satisfied provided that $n \times \sum_{j=k(n)+1}^{\infty} \mu_j^{-2} b_j^2 \times \delta_n^2 = o(1)$, which is the condition imposed in Ai and Chen (2007) for their root-n estimation of an average derivative of NPIV example, and is shown to allow for severely ill-posed inverse case in Ai and Chen (2007).

(ii) When the weighted derivative is irregular (i.e., $\sum_{j=1}^{\infty} \mu_j^{-2} b_j^2 = \infty$) for the model (3.15), Condition (3.20) is satisfied provided that

$$\frac{||v_n^*||^2}{n} = \frac{\sum_{j=1}^{k(n)} \mu_j^{-2} b_j^2}{n} = o(1) \quad \text{and} \quad \frac{\{k(n)\}^{-2\varsigma}}{\frac{1}{n}||v_n^*||^2} = \frac{\{k(n)\}^{-2\varsigma}}{\frac{1}{n}\sum_{j=1}^{k(n)} \mu_j^{-2} b_j^2} = o(1). \tag{3.23}$$

Condition (3.23) allows for both mildly and severely ill-posed cases. To provide concrete sufficient conditions for (3.23) we assume $b_j^2 \asymp (j\ln(j))^{-1}$ in the following calculations.

(a) *Mildly ill-posed*: $\mu_k \asymp k^{-a}$ for a finite $a > 0$. Then $||v_n^*||^2 \in [c\frac{k(n)^{2a}}{\ln(k(n))}, c'k(n)^{2a}]$ for some $0 < c \le c' < \infty$. Condition (3.23) and $\delta_{s,n} = o(1)$ are jointly satisfied by a wide range of sieve dimensions, such as $k(n) \asymp n^{\frac{1}{2(\varsigma+a)}}(\ln n)^{\varpi}$ for any finite $\varpi > \frac{1}{2(\varsigma+a)}$, or $k(n) \asymp n^{\epsilon}$ for any $\epsilon \in (\frac{1}{2(\varsigma+a)}, \frac{1}{2a+1})$ and $\varsigma > 1/2$.

(b) *Severely ill-posed*: $\mu_k \asymp \exp\{-0.5ak\}$ for $a > 0$. Then $||v_n^*||^2 \in [c\frac{\exp\{ak(n)\}}{k(n)\ln(k(n))}, c'\frac{\exp\{ak(n)\}}{\ln(k(n))}]$ for some $0 < c \le c' < \infty$. Condition (3.23) and $\delta_{s,n} = o(1)$ are jointly satisfied by $k(n) \asymp a^{-1}[\ln(n) - \varpi \ln(\ln(n))]$ for $\varpi \in (1, 2\varsigma - 1)$ and $\varsigma > 1$.

Nonlinear functionals For a nonlinear functional $\phi(h)$ of the model (3.15), Assumption 3.1 is satisfied provided that the sieve dimension $k(n)$ satisfies (3.20) (or (3.21) if $\phi(h)$ is regular) and Assumption 3.1(ii)' to control for the nonlinearity bias. Assumption 3.1(ii)' typically rules out nonlinear regular functionals of severely illposed inverse problems, but allows for nonlinear irregular functionals of severely illposed inverse problems.

Example 3 (weighted quadratic functional). For $\phi(h) = \frac{1}{2}\int w(y)|h(y)|^2 dy$, we have $||v_n^*||^2 = \sum_{j=1}^{k(n)} \mu_j^{-2} b_j^2$ with $b_j = \int h_0(y)w(y)\psi_j(y)dy$ for all j, and

$$\left|\frac{d\phi(h_0)}{dh}[\Pi_n h_0 - h_0]\right| = \left|\int w(y)h_0(y)[\Pi_n h_0(y) - h_0(y)]dy\right| \le const. \times ||\Pi_n h_0 - h_0||_{L^2(f_{Y_2})}$$

provided that $\sup_y \frac{w(y)}{f_{Y_2}(y)} < \infty$. This and $E\left([h_0(Y_2)]^2\right) < \infty$ imply that $\sum_{j=1}^{\infty} b_j^2 < \infty$. That is, the weighted quadratic functional is regular for the model (3.19) without endogeneity. Also,

$$\left|\phi(h) - \phi(h_0) - \frac{d\phi(h_0)}{dh}[h - h_0]\right| = \frac{1}{2}\int w(y)|h(y) - h_0(y)|^2 dy \le const. \times ||h - h_0||^2_{L^2(f_{Y_2})}.$$

(i) When the weighted quadratic functional is regular (i.e., $\sum_{j=1}^{\infty} \mu_j^{-2} b_j^2 < \infty$) for the model (3.15), Condition (3.21) is satisfied provided that $n \times \sum_{j=k(n)+1}^{\infty} \mu_j^{-2} b_j^2 \times \delta_n^2 = o(1)$, which allows for severely ill-posed cases. But Assumption 3.1(ii)' requires that $\sqrt{n} \times \delta_{s,n}^2 = \sqrt{n} \times \left(\{k(n)\}^{-\varsigma} + \mu_{k(n)}^{-1}\sqrt{\frac{k(n)}{n}}\right)^2 = o(1)$, which clearly rules out severely ill-posed inverse case where $\mu_k \asymp \exp\{-0.5ak\}$ for some finite $a > 0$.

(ii) When the weighted quadratic functional is irregular (i.e., $\sum_{j=1}^{\infty} \mu_j^{-2} b_j^2 = \infty$) for the model (3.15), Condition (3.20) is satisfied provided that Condition (3.23) holds with $b_j = \int h_0(y)w(y)\psi_j(y)dy$ for Example 3. Assumption 3.1(ii)' is satisfied provided that

$$\sqrt{n}\frac{\delta_{s,n}^2}{||v_n^*||} = \frac{\sqrt{n} \times \left(\{k(n)\}^{-\varsigma} + \mu_{k(n)}^{-1}\sqrt{\frac{k(n)}{n}}\right)^2}{||v_n^*||} \le n^{-1/2}\frac{\mu_{k(n)}^{-2}k(n)}{\sqrt{\sum_{j=1}^{k(n)} \mu_j^{-2} b_j^2}} = o(1). \quad (3.24)$$

Any $k(n)$ satisfying Conditions (3.23) and (3.24) automatically satisfies $\delta_{s,n} = o(1)$. In addition, both conditions allow for mildly and severely ill-posed cases. To provide concrete sufficient conditions we assume $b_j^2 \asymp (j\ln(j))^{-1}$ in the following calculations.

(a) *Mildly ill-posed*: $\mu_k \asymp k^{-a}$ for a finite $a > 0$. Then $||v_n^*||^2 \in [c\frac{k(n)^{2a}}{\ln(k(n))}, c'k(n)^{2a}]$ for some $0 < c \leq c' < \infty$. Conditions (3.23) and (3.24) are satisfied by a wide range of sieve dimensions, such as $k(n) \asymp n^{\frac{1}{2(\varsigma+a)}}(\ln n)^{\varpi}$ for any finite $\varpi > \frac{1}{2(\varsigma+a)}$, or $k(n) \asymp n^{\epsilon}$ for any $\epsilon \in (\frac{1}{2(\varsigma+a)}, \frac{1}{2a+2})$ and $\varsigma > 1$.

(b) *Severely ill-posed*: $\mu_k \asymp \exp\{-0.5ak\}$ for $a > 0$. Then $||v_n^*||^2 \in [c\frac{\exp\{ak(n)\}}{k(n)\ln(k(n))}, c'\frac{\exp\{ak(n)\}}{\ln(k(n))}]$ for some $0 < c \leq c' < \infty$. Conditions (3.23) and (3.24) are satisfied with $k(n) \asymp a^{-1}[\ln(n) - \varpi\ln(\ln(n))]$ and $\varpi \in (3, 2\varsigma - 1)$ for $\varsigma > 2$.

3.5.2 Verification of Assumption 3.2(i)

By Lemma 3.2, to verify Assumption 3.2(i), it suffices to verify Assumptions A.1 - A.4 in Appendix A. Note that Assumptions A.1 and A.2 do not depend on sieve Riesz representer at all, and have already been verified in Chen and Pouzo (2009), Ai and Chen (2007) and others for (penalized) SMD estimators for the model (3.15). Assumptions A.3 and A.4 do depend on the scaled sieve Riesz representer $u_n^* \equiv v_n^*/||v_n^*||_{sd}$. Both these assumptions are also verified in Ai and Chen (2003), Chen and Pouzo (2009), Ai and Chen (2007) for examples of regular functionals of the model (3.15). Here, we present verifications of Assumptions A.3 and A.4 for irregular functionals of the NPIV and NPQIV examples.

Condition 3.1. *(i)* $\{E[h(Y_2)|\cdot] : h \in \mathcal{H}\} \subseteq \Lambda_c^{\gamma}(\mathcal{X})$, *with* $\gamma > 0.5$*; (ii)* $\sup_{x,y_2} \frac{f_{Y_2X}(y_2,x)}{f_{Y_2}(y_2)f_X(x)} \leq Const. < \infty$.

Proposition 3.1. *Let all conditions for Remark 3.3 hold. Under Condition 3.1, Assumptions A.3 and A.4 hold for the NPIV model (2.18).*

Proposition 3.1 allows for irregular functionals of the NPIV model with severely ill-posed case.

Condition 3.2. *(i)* $\{E[F_{Y_1|Y_2X}(h(Y_2), Y_2, \cdot)|\cdot] : h \in \mathcal{H}\} \subseteq \Lambda_c^{\gamma}(\mathcal{X})$, *with* $\gamma > 0.5$*; (ii)* $\sup_{y_1,y_2,x} |\frac{df_{Y_1|Y_2X}(y_1,y_2,x)}{dy_1}| \leq C < \infty$.

Condition 3.3. $n(\log\log n)^4\delta_{s,n}^4 = o(1)$

Proposition 3.2. *Let all conditions for Remark 3.3 hold. Under conditions 3.1(ii) and 3.2-3.3, Assumptions A.3 and A.4 hold for the NPQIV model (2.21).*

It is clear that Condition 3.3 rules out severely ill-posed case, and hence Proposition 3.2 only allows for irregular functionals of the NPQIV model with mildly ill-posed case.

4 Inference Based on Asymptotic Critical Values

In this section we provide two simple inference procedures for possibly irregular functionals of the general model (1.1). The first one is based on the asymptotic normality Theorem 3.1 with a consistent sieve variance estimator. The second one is based on Theorem 3.2 with the optimally weighted SQLR statistic.

4.1 Consistent estimators of sieve 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 (3.5). We now provide two simple consistent estimators of the sieve variance when the residual function $\rho()$ is pointwise smooth with respect to α_0.

The theoretical sieve Riesz representer v_n^* is unknown but can be estimated easily. Let $\|\cdot\|_{n,M}$ denote the empirical norm induced by the following empirical inner product

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

for any $v_1, v_2 \in \overline{\mathbf{V}}_{k(n)}$, where $M_{n,i}$ is some (almost surely) positive definite weighting matrix.

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

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

and

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

For $\|v_n^*\|_{sd}^2 = E\left(S_{n,i}^* S_{n,i}^{*\prime}\right)$ given in (3.5) we can define two simple plug-in sieve variance estimators, either

$$||\widehat{v}_n^*||_{n,sd}^2 = \frac{1}{n}\sum_{i=1}^{n}\widehat{S}_{n,i}^*\widehat{S}_{n,i}^{*\prime} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n)}{d\alpha}[\widehat{v}_n^*]\right)'\widehat{\Sigma}_i^{-1}\left(\widehat{\rho}_i\widehat{\rho}_i'\right)\widehat{\Sigma}_i^{-1}\left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n)}{d\alpha}[\widehat{v}_n^*]\right) \tag{4.4}$$

with $\widehat{\rho}_i = \rho(Z_i,\widehat{\alpha}_n)$ and $\widehat{\Sigma}_i = \widehat{\Sigma}(X_i)$, or

$$||\widehat{v}_n^*||_{n,sd}^2 = ||\widehat{v}_n^*||_{n,\widehat{\Sigma}^{-1}\widehat{\Sigma}_0\widehat{\Sigma}^{-1}}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n)}{d\alpha}[\widehat{v}_n^*]\right)'\widehat{\Sigma}_i^{-1}\widehat{\Sigma}_{0i}\widehat{\Sigma}_i^{-1}\left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n)}{d\alpha}[\widehat{v}_n^*]\right) \tag{4.5}$$

with $\widehat{\Sigma}_{0i} = \widehat{\Sigma}_0(X_i)$ where $\widehat{\Sigma}_0(x)$ is a consistent estimator of $\Sigma_0(x)$, e.g. $\widehat{E}_n[\rho(Z,\widehat{\alpha}_n)\rho(Z,\widehat{\alpha}_n)' \mid X = x]$, where $\widehat{E}_n[\cdot \mid X = x]$ is some consistent estimator of a conditional mean function of X, such as a series, kernel or local polynomial based estimator.

Let $\langle v_1, v_2\rangle_M \equiv E\left[\left(\frac{dm(X,\alpha_0)}{d\alpha}[v_1]\right)' M \left(\frac{dm(X,\alpha_0)}{d\alpha}[v_2]\right)\right]$. Then $\langle v_1, v_2\rangle_{\Sigma^{-1}} \equiv \langle v_1, v_2\rangle$ and $\langle v_1, v_2\rangle_{\Sigma_0^{-1}} \equiv \langle v_1, v_2\rangle_0$ for all $v_1, v_2 \in \overline{\mathbf{V}}_{k(n)}$. Denote $\overline{\mathbf{V}}^1_{k(n)} \equiv \{v \in \overline{\mathbf{V}}_{k(n)} : ||v|| = 1\}$.

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* $\sup_{v_1,v_2\in\overline{\mathbf{V}}^1_{k(n)}} \left|\langle v_1, v_2\rangle_{n,\Sigma^{-1}} - \langle v_1, v_2\rangle_{\Sigma^{-1}}\right| = o_{P_{Z^\infty}}(1)$;
(iii) $\sup_{x\in\mathcal{X}} ||\widehat{\Sigma}(x) - \Sigma(x)||_e = o_{P_{Z^\infty}}(1)$;
(iv) $\sup_{x\in\mathcal{X}} E\left[\sup_{\alpha\in\mathcal{N}_{osn}} ||\rho(Z,\alpha)\rho(Z,\alpha)' - \rho(Z,\alpha_0)\rho(Z,\alpha_0)'||_e | X = x\right] = o(1)$.

Assumption 4.2. *either (a) or (b) holds:*
(a) $\sup_{v\in\overline{\mathbf{V}}^1_{k(n)}} |\langle v, v\rangle_{n,M} - \langle v, v\rangle_M| = o_{P_{Z^\infty}}(1)$ *with* $M = \Sigma^{-1}\rho(Z,\alpha_0)\rho(Z,\alpha_0)'\Sigma^{-1}$;
(b) (b.i) $\sup_{v\in\overline{\mathbf{V}}^1_{k(n)}} \left|\langle v, v\rangle_{n,\Sigma^{-1}\Sigma_0\Sigma^{-1}} - \langle v, v\rangle_{\Sigma^{-1}\Sigma_0\Sigma^{-1}}\right| = o_{P_{Z^\infty}}(1)$*; and*
(b.ii) $\sup_{\alpha\in\mathcal{N}_{osn}} \sup_{x\in\mathcal{X}} ||\widehat{E}_n[\rho(z,\alpha)\rho(z,\alpha)'|X = x] - E[\rho(z,\alpha)\rho(z,\alpha)'|X = x]||_e = o_{P_{Z^\infty}}(1)$.

Assumption 4.1(i) becomes vacuous if ϕ is linear; otherwise it requires smoothness of the family $\{\frac{d\phi(\alpha)}{d\alpha}[v] : \alpha \in \mathcal{N}_{osn}\}$ uniformly in $v \in \overline{\mathbf{V}}^1_{k(n)}$. Assumption 4.1(ii) implicitly assumes that the residual function $\rho(z,\cdot)$ 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)). Assumption 4.1(iii) assumes the existence of consistent estimators for Σ. In most applications, $\Sigma(\cdot)$ is either completely known (such as the identity matrix) or Σ_0; while $\Sigma_0(x)$ could be consistently estimated via kernel, series LS, local linear regression and other nonparametric procedures.

Theorem 4.1. *Let Assumption 4.1 and assumptions for Lemma 2.2 hold.*

(1) Let Assumption 4.2(a) hold for $||\widehat{v}^*_n||_{n,sd}$ *given in (4.4), or Assumption 4.2(b) hold for* $||\widehat{v}^*_n||_{n,sd}$ *given in (4.5). 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:

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

Theorem 4.1(2) 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 (2.21) example.

Remark 4.1. *Let* $\mathcal{W}_n \equiv \left(\sqrt{n}\frac{\phi(\widehat{\alpha}_n)-\phi_0}{||\widehat{v}_n^*||_{n,sd}}\right)^2 = \left(\widehat{W}_n + \sqrt{n}\frac{\phi(\alpha_0)-\phi_0}{||\widehat{v}_n^*||_{n,sd}}\right)^2$ *be the Wald test statistic. Then Theorem 4.1 (with* $\frac{||v_n^*||_{sd}}{\sqrt{n}} \asymp \frac{||v_n^*||}{\sqrt{n}} = o(1)$*) immediately implies the following results:*
Under $H_0 : \phi(\alpha_0) = \phi_0$, $\mathcal{W}_n = \left(\widehat{W}_n\right)^2 \Rightarrow \chi_1^2$.
Under $H_1 : \phi(\alpha_0) \neq \phi_0$, $\mathcal{W}_n = \left(O_P(1) + \sqrt{n}||v_n^*||_{sd}^{-1}[\phi(\alpha_0) - \phi_0]\,(1 + o_P(1))\right)^2 \to \infty$ *in probability.*
See Theorem A.4 in Appendix A for asymptotic properties of $\mathcal{W}_n$ *under local alternatives.*

4.1.1 Closed form expressions of sieve variance estimators

Under condition stated in Lemma 3.3, $\widehat{v}_n^*$ defined in (4.2-4.3) also has a closed form solution:

$$\widehat{v}_n^* = \overline{\psi}^{k(n)}(\cdot)'\widehat{\gamma}_n^*, \quad \text{and} \quad \widehat{\gamma}_n^* = \widehat{D}_n^{-1}\widehat{F}_n,$$

with $\widehat{D}_n = \frac{1}{n}\sum_{i=1}^n \left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n)}{d\alpha}[\overline{\psi}^{k(n)}(\cdot)']\right)' \widehat{\Sigma}_i^{-1} \left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n)}{d\alpha}[\overline{\psi}^{k(n)}(\cdot)']\right)$ and $\widehat{F}_n = \frac{d\phi(\widehat{\alpha}_n)}{d\alpha}[\overline{\psi}^{k(n)}(\cdot)]$.

Hence the sieve variance estimator given in (4.4) now becomes

$$||\widehat{v}_n^*||_{n,sd}^2 = \widehat{V}_1 \equiv \widehat{F}'_n\widehat{D}_n^{-1}\widehat{\mho}_n\widehat{D}_n^{-1}\widehat{F}_n \quad \text{with} \tag{4.6}$$

$$\widehat{\mho}_n = \frac{1}{n}\sum_{i=1}^n \left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n)}{d\alpha}[\overline{\psi}^{k(n)}(\cdot)']\right)' \widehat{\Sigma}_i^{-1}\left(\widehat{\rho}_i\widehat{\rho}_i'\right)\widehat{\Sigma}_i^{-1}\left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n)}{d\alpha}[\overline{\psi}^{k(n)}(\cdot)']\right).$$

In particular, with $\psi^{k(n)} = q^{k(n)}$ then the sieve variance estimator $||\widehat{v}_n^*||_{n,sd}^2$ given in (4.6) becomes the one given in (2.10) in Subsection 2.2.

Likewise the sieve variance estimator given in (4.5) becomes

$$||\widehat{v}_n^*||_{n,sd}^2 = \widehat{V}_2 \equiv \widehat{F}'_n\widehat{D}_n^{-1}\widehat{\Omega}_n\widehat{D}_n^{-1}\widehat{F}_n \quad \text{with} \tag{4.7}$$

$$\widehat{\Omega}_n = \frac{1}{n}\sum_{i=1}^n \left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n)}{d\alpha}[\overline{\psi}^{k(n)}(\cdot)']\right)' \widehat{\Sigma}_i^{-1}\widehat{\Sigma}_{0i}\widehat{\Sigma}_i^{-1}\left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n)}{d\alpha}[\overline{\psi}^{k(n)}(\cdot)']\right).$$

4.2 Optimally Weighted SQLR

For the specific regular functional $\phi(\alpha) = \lambda'\theta$ of the semi/nonparametric conditional moment model (1.1), Chen and Pouzo (2009) established that the optimally weighted SQLR statistic is asymptotically chi-square distributed under the null. 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 $\Sigma = \Sigma_0$. That is,

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

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 $||v_n^0||_{sd}^2 = Var(S_{n,i}^0) = ||v_n^0||_0^2$ 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 under the null. This result allows us to compute confidence sets for $\phi(\alpha)$ without the need of a consistent variance estimator for $\phi(\widehat{\alpha}_n)$.

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

$$\widehat{||v_n^0||_0^2} \equiv \left(\frac{\widehat{Q}_n(\widetilde{\alpha}_n) - \widehat{Q}_n(\widehat{\alpha}_n)}{\varepsilon_n^2}\right)^{-1}, \tag{4.8}$$

where $\widetilde{\alpha}_n$ is an approximate minimizer of $\widehat{Q}_n(\alpha)$ over $\{\alpha \in \mathcal{A}_{k(n)} \colon \phi(\alpha) = \phi(\widehat{\alpha}_n) - \varepsilon_n\}$.

We now formally state these results. Recall that we use $\widehat{QLR}_n^0(\phi_0)$ to denote the optimally weighted SQLR statistic in Subsection 2.2.

Theorem 4.2. *Let $\widehat{\alpha}_n$ be the optimally weighted PSMD estimator (2.2) with $\Sigma = \Sigma_0$, and conditions for Remark 2.1, Assumptions 3.1 and 3.2 hold with $||v_n^0||_{sd} = ||v_n^0||_0$ and $|B_n - 1| = o_{P_{Z^\infty}}(1)$.*
(1) If $\widehat{\alpha}_n^R \in \mathcal{N}_{osn}$ wpa1-P_{Z^∞}, then: under $H_0 : \phi(\alpha_0) = \phi_0$, $\widehat{QLR}_n^0(\phi_0) = (\sqrt{n}\mathbb{Z}_n)^2 + o_{P_{Z^\infty}}(1) \Rightarrow \chi_1^2$.
(2) Let $cn^{-1/2} \leq \frac{\varepsilon_n}{||v_n^0||_0} \leq C\delta_n$ for finite constants $c, C > 0$. Then: $\widetilde{\alpha}_n \in \mathcal{N}_{osn}$ wpa1-P_{Z^∞}, and

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

Theorem 4.2(1) recommends to construct an asymptotic $100(1-\tau)\%$ confidence set for $\phi(\alpha)$ by inverting the optimally weighted SQLR statistic: $\left\{r \in \mathbb{R} \colon \widehat{QLR}_n^0(r) \leq c_{\chi_1^2}(1-\tau)\right\}$. 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 (4.8) 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 $\rho(Z,\alpha)$ (such as the one in NPQIV), 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 (1.1), allowing for possibly irregular functionals.

5 Inference Based on Generalized Residual 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 the bootstrap sieve Wald and SQLR statistics under virtually the same conditions as those imposed for the original-sample sieve Wald and SQLR statistics.

A bootstrap procedure is described by an array of "weights" $\{\omega_{i,n}\}_{i=1}^n$ for each n, 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 $\{\omega_{i,n}\}_{i=1}^n$ but all satisfy $\omega_{i,n} \geq 0$ and $E[\omega_{i,n}] = 1$. For the time being we assume that $\lim_{n\to\infty} Var(\omega_{i,n}) = \sigma_\omega^2 \in (0,\infty)$ for all i.

In this paper we focus on two types of bootstrap weights:

Assumption Boot.1 (I.i.d Weights)**.** *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 $\int_0^\infty \sqrt{P(|\omega - 1| \geq t)}dt < \infty$.*

The condition $\int_0^\infty \sqrt{P(|\omega - 1| \geq t)}dt < \infty$ is implied by $E[|\omega - 1|^{2+\epsilon}] < \infty$ for some $\epsilon > 0$.

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

We sometimes omit the n subscript from the weight series. 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 p. 458 in Van der Vaart and Wellner (1996) (henceforth, VdV-W). We use these facts in the proofs.

Let $V_i \equiv (Z_i, \omega_{i,n})$ and

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

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 $\rho(Z_i, \alpha)$. In particular, $\widehat{m}^B(x, \alpha) = \sum_{i=1}^n \omega_{i,n}\rho(Z_i, \alpha)A_n(X_i, x)$ for any linear estimator $\widehat{m}(x, \alpha)$ (2.4) of $m(x, \alpha)$. For example, if $\widehat{m}(x, \alpha)$ is a series LS estimator (2.5), then $\widehat{m}^B(x, \alpha)$ is the bootstrap series LS estimator (2.16) defined in Subsection 2.2.

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 $\widehat{Q}_n(\alpha)$, and $\widehat{\alpha}_n^B$ be the bootstrap PSMD estimator, i.e., $\widehat{\alpha}_n^B$ is an approximate minimizer of $\left\{\widehat{Q}_n^B(\alpha) + \lambda_n Pen(h)\right\}$

on $\mathcal{A}_{k(n)}$. Denote $\widehat{\phi}_n \equiv \phi(\widehat{\alpha}_n)$. Then

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

is the (generalized residual) bootstrap SQLR test statistic. And $\mathcal{W}_{1,n}^B \equiv \left(\sqrt{n}\frac{\phi(\widehat{\alpha}_n^B)-\widehat{\phi}_n}{\sigma_\omega ||\widehat{v}_n^*||_{n,sd}}\right)^2$ is one simple bootstrap Wald test statistic (see Subsection 5.3 for other versions based on bootstrap sieve variances).

Additional notation. To be more precise, we introduce some definitions associated with the new random variables $V_i \equiv (Z_i, \omega_{i,n})$ 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 P_{V^∞} be the joint probability over $\mathcal{V}^\infty$. Finally, for each n, let $\mathcal{B}^n$ be the σ-algebra generated by $V^n \equiv Z^n \times (\omega_{1,n}, ..., \omega_{n,n})$, where each $\omega_{i,n}$ acts as a "weight" of Z_i. Let A_n be a random variable that is measurable with respect to $\mathcal{B}^n$, and $\mathcal{L}_{V^\infty|Z^\infty}(A_n|Z^n)$ (or $P_{V^\infty|Z^\infty}(A_n \leq \cdot \mid Z^n)$) be the conditional law (or conditional distribution) of A_n given Z^n. Let B_n be a random variable measurable with respect to $\mathcal{B}_Z^\infty$, and $\mathcal{L}(B_n)$ (or $P_{Z^\infty}(B_n \leq \cdot)$) be the law (or distribution) of B_n. 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

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

(i.e., $\sup_{f \in BL_1} |E[f(A_n)|Z^n] - E[f(B)]| = o_{P_{Z^\infty}}(1)$), 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 chapter 1.12 of VdV-W 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

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

5.1 Consistency and convergence rate of the bootstrap PSMD estimators

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 $\widehat{\alpha}_n$. We also consider a *restricted* bootstrap PSMD estimator, $\widehat{\alpha}_n^{R,B}$, defined as

$$\widehat{Q}_n^B(\widehat{\alpha}_n^{R,B}) + \lambda_n Pen(\widehat{h}_n^{R,B}) \leq \inf_{\alpha \in \mathcal{A}_{k(n)} : \phi(\alpha) = \phi(\widehat{\alpha}_n)} \left\{\widehat{Q}_n^B(\alpha) + \lambda_n Pen(h)\right\} + o_{P_{V^\infty|Z^\infty}}(\frac{1}{n}) \ wpa1(P_{Z^\infty}). \tag{5.1}$$

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)* $\widehat{Q}_n^B(\widehat{\alpha}_n) \leq c_0^* \widehat{Q}_n(\widehat{\alpha}_n) + o_{P_{V^\infty|Z^\infty}}(\frac{1}{n})$ *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 Assumption 5.1 and conditions for Lemma 2.1 hold. Then:*

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

(2) In addition, let Assumption 2.4 hold and $\widehat{Q}_n^B(\alpha) \geq c^* \widehat{Q}_n(\alpha) - O_{P_{V^\infty|Z^\infty}}(\delta^2_{m,n})$ *uniformly over* $\mathcal{A}_{osn}$ *wpa1(P_{Z^∞}). If* $\max\{\delta^2_{m,n}, Q(\Pi_n \alpha_0), \lambda_n, o(n^{-1})\} = \delta^2_{m,n}$*, then:*

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

(3) The above results remain true when $\widehat{\alpha}_n^B$ *is replaced by* $\widehat{\alpha}_n^{R,B}$*.*

Lemma 5.1(2) and (3) show that $\widehat{\alpha}_n^B \in \mathcal{N}_{osn}$ wpa1 and $\widehat{\alpha}_n^{R,B} \in \mathcal{N}_{osn}$ wpa1 regardless of whether the null $H_0 : \phi(\alpha_0) = \phi_0$ is true or not. Again, when $\widehat{m}^B(x, \alpha)$ is the bootstrap series LS estimator (2.16) of $m(x, \alpha)$, under virtually the same 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 $\widehat{Q}_n^B(\alpha) \geq c^* \widehat{Q}_n(\alpha) - O_{P_{V^\infty|Z^\infty}}(\delta^2_{m,n})$ uniformly over $\mathcal{A}_{osn}$ wpa1(P_{Z^∞}).[16]

[16] The verification is amounts to follow the proof of Lemma C.2 of Chen and Pouzo (2012a) except that the

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:*

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

Their theorem is proved under a high level assumption that the first step nonparametric bootstrap estimator $\widehat{h}^B_n$ of unknown function h_0 satisfies $||\widehat{h}^B_n-\widehat{h}_n||=o_{P_{V^\infty|Z^\infty}}\left(n^{-1/4}\right)$ wpa1(P_{Z^∞}). Our Lemmas 2.2 and 5.1 together imply that $||\widehat{h}^B_n-\widehat{h}_n||=O_{P_{V^\infty|Z^\infty}}(\delta_{m,n})$ 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 $\widehat{\Lambda}^B_n(\alpha(t_n),\alpha)\equiv 0.5\{\widehat{Q}^B_n(\alpha(t_n))-\widehat{Q}^B_n(\alpha)\}$ with $\alpha(t_n)\equiv\alpha+t_nu^*_n$. For any sequence of non-negative weights $(b_i)_i$, let

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

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 $r_n(t_n)=\left(\max\{t_n^2,t_nn^{-1/2},o(n^{-1})\}\right)^{-1}$,*

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

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

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

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.

original-sample series LS estimator $\widehat{m}(x,\alpha)$ is replaced by its bootstrap version $\widehat{m}^B(x,\alpha)$.

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* $\widehat{m}^B(\cdot, \alpha)$ *be the bootstrap series LS estimator (2.16), and conditions of Lemmas 3.2 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 3.2 and 5.2(1) indicate that the low level Assumptions A.1 - A.4 are 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 either Assumption Boot.1 or Assumption Boot.2 hold. If there is a positive real sequence* $(b_n)_n$ *such that* $b_n = o\left(\sqrt{n}\right)$ *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 Student t statistic

In this subsection we present two slightly different bootstrap sieve t statistics based on different sieve variance estimators. The first one is $\widehat{W}_{1,n}^B \equiv \sqrt{n}\frac{\phi(\widehat{\alpha}_n^B) - \phi(\widehat{\alpha}_n)}{\sigma_\omega ||\widehat{v}_n^*||_{n,sd}}$. The second one is $\widehat{W}_{2,n}^B \equiv \sqrt{n}\frac{\phi(\widehat{\alpha}_n^B) - \phi(\widehat{\alpha}_n)}{||\widehat{v}_n^*||_{B,sd}}$, where $||\widehat{v}_n^*||_{B,sd}^2$ is a bootstrap sieve variance estimator:

$$||\widehat{v}_n^*||_{B,sd}^2 \equiv \frac{1}{n}\sum_{i=1}^{n}\left(\frac{d\widehat{m}(X_i, \widehat{\alpha}_n)}{d\alpha}[\widehat{v}_n^*]\right)' \widehat{\Sigma}_i^{-1} \varrho(V_i, \widehat{\alpha}_n)\varrho(V_i, \widehat{\alpha}_n)' \widehat{\Sigma}_i^{-1}\left(\frac{d\widehat{m}(X_i, \widehat{\alpha}_n)}{d\alpha}[\widehat{v}_n^*]\right) \tag{5.3}$$

with $\varrho(V_i, \alpha) \equiv (\omega_{i,n} - 1)\rho(Z_i, \alpha) \equiv \rho^B(V_i, \alpha) - \rho(Z_i, \alpha)$ for any α. That is, $||\widehat{v}_n^*||_{B,sd}^2$ is an analog to $||\widehat{v}_n^*||_{n,sd}^2$ defined in (4.4) but using the bootstrapped generalized residual $\varrho(V_i, \widehat{\alpha}_n)$ instead of the original sample fitted residual $\rho(Z_i, \widehat{\alpha}_n)$. One could also define $||\widehat{v}_n^*||_{B,sd}^2$ using $\hat{\Sigma}_{0i}^B = \widehat{E}_n[\varrho(V, \widehat{\alpha}_n)\varrho(V, \widehat{\alpha}_n)'|X = X_i]$ instead of $\varrho(V_i, \widehat{\alpha}_n)\varrho(V_i, \widehat{\alpha}_n)'$, which will be a bootstrap analog to

$||\widehat{v}_n^*||_{n,sd}^2$ defined in (4.5). In Appendix A we provide additional bootstrap sieve t statistic that is based on yet another bootstrap standard errors.

Assumption 5.2. $\sup_{v\in\overline{\mathbf{V}}^1_{k(n)}} |\langle v,v\rangle_{n,\hat{M}^B} - \sigma_\omega^2 \langle v,v\rangle_{n,\hat{M}}| = o_{P_{V^\infty|Z^\infty}}(1)$ *wpa1*(P_{Z^∞}) *with* $\hat{M}_i^B = (\omega_{i,n}-1)^2\hat{M}_i$ *and* $\hat{M}_i = \hat{\Sigma}_i^{-1}\rho(Z_i,\widehat{\alpha}_n)\rho(Z_i,\widehat{\alpha}_n)'\hat{\Sigma}_i^{-1}$.

The following result is a bootstrap version of Theorem 4.1(1).

Theorem 5.1. *Let conditions for Remark 2.1 and Assumptions 4.1, 4.2(a), 5.2, Boot.1 or Boot.2 hold. Then:*

$$\left|\frac{||\widehat{v}_n^*||_{B,sd}}{\sigma_\omega ||v_n^*||_{sd}} - 1\right| = o_{P_{V^\infty|Z^\infty}}(1) \ wpa1(P_{Z^\infty}).$$

Recall that $\widehat{W}_n \equiv \sqrt{n}\frac{\phi(\widehat{\alpha}_n)-\phi(\alpha_0)}{||\widehat{v}_n^*||_{n,sd}}$, whose probability distribution $P_{Z^\infty}\left(\widehat{W}_n \leq \cdot\right)$ converges to the standard normal cdf $\Phi(\cdot)$. The next result is about the consistency of the bootstrap sieve t statistics $\widehat{W}_{j,n}^B$ for $j=1,2$.

Theorem 5.2. *Let $\widehat{\alpha}_n$ be the PSMD estimator (2.2) and $\widehat{\alpha}_n^B$ the bootstrap PSMD estimator. Let conditions for Remark 2.1 and Lemma 5.1 hold. Let Assumptions 3.1, 3.2 and 5.2 hold.*

(1) Let Assumptions 4.1 and 4.2 hold, and Assumption 5.2 hold for $\widehat{W}_{2,n}^B$. Then: for $j=1,2$,

$$\sup_{t\in\mathbb{R}}\left|P_{V^\infty|Z^\infty}\left(\widehat{W}_{j,n}^B \leq t \mid Z^n\right) - P_{Z^\infty}\left(\widehat{W}_n \leq t\right)\right| = o_{P_{V^\infty|Z^\infty}}(1) \ wpa1(P_{Z^\infty}).$$

(2) If ϕ is regular, without imposing Assumptions 4.1, 4.2 and 5.2, we have:

$$\sup_{t\in\mathbb{R}}\left|P_{V^\infty|Z^\infty}\left(\sqrt{n}\frac{\phi(\widehat{\alpha}_n^B)-\phi(\widehat{\alpha}_n)}{\sigma_\omega} \leq t \mid Z^n\right) - P_{Z^\infty}\left(\sqrt{n}\left(\phi(\widehat{\alpha}_n)-\phi(\alpha_0)\right) \leq t\right)\right| = o_{P_{V^\infty|Z^\infty}}(1) \ wpa1(P_{Z^\infty}).$$

For a regular functional, Theorem 5.2(2) provides one way to construct its confidence sets without the need to compute any variance estimator. This extends the result in Chen and Pouzo (2009) for a regular Euclidean parameter $\lambda'\theta$ to a general regular functional $\phi(\alpha)$. Unfortunately for an irregular functional, we need to compute a consistent sieve variance estimator $||\widehat{v}_n^*||_{n,sd}$ or a bootstrap sieve variance estimator $||\widehat{v}_n^*||_{B,sd}^2$ to apply Theorem 5.2(1). Both $||\widehat{v}_n^*||_{n,sd}$ and $||\widehat{v}_n^*||_{B,sd}^2$ are easy to compute when the residual function $\rho(Z_i,\alpha)$ is pointwise smooth in α_0. Note that the bootstrap sieve variance $||\widehat{v}_n^*||_{B,sd}^2$ has a closed form expression: $||\widehat{v}_n^*||_{B,sd}^2 = \widehat{F}'_n\widehat{D}_n^{-1}\widehat{\mho}_n^B\widehat{D}_n^{-1}\widehat{F}_n$ with

$$\widehat{\mho}_n^B = \frac{1}{n}\sum_{i=1}^n\left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n)}{d\alpha}[\overline{\psi}^{k(n)}(\cdot)']\right)'\widehat{\Sigma}_i^{-1}(\omega_{i,n}-1)^2\rho(Z_i,\widehat{\alpha}_n)\rho(Z_i,\widehat{\alpha}_n)'\widehat{\Sigma}_i^{-1}\left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n)}{d\alpha}[\overline{\psi}^{k(n)}(\cdot)']\right)$$

That is, $||\widehat{v}_n^*||_{B,sd}^2$ is computed in the same way as $||\widehat{v}_n^*||_{n,sd}^2 = \widehat{F}'_n\widehat{D}_n^{-1}\widehat{\mho}_n\widehat{D}_n^{-1}\widehat{F}_n$ given in (4.6) except using $\widehat{\mho}_n^B$ instead of $\widehat{\mho}_n$. Since

$$\begin{aligned} E\left(||\widehat{v}_n^*||_{B,sd}^2 \mid Z^n\right) &= \frac{\sigma_\omega^2}{n}\sum_{i=1}^n \left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n)}{d\alpha}[\widehat{v}_n^*]\right)' \widehat{\Sigma}_i^{-1}\rho(Z_i,\widehat{\alpha}_n)\rho(Z_i,\widehat{\alpha}_n)'\widehat{\Sigma}_i^{-1}\left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n)}{d\alpha}[\widehat{v}_n^*]\right) \\ &= \sigma_\omega^2 ||\hat{v}_n^*||_{n,sd}^2 \end{aligned}$$

we suspect that the bootstrap sieve t statistic $\widehat{W}_{2,n}^B = \sqrt{n}\frac{\phi(\widehat{\alpha}_n^B)-\phi(\widehat{\alpha}_n)}{||\widehat{v}_n^*||_{B,sd}}$ might have second order refinement property by choices of bootstrap weights $\{\omega_{i,n}\}$, which will be a subject of future research.

Both bootstrap sieve t statistics $\widehat{W}_{1,n}^B$ and $\widehat{W}_{2,n}^B$ require to compute the original sample PSMD estimator $\widehat{\alpha}_n$ and the bootstrap PSMD estimator $\widehat{\alpha}_n^B$. In Online Appendix D we present a sieve score test and its bootstrap version, which only use the original sample restricted PSMD estimator $\widehat{\alpha}_n^R$ and do not use $\widehat{\alpha}_n^B$, and hence are computationally simple.

Remark 5.2. *Theorems 4.1(2) and 5.2(1) imply that the bootstrap Wald test statistic* $\mathcal{W}_{j,n}^B \equiv \left(\widehat{W}_{j,n}^B\right)^2$, $j=1,2$, *always has the same limiting distribution* χ_1^2 *(conditional on the data) under the null and the alternatives. Let* $\widehat{c}_{j,n}(a)$ *be the* $a-th$ *quantile of the distribution of* $\mathcal{W}_{j,n}^B$ *(conditional on the data* $\{Z_i\}_{i=1}^n$*). Let* $\mathcal{W}_n \equiv \left(\sqrt{n}\frac{\phi(\widehat{\alpha}_n)-\phi_0}{||\widehat{v}_n^*||_{n,sd}}\right)^2$ *be the original sample Wald test statistic. Then Remark 4.1 and Theorem 5.2(1) immediately imply that for* $j=1,2$ *and for any* $\tau\in(0,1)$,

under $H_0:\phi(\alpha_0)=\phi_0$, $\lim_{n\to\infty}\Pr\left(\mathcal{W}_n \geq \widehat{c}_{j,n}(1-\tau)\right)=\tau$;

under $H_1:\phi(\alpha_0)\neq\phi_0$, $\lim_{n\to\infty}\Pr\left(\mathcal{W}_n \geq \widehat{c}_{j,n}(1-\tau)\right)=1$.

See Theorem A.5 in Appendix A for properties under local alternatives.

5.4 Bootstrap SQLR statistic

If $\Sigma\neq\Sigma_0$, the SQLR statistic $\widehat{QLR}_n(\phi_0)=n\left(\widehat{Q}_n(\widehat{\alpha}_n^R)-\widehat{Q}_n(\widehat{\alpha}_n)\right)$ is no longer asymptotically chi-square even under the null; Theorem 3.2, however, implies that the SQLR statistic converges weakly to a tight limit under the null. In this subsection we show that the asymptotic null distribution of the SQLR can be consistently approximated by that of the (generalized residual) bootstrap SQLR statistic $\widehat{QLR}_n^B(\widehat{\phi}_n)$. Recall that

$$\widehat{QLR}_n^B(\widehat{\phi}_n) = n\left(\widehat{Q}_n^B(\widehat{\alpha}_n^{R,B})-\widehat{Q}_n^B(\widehat{\alpha}_n^B)\right)+o_{P_{V^\infty|Z^\infty}}(1)\ wpa1(P_{Z^\infty})$$

where $\widehat{\phi}_n\equiv\phi(\widehat{\alpha}_n)$, and $\widehat{\alpha}_n^{R,B}$ is the restricted bootstrap PSMD estimator (5.1).

Lemma 5.1 implies that $\widehat{\alpha}_n^{R,B}$, $\widehat{\alpha}_n^B\in\mathcal{N}_{osn}$ wpa1 under both the null $H_0:\phi(\alpha_0)=\phi_0$ and the alternatives $H_1:\phi(\alpha_0)\neq\phi_0$. This indicates that the bootstrap SQLR statistic $\widehat{QLR}_n^B(\widehat{\phi}_n)$ is always properly centered and should be stochastically bounded under both the null and the alternatives, as shown in the next theorem. Let $P_{Z^\infty}\left(\widehat{QLR}_n(\phi_0)\leq\cdot\mid H_0\right)$ denote the probability distribution of $\widehat{QLR}_n(\phi_0)$ under the null $H_0 : \phi(\alpha_0) = \phi_0$, which would converge to the cdf of χ^2_1 when $\widehat{QLR}_n(\phi_0) = \widehat{QLR}^0_n(\phi_0)$ (the optimally weighted SQLR).

Theorem 5.3. *Let conditions for Remark 2.1 and Lemma 5.1 hold. Let Assumptions 3.1, 3.2 and 5.2 hold with* $\left|B^\omega_n - ||u^*_n||^2\right| = o_{P_{V^\infty|Z^\infty}}(1)$ *wpa1*(P_{Z^∞})*. Then:*

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

$$\text{(2)}\ \sup_{t\in\mathbb{R}}\left|P_{V^\infty|Z^\infty}\left(\frac{\widehat{QLR}^B_n(\widehat{\phi}_n)}{\sigma^2_\omega} \le t \mid Z^n\right) - P_{Z^\infty}\left(\widehat{QLR}_n(\phi_0) \le t \mid H_0\right)\right| = o_{P_{V^\infty|Z^\infty}}(1)\ wpa1(P_{Z^\infty}).$$

Theorem 5.3 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. We recommend this procedure 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. See, e.g., Andrews and Buchinsky (2000) for a thorough discussion about how to construct CS via bootstrap.

Remark 5.3. *Let* $\widehat{c}_n(a)$ *be the* $a - th$ *quantile of the distribution of* $\frac{\widehat{QLR}^B_n(\widehat{\phi}_n)}{\sigma^2_\omega}$ *(conditional on the data* $\{Z_i\}^n_{i=1}$*). Then Theorems 3.2, 3.3 and 5.3 immediately imply that for any* $\tau \in (0, 1)$*,*

under $H_0 : \phi(\alpha_0) = \phi_0$, $\lim_{n\to\infty} \Pr\left(\widehat{QLR}_n(\phi_0) \ge \widehat{c}_n(1-\tau)\right) = \tau$;

under $H_1 : \phi(\alpha_0) \neq \phi_0$, $\lim_{n\to\infty} \Pr\left(\widehat{QLR}_n(\phi_0) \ge \widehat{c}_n(1-\tau)\right) = 1$.

See Theorem A.3 in Appendix A for properties under local alternatives.

6 Simulation Studies and An Empirical Illustration

In this section, we first present four simulation studies of the PSMD estimation, sieve t and SQLR based confidence sets for the NPQIV and NPIV regressions. We then provide an empirical illustration of the SQLR based confidence sets for the NPQIV Engel curve estimation. We use the series LS estimator (2.5) of $m(X, h)$ 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 the NPQIV model (2.21) and the NPIV model (2.18). MC studies 1 and 2 consider the NPQIV model, while MC studies 3 and 4 are about the NPIV model.

MC Study 1: asymptotic normality of PSMD estimators of NPQIV.

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 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 study, 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: Study 1: MC bias and standard deviation of the PSMD estimator for $\nabla h_0(\mu_2)$.

MC Study 2: SQLR test for an irregular linear functional of NPQIV.

Our second simulation design is based on the MC design of Newey and Powell (2003) and Santos (2012) for a NPIV model, except that we consider a NPQIV model. Specifically, we generate 450 i.i.d. samples of $n = 750$ observations from the NPQIV model (2.21): $Y_1 = 2\sin(\pi Y_2) + 0.76U$, $U = 2(\Phi(U^*) - \gamma)$, $Y_2 = 2(\Phi(Y_2^*/3) - 0.5)$ and $X = 2(\Phi(X^*/3) - 0.5)$, where

$$\begin{bmatrix} Y_2^* \\ X^* \\ U^* \end{bmatrix} \sim N\left(0, \begin{bmatrix} 1 & 0.8 & 0.5 \\ 0.8 & 1 & 0 \\ 0.5 & 0 & 1 \end{bmatrix}\right),$$

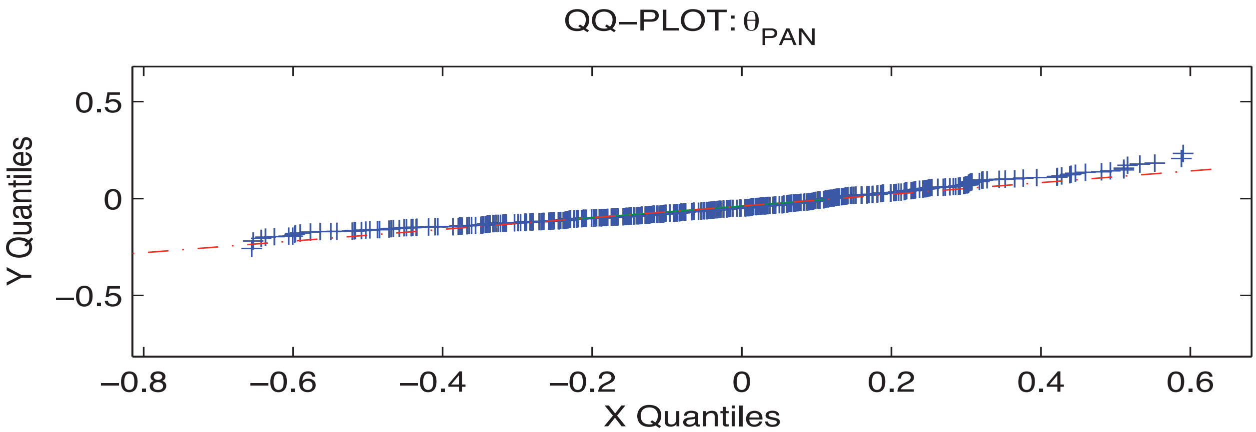

Figure 6.1: Study 1: QQ-Plot for $\nabla\widehat{h}(\mu_2)$ (appropriately centered and scaled), $n = 250$.

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: Study 2: Size of the SQLR test of $\phi(h_0) = 0$.

and finally $h_0(Y_2) = 2\sin(\pi Y_2)$. 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 6.2 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 $\lambda_n = 0.002$. The MC size is close to the norminal size (NS) for all cases.

We also compute the rejection probabilities of the null hypothesis as a function of $r \in \{2/\sqrt{n}, 4/\sqrt{n}\}$, where $r\colon \phi(h_0) = r$; 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 generalized residual bootstrap SQLR corresponding to Column (I), using multinomial bootstrap weights. We employ 450 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.

MC Study 3: sieve variance estimators for an irregular linear functional of NPIV.

This simulation design is the same as that of Newey and Powell (2003) and Santos (2012) for the NPIV model: $Y_1 = h_0(Y_2) + 0.76U$ with $h_0(\cdot) = 2\sin(\pi\cdot)$ (see MC Study 2 for details about the design). The parameter of interest is $\phi(h_0) = h_0(0)$, and the null hypothesis is $H_0\colon \phi(h_0) = 0$. This MC study focuses on the finite sample performance of the sieve variance estimators proposed in Subsection 4.1 for irregular linear functionals.

We generate 5000 i.i.d. samples of $n \in \{750, 1000\}$ observations from the NPIV model. We estimate $h_0(\cdot)$ via the PSMD procedure, using a polynomial spline (P-spline) and polynomial (Pol) sieve $\mathcal{H}_{k(n)}$ for different values of $k(n)$, $Pen(h) = ||h||_{L^2} + ||\nabla h||_{L^2}$ with $\lambda_n = 0.00001$, and $p^{J_n}(X)$ is a P-spline basis[17] for different values of $J_n \geq k(n)$. We compute two sieve variance estimators:

$$\widehat{V}_1 = q^{k(n)}(0)'\widehat{D}_n^{-1}\widehat{\mho}_n\widehat{D}_n^{-1}q^{k(n)}(0) \quad \text{and} \quad \widehat{V}_2 = q^{k(n)}(0)'\widehat{D}_n^{-1}\widehat{\Omega}_n\widehat{D}_n^{-1}q^{k(n)}(0),$$

where $\widehat{D}_n = n^{-1}\left(\widehat{C}_n(P'P)^{-}\widehat{C}_n' + \lambda_{1,n}I_{k(n)}\right)$ for a small $\lambda_{1,n} \in [0, 10^{-5}]$, $\widehat{C}_n \equiv \sum_{i=1}^n q^{k(n)}(Y_{2i})p^{J_n}(X_i)'$, $\widehat{\mho}_n$ is given in (2.20), and $\widehat{\Omega}_n = \frac{1}{n}\widehat{C}_n(P'P)^{-}\left(\sum_{i=1}^n p^{J_n}(X_i)\widehat{\Sigma}_0(X_i)p^{J_n}(X_i)'\right)(P'P)^{-}\widehat{C}_n'$ with $\widehat{U}_j = Y_{1j} - \widehat{h}(Y_{2j})$ and $\widehat{\Sigma}_0(x) = \left(\sum_{j=1}^n \widehat{U}_j^2 p^{J_n}(X_j)'\right)(P'P)^{-}p^{J_n}(x)$.

PSMD \ n	750		1000	
	$\hat{V}_1$	$\hat{V}_2$	$\hat{V}_1$	$\hat{V}_2$
(I)	0.13	0.14	0.08	0.07
(II)	0.10	0.09	0.08	0.08
(III)	0.10	0.10	0.08	0.08
(IV)	0.09	0.09	0.09	0.09
(V)	0.11	0.08	0.09	0.07
(VI)	0.12	0.11	0.09	0.09

Table 6.3: Study 3: Relative performance of $\hat{V}_1$ and $\hat{V}_2$: $Med_{MC}\left[\left|\frac{\hat{V}_j}{||v_n^*||_{sd}^2} - 1\right|\right]$ for $j = 1, 2$.

n	750				1000			
Estimator	$\hat{V}_1$		$\hat{V}_2$		$\hat{V}_1$		$\hat{V}_2$	
Size	5%	10%	5%	10%	5%	10%	5%	10%
(I)	6.6	11.0	6.8	11.2	4.5	8.9	4.6	8.9
(II)	4.9	9.7	5.0	10.0	4.8	9.7	4.9	9.9
(III)	4.3	8.9	4.3	8.9	4.4	9.0	4.3	9.1
(IV)	4.1	8.4	4.7	8.4	4.5	8.8	4.7	8.8
(V)	4.5	9.5	5.7	10.7	4.6	10.0	6.0	11.1
(VI)	4.1	7.9	4.1	8.0	4.0	8.9	4.2	9.0

Table 6.4: Study 3: Nominal size and MC rejection frequencies for t tests $\hat{t}_j$ for $j = 1, 2$.

[17]We also considered other bases such as B-splines and polynomial and results remained essentially the same.

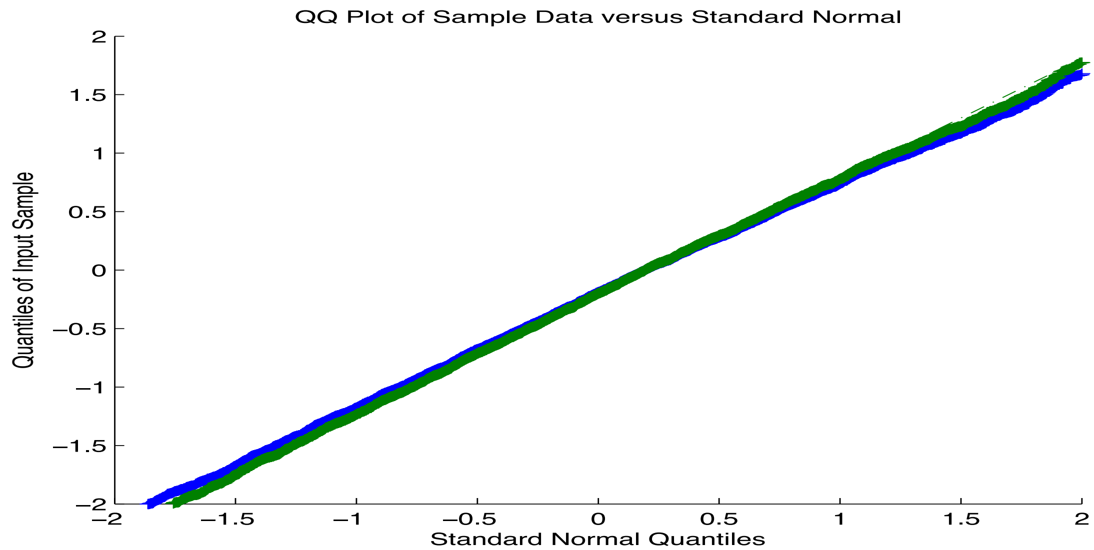

Figure 6.2: Study 3: QQ-Plot for t tests $\hat{t}_j$ for $j = 1, 2$.

Table 6.3 shows $Med_{MC}\left[\left|\frac{\widehat{V}_j}{||v_n^*||_{sd}^2} - 1\right|\right]$ for $j = 1, 2$, where $||v_n^*||_{sd}$ is computed using the MC variance of $\sqrt{n}\widehat{h}_n(0)$ and $Med_{MC}[\cdot]$ is the MC median for different choices of the tuning parameters $(k(n), J_n)$ and bases. Table 6.4 shows the nominal size and MC rejection frequencies of the two sieve t tests $\widehat{t}_j = \sqrt{n}\frac{\widehat{h}_n(0)-0}{\sqrt{\widehat{V}_j}}$ for $j = 1, 2$. Row (I) corresponds to Pol with $k(n) = 4$ for $q^{k(n)}$ and Pol with $J_n = 4$ for p^{J_n}; row (II) corresponds to Pol with $k(n) = 4$ for $q^{k(n)}$ and Pol with $J_n = 5$ for p^{J_n}; row (III) corresponds to Pol with $k(n) = 4$ for $q^{k(n)}$ and Pol with $J_n = 6$ for p^{J_n}; row (IV) corresponds to Pol with $k(n) = 4$ for $q^{k(n)}$ and P-Spline with $J_n = 7$ for p^{J_n}; row (V) corresponds to P-spline with $k(n) = 5$ for $q^{k(n)}$ and pol with $J_n = 7$ for p^{J_n} with $\lambda_n = 0.00002$; row (VI) corresponds to P-spline with $k(n) = 6$ for $q^{k(n)}$ and P-spline with $J_n = 7$ for p^{J_n} with $\lambda_n = 0.00005$. The results seem to behave uniformly well across the different specifications, with the best specification being the one corresponding to rows (II) and (III).

Figure 6.2 shows the QQ-Plot for the sieve t tests $\widehat{t}_j = \sqrt{n}\frac{\widehat{h}_n(0)-0}{\sqrt{\widehat{V}_j}}$ under the null for $j = 1, 2$ and Case (V). Both sieve t tests are almost identical to each other and to the standard normal.

MC Study 4: sieve variance estimators for an irregular nonlinear functional of NPIV. This simulation design is identical to that in MC Study 3, except that the functional of interest is $\phi(h_0) = \exp\{h_0(0)\}$, and the null hypothesis is $H_0\colon \phi(h_0) = 1$. This choice of ϕ allows us to evaluate the finite sample performance of sieve t statistics for an irregular nonlinear functional of a NPIV model. In this MC study, the two sieve t statistics become $\widehat{t}_j = \sqrt{n}\frac{\exp\{\widehat{h}_n(0)\}-\exp\{h_0(0)\}}{\sqrt{\widehat{V}_j}}$ for $j = 1, 2$. Tables 6.5 and 6.6 show the results for cases (I),(II),(IV) and (VI). Overall the results are similar to those in MC Study 3, although the sieve t tests seem to yield slightly lower MC rejection frequencies.

PSMD \ n	750		1000	
	$\hat{V}_1$	$\hat{V}_2$	$\hat{V}_1$	$\hat{V}_2$
(I)	0.18	0.14	0.09	0.10
(II)	0.15	0.15	0.09	0.09
(IV)	0.14	0.14	0.09	0.09
(VI)	0.18	0.17	0.13	0.12

Table 6.5: Study 4: Relative performance of $\hat{V}_1$ and $\hat{V}_2$: $Med_{MC}\left[\left|\frac{\hat{V}_j}{||v_n^*||_{sd}^2} - 1\right|\right]$, $j = 1, 2$ for a nonlinear irregular ϕ.

n	750				1000			
Estimator	$\hat{V}_1$		$\hat{V}_2$		$\hat{V}_1$		$\hat{V}_2$	
Size	5%	10%	5%	10%	5%	10%	5%	10%
(I)	3.4	7.2	4.5	8.5	5.0	9.4	5.6	10.4
(II)	4.0	8.1	4.1	8.2	4.7	8.5	4.7	8.6
(IV)	4.4	9.4	4.6	9.4	4.5	8.4	4.5	8.4
(VI)	3.5	8.0	3.6	8.0	4.0	7.8	4.0	7.9

Table 6.6: Study 4: Nominal size and MC rejection frequencies for t tests $\hat{t}_j$, $j = 1, 2$ for a nonlinear irregular ϕ.

Figure 6.3 shows the QQ-Plot for the two sieve t tests $\widehat{t}_j = \sqrt{n}\frac{\exp\{\widehat{h}_n(0)\}-1}{\sqrt{\widehat{V}_j}}$ under the null for $j = 1, 2$ and Case (IV). Again both t tests are almost identical to each other, except that the quality of the normal approximation is slightly worse than that in Figure 6.2 for a linear irregular ϕ.

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.

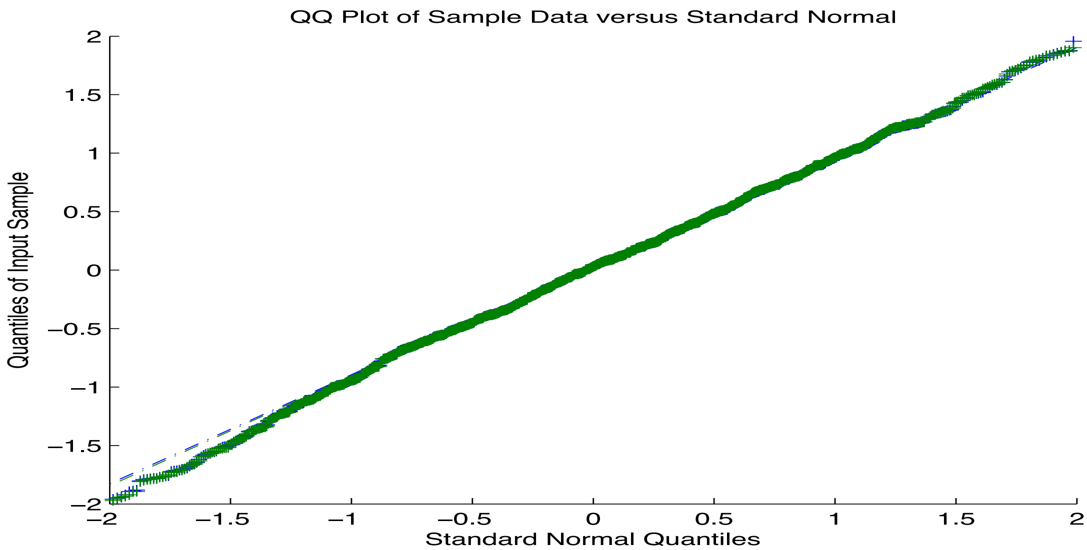

Figure 6.3: Study 4: QQ-Plot for t tests $\hat{t}_j$, $j = 1, 2$ for a nonlinear irregular ϕ.

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.[18] 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 PSMD based inferences on possibly irregular parameters $\phi(\alpha_0)$ of the general semi/nonparametric conditional moment restrictions $E[\rho(Y, X; \alpha_0)|X] = 0$. Under regularity conditions that allow for any consistent nonparametric estimator of the conditional mean function $m(X, \alpha) \equiv E[\rho(Y, X; \alpha)|X]$, 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

[18] 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.

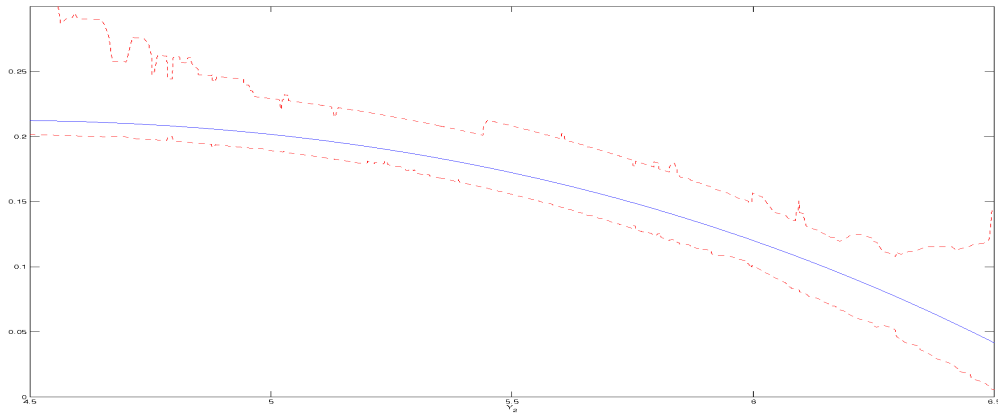

Figure 6.4: PSMD Estimate of the NPQIV food-in Engel curve (blue solid line), with the 95% pointwise confidence bands (red dash lines).

distribution of a possibly non-optimally weighted SQLR statistic under the 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 several simple consistent estimators of sieve variance of $\phi(\widehat{\alpha}_n)$, and establish the asymptotic chi-square distribution of sieve Wald statistics. We also establish local power properties of SQLR and sieve Wald tests in Appendix A. Under conditions that are virtually the same as those for the limiting distributions of the original-sample sieve Wald and SQLR statistics, we establish the consistency of the generalized residual bootstrap sieve Wald and SQLR statistics. All these results are valid regardless of whether $\phi(\alpha_0)$ is regular or not. While SQLR and bootstrap SQLR are useful for models with (pointwise) non-smooth $\rho(Z;\alpha)$, sieve Wald statistic is computationally attractive for models with smooth $\rho(Z;\alpha)$. Monte Carlo studies and an empirical illustration of a nonparametric quantile IV regression demonstrate the good finite sample performance of our inference procedures.

This paper assumes that the semi/nonparametric conditional moment restrictions $E[\rho(Y,X;\alpha_0)|X] = 0$ 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 $E[Y_1 - h_0(Y_2)|X] = 0$ 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 $E[\rho(Y,X;\alpha_0)|X] = 0$.

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*, 5–43.

Ai, C. and X. Chen (2012). Semiparametric efficiency bound for models of sequential moment restrictions containing unknown functions. *Journal of Econometrics 170*, 442–457.

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.

Bontemps, C. and D. Martimort (2013). Identification and estimation of incentive contracts under asymmetric information: An application to the french water sector. *Working paper, Toulouse School of Economics*.

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.

Chamberlain, G. (2010). Binary response models for panel data: Identification and information. *Econometrica 78*, 159–168.

Chen, X., V. Chernozhukov, S. Lee, and W. Newey (2013). Local identification of nonparametric and semiparametric models. *Econometrica, forthcoming*.

Chen, X., J. Favilukis, and S. Ludvigson (2013). An estimation of economic models with recursive preferences. *Quantitative Economics 4*, 39–83.

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*.

Chen, X. and M. Reiß(2011). On rate optimality for nonparametric ill-posed inverse problems in econometrics. *Econometric Theory*.

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.

Davidson, R. and G. MacKinnon (2010). Uniform confidence bands for functions estimated nonparametrically with instrumental variables. *Journal of Business and Economic Statistics 28*, 128–144.

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

Gagliardini, P. and O. Scaillet (2012). Nonparametric instrumental variable estimation of structural quantile effects. *Econometrica 80*(4), 1533–1562.

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 average partial effects in irregular correlated random coefficient panel data models. *Econometrica 80*, 2105–2152.

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 (2012). Uniform confidence bands for functions estimated nonparametrically with instrumental variables. *Journal of Econometrics 168*, 175–188.

Kahn, S. and E. Tamer (2010). Irregular identification, support conditions, and inverse weight estimation. *Econometrica 78*, 2021–2042.

Kline, P. and A. Santos (2012). A score based approach to wild bootstrap inference. *Journal of Econometric Methods 1*, 23–41.

Kosorok, M. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. New York: Springer.

Liao, Y. and W. Jiang (2011). Posterior consistency of nonparametric conditional moment restricted models. *Annals of Statistics 39*, 3003–3031.

Merlo, A. and A. de Paula (2013). Identification and estimation of preference distributions when voters are ideological. *Cemmap working paper, CWP51/13*.

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.

Penaranda, F. and E. Sentana (2013). Duality in mean-variance frontiers with conditioning information. *Working paper, CEMFI, Spain*.

Pinkse, J., M. Slade, and C. Brett (2002). Spatial price competition: a semiparametric approach. *Econometrica 70*, 1111–1153.

Pollard, D. (2001). *A User's Guide to Measure Theoretic Probability*. Cambridge University Press.

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.

Souza-Rodrigues, E. (2012). Demand for deforestation in the amazon. *Working paper, Harvard University Center for the Environment*.

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

A Sufficient Conditions and Additional Results

Appendix A consists of several subsections. Subsection A.1 provides some low level sufficient conditions for the high level LQA assumption 3.2(i) and the bootstrap LQA assumption 5.2(i). Subsection A.2 states useful lemmas when the conditional mean function $m(\cdot,\alpha)$ is estimated by series LS estimators. Subsection A.3 provides consistency theorems for additional bootstrap sieve Student t statistics. Subsection A.4 presents asymptotic properties under local alternatives of the SQLR and the sieve Wald tests, and of their bootstrap versions. Subsection A.5 provides some inference results for functionals of increasing dimension. See online supplemental Appendix C for the proofs of all the results in this Appendix.

A.1 Sufficient conditions for LQA(i) and LQA^B(i) with series LS estimator $\widehat{m}$

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 $g\in\{m(\cdot,\alpha):\alpha\in\mathcal{A}^{M_0}_{k(n)}\}$.*

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 $\sup_{\mathcal{A}^{M_0}_{k(n)}}|\rho(Z,\alpha)|\leq\bar{\rho}_n(Z)$ 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 $\forall\delta>0$,*

$$E\left[\sup_{\alpha\in\mathcal{N}_{0sn}\colon||\alpha-\alpha'||_s\leq\delta}\left\|\rho(Z,\alpha)-\rho(Z,\alpha')\right\|_e^2|X=x\right]\leq K(x)^2\delta^{2\kappa},\ \forall\alpha'\in\mathcal{N}_{osn}\cup\{\alpha_0\}\text{ and all }n,$$

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 $\widetilde{m}(X,\alpha)\equiv\left(\sum_{i=1}^n m(X_i,\alpha)p^{J_n}(X_i)'\right)(P'P)^-p^{J_n}(X)$ 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) $\sup_{\mathcal{N}_{osn}}n^{-1}\sum_{i=1}^n\{||m(X_i,\alpha)||_e^2-E[||m(X_1,\alpha)||_e^2]\}=o_P(n^{-1/2})$;
(iv) $\sup_{\mathcal{N}_{osn}}n^{-1}\sum_{i=1}^n\{g(X_i,u_n^*)m(X_i,\alpha)-E[g(X_1,u_n^*)m(X_1,\alpha)]\}=o_P(n^{-1/2})$.

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* $\alpha_2\in\mathcal{N}_{osn}$*,*

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

Assumptions A.1 and A.2 are comparable to those imposed in Chen and Pouzo (2009) for a non-smooth residual function $\rho(Z,\alpha)$. These assumptions ensure 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 $u_n^*\equiv v_n^*/\|v_n^*\|_{sd}$. 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 (1.1).

A.2 Lemmas for series LS estimator $\widehat{m}(x,\alpha)$ and its bootstrap version

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 version. 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 (2.16). 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 $\tau_n^{-1}\equiv(\delta_n)^2(M_n\delta_{s,n})^{2\kappa}C_n$.

(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 $n\geq N(\delta)$,

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

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 of 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 of 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 Alternative bootstrap sieve t statistic

In this subsection we present additional consistent bootstrap sieve variance estimators and the corresponding bootstrap sieve t statistics. Recall that $\widehat{W}_n\equiv\sqrt{n}\frac{\phi(\widehat{\alpha}_n)-\phi(\alpha_0)}{||\widehat{v}_n^*||_{n,sd}}$ is the original sample sieve t statistic.

Let $\widehat{W}_{3,n}^B\equiv\sqrt{n}\frac{\phi(\widehat{\alpha}_n^B)-\phi(\widehat{\alpha}_n)}{||\widehat{v}_n^B||_{B,sd}}$ where $||\widehat{v}_n^B||_{B,sd}^2$ is a bootstrap sieve variance estimator that is constructed as follows. First, we define

$$||v||_{B,M}^2\equiv n^{-1}\sum_{i=1}^{n}\left(\frac{d\widehat{m}^B(X_i,\widehat{\alpha}_n^B)}{d\alpha}[\cdot]\right)'M_{n,i}\left(\frac{d\widehat{m}^B(X_i,\widehat{\alpha}_n^B)}{d\alpha}[\cdot]\right),$$

where $M_{n,i}$ is some (almost surely) positive definite weighting matrix. Let $\widehat{v}_n^B$ be a *bootstrapped empirical Riesz representer* of the linear functional $\frac{d\phi(\widehat{\alpha}_n^B)}{d\alpha}[\cdot]$ under $||\cdot||_{B,\hat{\Sigma}^{-1}}$. We compute a bootstrap sieve variance estimator as:

$$||\widehat{v}_n^B||_{B,sd}^2\equiv\frac{1}{n}\sum_{i=1}^{n}\left(\frac{d\widehat{m}^B(X_i,\widehat{\alpha}_n^B)}{d\alpha}[\widehat{v}_n^B]\right)'\hat{\Sigma}_i^{-1}\varrho(V_i,\widehat{\alpha}_n^B)\varrho(V_i,\widehat{\alpha}_n^B)'\hat{\Sigma}_i^{-1}\left(\frac{d\widehat{m}^B(X_i,\widehat{\alpha}_n^B)}{d\alpha}[\widehat{v}_n^B]\right)\quad\text{(A.1)}$$

with $\varrho(V_i, \alpha) \equiv (\omega_{i,n} - 1)\rho(Z_i, \alpha) \equiv \rho^B(V_i, \alpha) - \rho(Z_i, \alpha)$ for any α. That is, $||\widehat{v}_n^B||_{B,sd}^2$ is a bootstrap analog to $||\widehat{v}_n^*||_{n,sd}^2$ defined in (4.4). One could also define $||\widehat{v}_n^B||_{B,sd}^2$ using $\widehat{E}_n[\varrho(V, \widehat{\alpha}_n^B)\varrho(V, \widehat{\alpha}_n^B)'|X = X_i]$ instead of $\varrho(V_i, \widehat{\alpha}_n^B)\varrho(V_i, \widehat{\alpha}_n^B)'$, which will be a bootstrap analog to $||\widehat{v}_n^*||_{n,sd}^2$ defined in (4.5). In addition, one could also define $||\widehat{v}_n^B||_{B,sd}^2$ using $\widehat{\alpha}_n$ instead of $\widehat{\alpha}_n^B$. In terms of the first order asymptotic approximation, this alternative definition yields the same asymptotic results. Due to space considerations, we omit these alternative bootstrap sieve variance estimators.

Recall that $\hat{M}_i^B = (\omega_{i,n} - 1)^2\hat{M}_i$ and $\hat{M}_i = \hat{\Sigma}_i^{-1}\rho(Z_i, \widehat{\alpha}_n)\rho(Z_i, \widehat{\alpha}_n)'\hat{\Sigma}_i^{-1}$. We impose the following assumption to ensure that $\widehat{V}_{4,B}$ is a consistent estimator of $\sigma_\omega^2||\widehat{v}_n^*||_{n,sd}^2$ conditional on the original data $\{Z_i\}_{i=1}^n$.

Assumption A.5. *(i)* $\sup_{v_1,v_2\in\overline{\mathbf{V}}_{k(n)}^1} |\langle v_1, v_2\rangle_{B,\Sigma^{-1}} - \langle v_1, v_2\rangle_{n,\Sigma^{-1}}| = o_{P_{V^\infty|Z^\infty}}(1)$ *wpa1*(P_{Z^∞})*;*
(ii) $\sup_{v\in\overline{\mathbf{V}}_{k(n)}^1} |\langle v, v\rangle_{B,\hat{M}^B} - \sigma_\omega^2\langle v, v\rangle_{n,\hat{M}}| = o_{P_{V^\infty|Z^\infty}}(1)$ *wpa1*(P_{Z^∞})*;*
(iii) $\sup_{v\in\overline{\mathbf{V}}_{k(n)}^1} n^{-1}\sum_{i=1}^n(\omega_{i,n} - 1)^2 \left\|\frac{d\widehat{m}^B(X_i,\hat{\alpha}_n^B)}{d\alpha}[v]\right\|_e^2 = O_{P_{V^\infty|Z^\infty}}(1)$ *wpa1*(P_{Z^∞})*.*

Assumption A.5(i)(ii) is analogous to assumptions 4.1(ii) and 4.2(a). Assumption A.5(iii) is a mild assumption that follows from the other assumptions in the theorem if $|\omega_{i,n}| \leq C < \infty$ for all i for the IID weights case.

The following result is a bootstrap version of theorem 4.1.

Theorem A.1. *Let Conditions for Theorem 4.1(1) and Lemma 5.1, Assumptions A.5, Boot.1 or Boot.2 hold. Then:*

$$\text{(1)} \quad \left|\frac{||\widehat{v}_n^B||_{B,sd}}{\sigma_\omega||v_n^*||_{sd}} - 1\right| = o_{P_{V^\infty|Z^\infty}}(1) \ wpa1(P_{Z^\infty}).$$

(2) If further, conditions for Theorem 5.2(1) hold, then:

$$\widehat{W}_{3,n}^B = -\sqrt{n}\frac{\mathbb{Z}_n^{\omega-1}}{\sigma_\omega} + o_{P_{V^\infty|Z^\infty}}(1) \ wpa1(P_{Z^\infty}),$$

$$\left|\mathcal{L}_{V^\infty|Z^\infty}\left(\widehat{W}_{3,n}^B \mid Z^n\right) - \mathcal{L}\left(\widehat{W}_n\right)\right| = o_{P_{Z^\infty}}(1), \quad \textit{and}$$

$$\sup_{t\in\mathbb{R}}\left|P_{V^\infty|Z^\infty}(\widehat{W}_{3,n}^B \leq t|Z^n) - P_{Z^\infty}(\widehat{W}_n \leq t)\right| = o_{P_{V^\infty|Z^\infty}}(1) \ wpa1(P_{Z^\infty}).$$

This bootstrap sieve variance estimator $||\widehat{v}_n^B||_{B,sd}^2$ also has a closed form expression: $||\widehat{v}_n^B||_{B,sd}^2 = (\widehat{F}_n^B)'(\widehat{D}_n^B)^{-1}\widehat{\mho}_{3,n}^B(\widehat{D}_n^B)^{-1}\widehat{F}_n^B$ with

$$\begin{aligned} \widehat{F}_n^B &= \frac{d\phi(\widehat{\alpha}_n^B)}{d\alpha}[\overline{\psi}^{k(n)}(\cdot)'], \ \widehat{D}_n^B = \frac{1}{n}\sum_{i=1}^n\left(\frac{d\widehat{m}^B(X_i, \widehat{\alpha}_n^B)}{d\alpha}[\overline{\psi}^{k(n)}(\cdot)']\right)'\hat{\Sigma}_i^{-1}\left(\frac{d\widehat{m}^B(X_i, \widehat{\alpha}_n^B)}{d\alpha}[\overline{\psi}^{k(n)}(\cdot)']\right), \\ \widehat{\mho}_{3,n}^B &= \frac{1}{n}\sum_{i=1}^n\left(\frac{d\widehat{m}^B(X_i, \widehat{\alpha}_n^B)}{d\alpha}[\overline{\psi}^{k(n)}(\cdot)']\right)'\hat{\Sigma}_i^{-1}(\omega_{i,n} - 1)^2\rho(Z_i, \widehat{\alpha}_n^B)\rho(Z_i, \widehat{\alpha}_n^B)'\hat{\Sigma}_i^{-1}\left(\frac{d\widehat{m}^B(X_i, \widehat{\alpha}_n^B)}{d\alpha}[\overline{\psi}^{k(n)}(\cdot)']\right). \end{aligned}$$

This expression is computed in the same way as $||\widehat{v}_n^*||_{n,sd}^2 = \widehat{F}_n'\widehat{D}_n^{-1}\widehat{\mho}_n\widehat{D}_n^{-1}\widehat{F}_n$ given in (4.6) but using bootstrap analogs. Note that this bootstrap sieve variance only uses $\widehat{\alpha}_n^B$, and is easy to compute.

When specialized to the NPIV model (2.18) in subsection 2.2.1, the expression $||\widehat{v}_n^B||_{B,sd}^2$ simplifies further, with $\widehat{F}_n^B = \frac{d\phi(\widehat{h}_n^B)}{d\alpha}[q^{k(n)}(\cdot)']$, $\widehat{D}_n^B = \frac{1}{n}\widehat{C}_n^B(P'P)^-(\widehat{C}_n^B)'$, $\widehat{C}_n^B = \sum_{j=1}^n \omega_{j,n} q^{k(n)}(Y_{2j})p^{J_n}(X_j)'$,

$$\widehat{\mho}_{3,n}^B = \frac{1}{n}\widehat{C}_n^B(P'P)^-\left(\sum_{i=1}^n p^{J_n}(X_i)[(\omega_{i,n}-1)\widehat{U}_i^B]^2 p^{J_n}(X_i)'\right)(P'P)^-(\widehat{C}_n^B)', \quad \text{with } \widehat{U}_i^B = Y_{1i} - \widehat{h}_n^B(Y_{2i}).$$

This expression is analogous to that for a 2SLS t-bootstrap test; see Davidson and MacKinnon (2010). We leave it to further work to study whether this bootstrap sieve t statistic might have second order refinement by choice of some IID bootstrap weights.

A.4 Asymptotic behaviors under local alternatives

In this subsection we consider the behavior of SQLR, sieve Wald and their bootstrap versions under local alternatives. That is, we consider local alternatives along the curve $\{\boldsymbol{\alpha}_n \in \mathcal{N}_{osn} : n \in \{1,2,...\}\}$, where

$$\boldsymbol{\alpha}_n = \alpha_0 + d_n\Delta_n \quad with \quad \frac{d\phi(\alpha_0)}{d\alpha}[\Delta_n] = \kappa \times (1 + o(1)) \neq 0 \tag{A.2}$$

for any $(d_n, \Delta_n) \in \mathbb{R}_+ \times \overline{\mathbf{V}}_{k(n)}$ such that $d_n||\Delta_n|| \leq M_n\delta_n$, $d_n||\Delta_n||_s \leq M_n\delta_{s,n}$ for all n. The restriction on the rates under both norms is to ensure that the required assumptions for studying the asymptotic behavior under these alternatives (Assumption 3.1 in particular) hold. This choice of local alternatives is to simplify the presentation and could be relaxed somewhat.

Since we are now interested in the behavior of the test statistics under local alternatives, we need to be more explicit about the underlying probability, in a.s. or in probability statements. Henceforth, we use P_{n,Z^∞} to denote the probability measure over sequences Z^∞ induced by the model at $\boldsymbol{\alpha}_n$ (we leave P_{Z^∞} to denote the one associated to α_0).

A.4.1 SQLR and SQLR^B under local alternatives

In this subsection we consider the behavior of the SQLR and the bootstrap SQLR, under local alternatives along the curve $\{\boldsymbol{\alpha}_n \in \mathcal{N}_{osn} : n \in \{1,2,...\}\}$ defined in (A.2).

Theorem A.2. *Let conditions for Remark 2.1 and Proposition B.1 and Assumption 3.2 (with $\left|B_n - ||u_n^*||^2\right| = o_{P_{n,Z^\infty}}(1)$) hold under the local alternatives $\boldsymbol{\alpha}_n$ defined in (A.2). Let Assumption 3.1 hold. Then, under the* local alternatives *$\boldsymbol{\alpha}_n$,*

(1) if $d_n = n^{-1/2}||v_n^*||_{sd}$, then $||u_n^*||^2 \times \widehat{QLR}_n(\phi_0) \Rightarrow \chi_1^2(\kappa^2)$;

(2) if $n^{1/2}||v_n^*||_{sd}^{-1}d_n \to \infty$, then $\lim_{n\to\infty}\left(||u_n^*||^2 \times \widehat{QLR}_n(\phi_0)\right) = \infty$ in probability.

The statement that assumptions hold under the local alternatives $\boldsymbol{\alpha}_n$ really means that the assumptions hold when the true DGP model is indexed by $\boldsymbol{\alpha}_n$ (as opposed to α_0). For instance, this change impacts on Assumption 3.2 by changing the "centering" of the expansion to $\boldsymbol{\alpha}_n$ and also changing "in probability" statements to hold under P_{n,Z^∞} as opposed to P_{Z^∞}.

If we had a likelihood function instead of our criterion function, we could adapt Le Cam's 3rd Lemma to show that Assumption 3.2 under local alternatives holds directly. Since our criterion function is not a likelihood we cannot proceed in this manner, and we directly assume it. Also, if we only consider contiguous alternatives, i.e., curves $\{\boldsymbol{\alpha}_n\}_n$ that yield probability measures P_{n,Z^∞} that are contiguous to P_{Z^∞}, then any statement in a.s. or wpa1 under P_{Z^∞} holds automatically under P_{n,Z^∞}.

The next proposition presents the relative efficiency under local alternatives of tests based on the non- and optimally weighted SQLR statistics. We show —aligned with the literature for regular cases— that optimally weighted SQLR statistic is more efficient than the non-optimally weighted one.

Proposition A.1. *Let all conditions for Theorem A.2 hold. Then, under the local alternatives* $\boldsymbol{\alpha}_n$ *defined in (A.2) with* $d_n = n^{-1/2}||v_n^*||_{sd}$*, we have: for any* t*,*

$$\lim_{n\to\infty} P_{n,Z^\infty}(||u_n^*||^2 \times \widehat{QLR}_n(\phi_0) \geq t) \leq \liminf_{n\to\infty} P_{n,Z^\infty}(\widehat{QLR}_n^0(\phi_0) \geq t).$$

The next theorem shows the consistency of our bootstrap SQLR statistic under the local alternatives $\boldsymbol{\alpha}_n$ in (A.2). This result completes that in Remark 5.3.

Theorem A.3. *Let conditions for Theorem 5.3 hold under local alternatives* $\boldsymbol{\alpha}_n$ *defined in (A.2). Then: (1)*

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

$$\sup_{t\in\mathbb{R}} \left| P_{V^\infty|Z^\infty}\left(\frac{\widehat{QLR}_n^B(\widehat{\phi}_n)}{\sigma_\omega^2} \leq t \mid Z^n\right) - P_{Z^\infty}\left(\widehat{QLR}_n(\phi_0) \leq t \mid H_0\right)\right| = o_{P_{V^\infty|Z^\infty}}(1) \ wpa1(P_{n,Z^\infty}).$$

(2) In addition, let conditions for Theorem A.2 hold. Then: for any $\tau \in (0,1)$*,*
$\tau < \lim_{n\to\infty} P_{n,Z^\infty}\left(\widehat{QLR}_n(\phi_0) \geq \widehat{c}_n(1-\tau)\right) < 1$ *under* $d_n = n^{-1/2}||v_n^*||_{sd}$*;*
$\lim_{n\to\infty} P_{n,Z^\infty}\left(\widehat{QLR}_n(\phi_0) \geq \widehat{c}_n(1-\tau)\right) = 1$ *under* $n^{1/2}||v_n^*||_{sd}^{-1} d_n \to \infty$*,*
where $\widehat{c}_n(a)$ *is the* $a-th$ *quantile of the distribution of* $\frac{\widehat{QLR}_n^B(\widehat{\phi}_n)}{\sigma_\omega^2}$ *(conditional on data* $\{Z_i\}_{i=1}^n$*).*

A.4.2 Sieve Wald and bootstrap sieve Wald tests under local alternatives

The next result establishes the asymptotic behavior of the sieve Wald test statistic $\mathcal{W}_n = \left(\sqrt{n}\frac{\phi(\widehat{\alpha}_n)-\phi_0}{||\widehat{v}_n^*||_{n,sd}}\right)^2$ under the local alternative along the curve $\boldsymbol{\alpha}_n$ defined in (A.2).

Theorem A.4. *Let* $\widehat{\alpha}_n$ *be the PSMD estimator (2.2), conditions for Remark 2.1 and Theorem 4.1 and Assumption 3.2 hold under the local alternatives* $\boldsymbol{\alpha}_n$ *defined in (A.2). Let Assumption 3.1 hold. Then, under the* local alternatives $\boldsymbol{\alpha}_n$*,*
(1) if $d_n = n^{-1/2}||v_n^*||_{sd}$*, then* $\mathcal{W}_n \Rightarrow \chi_1^2(\kappa^2)$*;*
(2) if $n^{1/2}||v_n^*||_{sd}^{-1} d_n \to \infty$*, then* $\lim_{n\to\infty} \mathcal{W}_n = \infty$ *in probability.*

Remark A.1. *By the same proof as that of Proposition A.1, one can establish the asymptotically relative efficiency results for the sieve Wald test statistic.*

The next theorem shows the consistency of our bootstrap sieve Wald test statistic under the local alternatives $\boldsymbol{\alpha}_n$ in (A.2). This result completes that in Remark 5.2.

Theorem A.5. *Let all conditions for Theorem 5.2(1) hold under local alternatives* $\boldsymbol{\alpha}_n$ *defined in (A.2). Then: (1) for* $j = 1, 2$*,*

$$\sup_{t\in\mathbb{R}} \left| P_{V^\infty|Z^\infty}\left(\widehat{W}_{j,n}^B \leq t \mid Z^n\right) - P_{Z^\infty}\left(\widehat{W}_n \leq t\right)\right| = o_{P_{V^\infty|Z^\infty}}(1) \ wpa1(P_{n,Z^\infty}).$$

(2) In addition, let conditions for Theorem A.4 hold. Then: for any $\tau \in (0,1)$,

(2a) If $d_n = n^{-1/2}||v_n^*||_{sd}$ then:

$$P_{n,Z^\infty}\left(\mathcal{W}_n \geq \widehat{c}_{j,n}(1-\tau)\right) = \tau + \Pr\left(\chi_1^2(\kappa^2) \geq \widehat{c}_{j,n}(1-\tau)\right) - \Pr\left(\chi_1^2 \geq \widehat{c}_{j,n}(1-\tau)\right) + o_{P_{V^\infty|Z^\infty}}(1)\ wpa1(P_{n,Z^\infty})$$

and $\tau < \lim_{n\to\infty} P_{n,Z^\infty}\left(\mathcal{W}_n \geq \widehat{c}_{j,n}(1-\tau)\right) < 1$,

(2b) If $\sqrt{n}||v_n^*||_{sd}^{-1} d_n \to \infty$ then: $\lim_{n\to\infty} P_{n,Z^\infty}\left(\mathcal{W}_n \geq \widehat{c}_{j,n}(1-\tau)\right) = 1$.

where $\widehat{c}_{j,n}(a)$ be the $a-th$ quantile of the distribution of $\mathcal{W}_{j,n}^B \equiv \left(\widehat{W}_{j,n}^B\right)^2$ (conditional on the data $\{Z_i\}_{i=1}^n$).

A.5 Local asymptotic theory under increasing dimension of ϕ

In this section we extend some inference results to the case of vector-valued functional ϕ (i.e., $d_\phi \equiv d(n) > 1$). These results would be the basis for uniform confidence bands for nonparametric part, but they are also of independent interest. For instance, Theorem A.7 shows that the Wilks phenomenon extends to our setting, even when $d(n)$ could grow with n.

We first introduce some notation. Let $v_{j,n}^*$ be the sieve Riesz representer corresponding to ϕ_j for $j = 1, ..., d(n)$ and let $\mathbf{v}_n^* \equiv (v_{1,n}^*, ..., v_{d(n),n}^*)$. For each x, we use $\frac{dm(x,\alpha_0)}{d\alpha}[\mathbf{v}_n^*]$ to denote a $d_\rho \times d(n)-$matrix with $\frac{dm(x,\alpha_0)}{d\alpha}[v_{j,n}^*]$ as its $j-$th column for $j = 1, ..., d(n)$. Finally, let

$$\Omega_{sd,n} \equiv E\left[\left(\frac{dm(X,\alpha_0)}{d\alpha}[\mathbf{v}_n^*]\right)' \Sigma^{-1}(X)\Sigma_0(X)\Sigma^{-1}(X)\left(\frac{dm(X,\alpha_0)}{d\alpha}[\mathbf{v}_n^*]\right)\right] \in \mathbb{R}^{d(n)\times d(n)}$$

and

$$\Omega_n \equiv \langle \mathbf{v}_n^{*\prime}, \mathbf{v}_n^* \rangle \equiv E\left[\left(\frac{dm(X,\alpha_0)}{d\alpha}[\mathbf{v}_n^*]\right)' \Sigma^{-1}(X)\left(\frac{dm(X,\alpha_0)}{d\alpha}[\mathbf{v}_n^*]\right)\right] \in \mathbb{R}^{d(n)\times d(n)}.$$

Observe that for $d(n) = 1$, $\Omega_{sd,n} = ||v_n^*||_{sd}^2$ and $\Omega_n = ||v_n^*||^2$. Also, for the case $\Sigma = \Sigma_0$, we would have

$$\Omega_n = \Omega_{sd,n} = \Omega_{0,n} \equiv E\left[\left(\frac{dm(X,\alpha_0)}{d\alpha}[\mathbf{v}_n^*]\right)' \Sigma_0^{-1}(X)\left(\frac{dm(X,\alpha_0)}{d\alpha}[\mathbf{v}_n^*]\right)\right].$$

Let

$$\mathcal{T}_n^M \equiv \{t \in \mathbb{R}^{d(n)} : ||t||_e \leq M_n n^{-1/2}\sqrt{d(n)}\} \quad \text{and} \quad \alpha(t) \equiv \alpha + \mathbf{v}_n^*(\Omega_{sd,n})^{-1/2} t.$$

Let $(c_n)_n$ be a real-valued positive sequence that converges to zero as $n \to \infty$. The following assumption is analogous to Assumption 3.1 but for vector-valued ϕ. Under Assumption 2.1(iv), we could use Ω_n instead of $\Omega_{sd,n}$ in Assumption A.6(ii)(iii) below.

Assumption A.6. *(i) for each $j = 1, ..., d(n)$, $\frac{d\phi_j(\alpha_0)}{d\alpha}$ satisfies Assumption 3.1(i); and for each $v \neq 0$, $\frac{d\phi(\alpha_0)}{d\alpha}[v] \equiv \left(\frac{d\phi_1(\alpha_0)}{d\alpha}[v], ..., \frac{d\phi_{d(n)}(\alpha_0)}{d\alpha}[v]\right)'$ is linearly independent;*

$$\text{(ii)} \quad \sup_{(\alpha,t)\in\mathcal{N}_{osn}\times\mathcal{T}_n^M} \left\| (\Omega_{sd,n})^{-1/2}\left\{\phi(\alpha(t)) - \phi(\alpha_0) - \frac{d\phi(\alpha_0)}{d\alpha}[\alpha(t) - \alpha_0]\right\}\right\|_e = O(c_n);$$

(iii) $\left\|(\Omega_{sd,n})^{-1/2}\frac{d\phi(\alpha_0)}{d\alpha}[\alpha_{0,n} - \alpha_0]\right\|_e = O(c_n)$; (iv) $c_n = o(n^{-1/2})$.

For any $v \in \overline{\mathbf{V}}_{k(n)}$, we use $\langle \mathbf{v}_n^{*\prime}, v\rangle$ to denote a $d(n) \times 1$ vector with components $\langle v_{j,n}^*, v\rangle$ for

$j=1,...,d(n)$. Then $\frac{d\phi(\alpha_0)}{d\alpha}[v]=\langle \mathbf{v}_n^{*\prime}, v\rangle$ with $\frac{d\phi_j(\alpha_0)}{d\alpha}[v]=\langle v_{j,n}^*, v\rangle$ for $j=1,...,d(n)$. Let $\mathbf{Z}_n \equiv (\mathbb{Z}_{1,n}||v_{1,n}^*||_{sd},...,\mathbb{Z}_{d(n),n}||v_{d(n),n}^*||_{sd})'$, where $\mathbb{Z}_{j,n}$ is the notation for $\mathbb{Z}_n$ defined in (3.8) corresponding to the $j-$th sieve Riesz representer.

The next assumption is analogous to Assumption 3.2(i) but for the vector valued case. Let $(a_n,b_n,s_n)_n$ be real-valued positive sequences that converge to zero as $n\to\infty$.

Assumption A.7. *(i) For all n, for all $(\alpha,t)\in\mathcal{N}_{osn}\times\mathcal{T}_n^M$ with $\alpha(t)\in\mathcal{A}_{k(n)}$,*

$$\sup_{(\alpha,t_n)\in\mathcal{N}_{osn}\times\mathcal{T}_n^M} r_n(t_n)\left|\widehat{\Lambda}_n(\alpha(t_n),\alpha)-t_n'(\Omega_{sd,n})^{-1/2}\left\{\mathbf{Z}_n+\langle\mathbf{v}_n^{*\prime},\alpha-\alpha_0\rangle\right\}-t_n'\frac{\mathbb{B}_n}{2}t_n\right|=O_{P_{Z^\infty}}(1)$$

where $r_n(t_n)=\left(\max\{||t_n||_e^2 b_n,||t_n||_e a_n,s_n\}\right)^{-1}$ and $(\mathbb{B}_n)_n$ is such that, for each n, $\mathbb{B}_n$ is a Z^n measurable positive definite matrix in $\mathbb{R}^{d(n)\times d(n)}$ and $\mathbb{B}_n=O_{P_{Z^\infty}}(1)$; (ii) $s_n nd(n)=o(1)$, $b_n\sqrt{d(n)}=o(1)$, $\sqrt{nd(n)}\times a_n=o(1)$.

In the rest of this section as well as in its proofs, since there is no risk of confusion, we use o_P and O_P to denote $o_{P_{Z^\infty}}$ and $O_{P_{Z^\infty}}$ respectively.

The next theorem extends Theorem 3.1 for the sieve Wald statistic to the case of vector-valued functionals ϕ (of increasing dimension). Let $\mu_{3,n}\equiv E\left[\left\|\Omega_{sd,n}^{-1/2}\left(\frac{dm(X,\alpha_0)}{d\alpha}[\mathbf{v}_n^*]\right)'\rho(Z,\alpha_0)\right\|_e^3\right]$.

Theorem A.6. *Let Conditions for Remark 2.1, Assumptions A.6 and A.7 hold. Then:*

(1) $n(\phi(\widehat{\alpha}_n)-\phi(\alpha_0))'\Omega_{sd,n}^{-1}(\phi(\widehat{\alpha}_n)-\phi(\alpha_0))=n\mathbf{Z}_n'\Omega_{sd,n}^{-1}\mathbf{Z}_n+o_P\left(\sqrt{d(n)}\right)$;

(2) for a fixed $d(n)=d$, if $\sqrt{n}\Omega_{sd,n}^{-1/2}\mathbf{Z}_n\Rightarrow N(0,I_d)$ then

$$n(\phi(\widehat{\alpha}_n)-\phi(\alpha_0))'\Omega_{sd,n}^{-1}(\phi(\widehat{\alpha}_n)-\phi(\alpha_0))\Rightarrow\chi_d^2;$$

(3) if $d(n)\to\infty$, $d(n)=o(\sqrt{n}\mu_{3,n}^{-1})$, then:

$$\frac{n(\phi(\widehat{\alpha}_n)-\phi(\alpha_0))'\Omega_{sd,n}^{-1}(\phi(\widehat{\alpha}_n)-\phi(\alpha_0))-d(n)}{\sqrt{2d(n)}}\Rightarrow N(0,1).$$

Theorem A.6(3) essentially states that the asymptotic distribution of $n(\phi(\widehat{\alpha}_n)-\phi(\alpha_0))'\Omega_{sd,n}^{-1}(\phi(\widehat{\alpha}_n)-\phi(\alpha_0))$ is close to $\chi^2_{d(n)}$. Moreover, as $N(d(n),2d(n))$ is close to $\chi^2_{d(n)}$ for large $d(n)$ one could simulate from either distribution. However, since $d(n)$ grows slowly (depends on the rate of $\mu_{3,n}$),[19] it might be more convenient to use $\chi^2_{d(n)}$ in finite samples.

Let

$$\mathbb{D}_n\equiv\Omega_{sd,n}^{1/2}\Omega_n^{-1}\Omega_{sd,n}^{1/2}$$

which, under Assumption 2.1(iv), is bounded in the sense that $\mathbb{D}_n\asymp I_{d(n)}$ (see Lemma C.2 in Appendix C). It is obvious that if $\Sigma=\Sigma_0$ then $\mathbb{D}_n=I_{d(n)}$. Note that $\mathbb{D}_n$ becomes $||u_n^*||^{-2}$ for a scalar-valued functional ϕ.

The next result extends Theorem 3.2 for the SQLR statistic to the case of vector-valued functionals ϕ (of increasing dimension). Recall that $\widehat{QLR}_n^0(\phi_0)$ is the SQLR statistic $\widehat{QLR}_n(\phi_0)$ when $\Sigma=\Sigma_0$.

[19] The condition $d(n)=o(\sqrt{n}\mu_{3,n}^{-1})$ is used for a coupling argument regarding $\Omega_{sd,n}^{-1/2}\sqrt{n}\mathbf{Z}_n$ and a multivariate Gaussian $N(0,I_{d(n)})$. See, e.g., Section 10.4 of Pollard (2001).

Theorem A.7. *Let Conditions for Remark 2.1 and Proposition B.1 (in Appendix B) hold. Let Assumptions A.6 and A.7 hold with* $\max_{t:||t||_e=1} |t'\{\mathbb{B}_n - \mathbb{D}_n^{-1}\}t| = O_P(b_n)$*. Then: under the null hypothesis of* $\phi(\alpha_0) = \phi_0$*,*

(1) $\widehat{QLR}_n(\phi_0) = (\sqrt{n}\Omega_{sd,n}^{-1/2}\mathbf{Z}_n)'\mathbb{D}_n(\sqrt{n}\Omega_{sd,n}^{-1/2}\mathbf{Z}_n) + o_P(\sqrt{d(n)})$*;*

(2) if $\Sigma = \Sigma_0$*, then* $\widehat{QLR}_n^0(\phi_0) = n\mathbf{Z}_n'\Omega_{0,n}^{-1}\mathbf{Z}_n + o_P\left(\sqrt{d(n)}\right)$ *; for a fixed* $d(n) = d$ *if* $\sqrt{n}\Omega_{0,n}^{-1/2}\mathbf{Z}_n \Rightarrow N(0, I_d)$ *then* $\widehat{QLR}_n^0(\phi_0) \Rightarrow \chi_d^2$*;*

(3) if $\Sigma = \Sigma_0$ *and* $d(n) \to \infty$*,* $d(n) = o(\sqrt{n}\mu_{3,n}^{-1})$*, then:* $\frac{\widehat{QLR}_n^0(\phi_0) - d(n)}{\sqrt{2d(n)}} \Rightarrow N(0,1)$.

Theorem A.7(2) is a multivariate version of Theorem 4.2(1). Theorem A.7(3) shows that the optimally weighted SQLR preserves the Wilks phenomenon that is previously shown for the likelihood ratio statistic for semiparametric likelihood models. Again, as $d(n)$ grows slowly with n, Theorem A.7(3) essentially states that the asymptotic null distribution of $\widehat{QLR}_n^0(\phi_0)$ is close to $\chi^2_{d(n)}$.

Given Theorems A.6 and A.7 and their proofs, it is obvious that we can repeat the results on the consistency of the bootstrap SQLR and sieve Wald as well as the local power properties of SQLR and sieve Wald tests to vector-valued ϕ (of increasing dimension). We do not state these results here due to the length of the paper. We suspect that one could slightly improve Assumptions A.6 and A.7 and the coupling condition $d(n) = o(\sqrt{n}\mu_{3,n}^{-1})$ so that the dimension $d(n)$ might grow faster with n, but this will be a subject of future research.