cemmap
centre for microdata methods and practice

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

Xiaohong Chen and Demian Pouzo

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP38/14

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

An ESRC Research Centre

Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE **Tel:** 44 (0)20 7291 4800 **Email:** info@cemmap.ac.uk **www.cemmap.ac.uk**

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

Xiaohong Chen[2] and Demian Pouzo[3]

First version: March 2009. Revised version: August 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. There models are often illposed and hence it is difficult to verify whether a (possibly nonlinear) functional is root-n estimable or not. We provide computationally simple, unified inference procedures that are asymptotically valid regardless of whether a functional is root-n estimable 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 nonlinear) functional; (2) the consistency of simple sieve variance estimators for 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) asymptotic properties of sieve Wald and SQLR for functionals of 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.

[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 and especially A. Santos for helpful comments. We thank Y. Qiu for excellent research assistant in simulations. Chen acknowledges financial support from National Science Foundation grant SES-0838161 and Cowles Foundation. 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. (2014) 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 (possibly nonlinear) functional with a finite $d_\phi \geq 1$. Typical linear 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 $\phi(h) = \int w(y_2)\nabla h(y_2)dy_2$ and many others. Typical nonlinear functionals include a quadratic functional $\int w(y_2) |h(y_2)|^2 dy_2$, a quadratic derivative functional $\int w(y_2) |\nabla h(y_2)|^2 dy_2$, a consumer surplus or an average consumer surplus functional of an endogenous demand function h. 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 consistent 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 non-optimally weighed PSMD procedures for estimation of and inference on $\phi(\alpha_0)$. We show that the non-optimally weighted SQLR statistic

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

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 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). 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 (possibly nonlinear) $\phi(\alpha_0)$ 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 and their bootstrap versions of general nonlinear hypotheses for the model (1.1).

Roughly speaking, our results extend the classical theories on Wald and QLR tests of nonlinear hypothesis based on root-n consistent parametric minimum distance estimator $\widehat{\alpha}_n$ to those based on slower than root-n consistent nonparametric minimum distance estimator $\widehat{\alpha}_n \equiv (\widehat{\theta}'_n, \widehat{h}_n)$ of $\alpha_0 \equiv (\theta'_0, h_0)$ satisfying the model (1.1). The implementations of the sieve Wald and SQLR also resemble the classical Wald and QLR based on parametric extreme estimators and hence are computationally attractive. For example, our sieve t (Wald) test on a general nonlinear hypothesis $\phi(h_0) = \phi_0$ of the NPIV model $E[Y_1 - h_0(Y_2)|X] = 0$ can be implemented as a standard t (Wald) test for a parametric linear IV model using two stage least squares (see Subsection 2.2). The proof techniques are quite different, however, because one is no longer able to rely on the root-n asymptotic normality of $\widehat{\alpha}_n$ and then a standard "delta-method" to establish the asymptotic normality of $\sqrt{n}\,(\phi(\widehat{\alpha}_n) - \phi(\alpha_0))$. In our framework (1.2), $\sqrt{n}\,(\phi(\widehat{\alpha}_n) - \phi(\alpha_0))$ could diverge to infinity under the combined effects of (i) slower convergence rate of $\widehat{\alpha}_n$ to α_0 due to the illposed inverse problem and (ii) nonlinearity in either the functional $\phi()$ or the residual function $\rho()$. Our proof strategy relies on the convergence rates of the PSMD estimator $\widehat{\alpha}_n$ to α_0 in both weak and strong metrics, and then the local curvatures of the functional $\phi()$ and the criterion function under these two metrics. The weak metric is

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

intrinsic to the variance of the linear approximation to $\phi(\widehat{\alpha}_n) - \phi(\alpha_0)$, while the strong metric controls the nonlinearity (in α) of the functional $\phi()$ and of the conditional mean function $m(\cdot, \alpha) = E[\rho(Y, X; \alpha)|X = \cdot]$. Unfortunately the convergence rate in the strong metric could be very slow due to the illposed inverse problem. This explains why it is difficult to establish the asymptotic normality of $\phi(\widehat{\alpha}_n)$ for a nonlinear functional $\phi()$ even in the NPIV model. Our paper builds upon the recent results on convergence rates in Chen and Pouzo (2012a) and others. 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 rates as that of the original-sample PSMD estimator $\widehat{\alpha}_n$. 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.[6]

There are some published work about estimation of and inference on a particular linear functional, the Euclidean parameter $\phi(\alpha) = \theta$, of the general model (1.1) when θ_0 is assumed to be regular (i.e., root-n estimable); see Ai and Chen (2003), Chen and Pouzo (2009), Otsu (2011) and others. None of the existing work allows for irregular θ_0 identified by the model (1.1), however. 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.[7]

When specializing our theory to inference on a particular irregular linear functional, the point evaluation functional $\phi(\alpha) = h(\overline{y}_2)$, of the 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 $\rho(Y_1; \theta_0, h_0(Y_2)) = 1\{Y_1 \leq h_0(Y_2)\} - \gamma$. Previously, 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. Immediately after our paper was first presented in April 2009 Banff/Canada conference on semiparametrics, the authors of Horowitz and Lee (2012) informed us that they were concurrently working on confidence bands for h_0 using a particular SMD estimator of the NPIV example. To the best of our knowledge, there is no inference results, in the existing literature, on any nonlinear functional of h_0 even for the NPIV

[6]The convergence rate of the bootstrap PSMD estimator is also very useful for the consistency of the bootstrap Wald statistic for semiparametric two-step GMM estimation of regular functionals when the first-step unknown functions are estimated via a PSMD procedure. See e.g., Chen et al. (2003)

[7]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. Many applied papers on complicated semi/nonparametric models simply assume that θ_0 is root-n estimable.

and NPQIV examples. Our paper is the first to provide simple sieve Wald and SQLR tests for (possibly) nonlinear functionals satisfying the general semi/nonparametric IV model (1.2).

The rest of the paper is organized as follows. Section 2 presents the plug-in PSMD estimator $\phi(\widehat{\alpha}_n)$ of a (possibly nonlinear) 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 states the basic regularity conditions. Section 4 provides the asymptotic properties of sieve t (Wald) and sieve QLR statistics. Section 5 establishes the consistency of the bootstrap sieve t (Wald) and the bootstrap SQLR statistics. Section 6 verifies the key regularity conditions for the asymptotic theories via the three functionals of the NPIV and NPQIV examples presented in Section 2. Section 7 presents simulation studies and an empirical illustration. Section 8 briefly concludes. Appendix A consists of several subsections, presenting (1) further results on sieve Riesz representation of a functional of interest; (2) the convergence rates of the bootstrap PSMD estimator $\widehat{\alpha}_n^B$ for model (1.1); (3) the local power properties of sieve Wald and SQLR tests and of their bootstrap versions; (4) asymptotic properties of sieve Wald and SQLR for functionals of increasing dimension; (5) low level sufficient conditions with a series least squares (LS) estimated conditional mean function $m(\cdot,\alpha) = E[\rho(Y,X;\alpha)|X=\cdot]$; and (6) additional useful lemmas with series LS estimated $m(\cdot,\alpha)$. Online supplemental materials consist of Appendices B, C and D. Appendix B contains additional theoretical results (including other consistent variance estimators and other bootstrap sieve Wald tests) and proofs of all the results stated in the main text. Appendix C contains proofs of all the results stated in Appendix A. 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 $||A||_W^2 \equiv A'WA$ for a positive definite weighting matrix W. Let $\lambda_{\max}(W)$ and $\lambda_{\min}(W)$ denote the maximal and minimal eigenvalues of W respectively. All random variables $Z \equiv (Y', X')'$, $Z_i \equiv (Y_i', X_i')'$ are defined on a complete probability space $(\mathcal{Z}, \mathcal{B}_Z, P_Z)$, where P_Z is the joint probability distribution of (Y', X'). We define $(\mathcal{Z}^\infty, \mathcal{B}_Z^\infty, P_{Z^\infty})$ as the probability space of the sequences $(Z_1, Z_2, ...)$. For simplicity we assume that Y and X are continuous random variables. Let f_X (F_X) be the marginal density (cdf) of X with support $\mathcal{X}$, and $f_{Y|X}$ $(F_{Y|X})$ be the conditional density (cdf) of Y given X. Let $E_P[\cdot]$ 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 any (possibly random) positive sequences $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$, $a_n = O_P(b_n)$ means that $\lim_{c\to\infty}\limsup_n \Pr(a_n/b_n > c) = 0$; $a_n = o_P(b_n)$ means that for all $\varepsilon > 0$, $\lim_{n\to\infty}\Pr(a_n/b_n > \varepsilon) = 0$; and $a_n \asymp b_n$ means that there exist two constants $0 < c_1 \leq c_2 < \infty$ such that $c_1 a_n \leq b_n \leq c_2 a_n$. Also, we 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. We assume $\dim(\mathcal{A}_{k(n)}) \asymp \dim(\mathcal{H}_{k(n)}) \asymp k(n)$ for simplicity, 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 $F : \mathbf{H}_1 \to \mathbf{H}_2$ between two generic Banach spaces, $\frac{dF(\alpha_0)}{d\alpha}[v] \equiv \left.\frac{\partial F(\alpha_0+\tau v)}{\partial\tau}\right|_{\tau=0}$ is the pathwise (or Gateaux) derivative at α_0 in the direction $v \in \mathbf{H}_1$. And $\frac{dF(\alpha_0)}{d\alpha}[\mathbf{v}'] \equiv \left(\frac{dF(\alpha_0)}{d\alpha}[v_1], \cdots, \frac{dF(\alpha_0)}{d\alpha}[v_k]\right)$ for $\mathbf{v}' = (v_1, \cdots, v_k)$ with $v_j \in \mathbf{H}_1$ for all $j = 1, ..., k$.

2 PSMD Estimation and Inferences: An Overview

2.1 The Penalized Sieve Minimum Distance Estimator

Let $m(X,\alpha) \equiv E[\rho(Y,X;\alpha)|X] = \int \rho(y,X;\alpha)dF_{Y|X}(y)$ be a $d_\rho \times 1$ vector valued conditional mean function, $\Sigma(X)$ be a $d_\rho \times d_\rho$ positive definite ($a.s.-X$) 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

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

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 $\widehat{Q}_n^0(\alpha)$.

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., $\dim(\mathcal{H}_{k(n)}) = k(n) \leq \infty$). 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 a tiny penalty (i.e., $k(n) \to \infty, \frac{k(n)}{n} \to 0$ as $n \to \infty$ and $\lambda_n = o(n^{-1})$, and hence 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 $\{X_j\}_{j=1}^n$ for all x, 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 $P'P$. Following Blundell et al. (2007) and Chen and Pouzo (2009), 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 $J_n \geq d_\theta + k(n) \to \infty$ and $\frac{J_n}{n} \to 0$ as $n \to \infty$.

2.2 Preview of the Main Results for Inference

For simplicity we let $\phi : \mathbb{R}^{d_\theta} \times \mathbf{H} \to \mathbb{R}$ 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, Theorem 4.1 shows 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 resembling the "delta-method" variance for a parametric MD problem:

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

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,

$$\frac{d\phi(\alpha_0)}{d\alpha}[\overline{q}^{k(n)}(\cdot)] \equiv \frac{\partial \phi(\theta_0+\theta, h_0+\beta' q^{k(n)}(\cdot))}{\partial \gamma'}|_{\gamma=0} \equiv \left(\frac{\partial\phi(\alpha_0)}{\partial\theta'}, \frac{d\phi(\alpha_0)}{dh}[q^{k(n)}(\cdot)']\right)' \tag{2.7}$$

and $\gamma \equiv (\theta', \beta')'$ are $(d_\theta + k(n)) \times 1$ vectors, $\frac{d\phi(\alpha_0)}{dh}[q^{k(n)}(\cdot)'] \equiv \frac{\partial\phi(\theta_0, h_0+\beta' q^{k(n)}(\cdot))}{\partial\beta}|_{\beta=0}$, and

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

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

where $\frac{dm(X,\alpha_0)}{d\alpha}[\overline{q}^{k(n)}(\cdot)'] \equiv \frac{\partial E[\rho(Z,\theta_0+\theta,h_0+\beta' q^{k(n)}(\cdot))|X]}{\partial\gamma}|_{\gamma=0}$ is a $d_\rho \times (d_\theta + k(n))$ matrix. The closed form expression of $||v_n^*||_{sd}^2$ immediately leads to simple consistent plug-in sieve variance estimators; one of which is

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

where $\frac{d\phi(\widehat{\alpha}_n)}{d\alpha}[\overline{q}^{k(n)}(\cdot)] \equiv \frac{\partial\phi(\widehat{\theta}_n+\theta,\widehat{h}_n+\beta' q^{k(n)}(\cdot))}{\partial\gamma'}|_{\gamma=0}$ and

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

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

Theorem 4.2 then presents the asymptotic normality of the sieve (Student's) t statistic:[8]

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

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, $\widehat{QLR}_n^0(\phi_0)$, when $\widehat{Q}_n(\alpha)$ is the optimally weighted MD criterion $\widehat{Q}_n^0(\alpha)$. Regardless of whether $\phi(\alpha_0)$ is $\sqrt{n}$ estimable or not, Theorems 4.3(2) and 4.4 show that $\widehat{QLR}_n^0(\phi_0)$ is asymptotically chi-square distributed under the null H_0, and diverges to infinity under the fixed alternatives H_1. Theorem A.1 in Appendix A states that $\widehat{QLR}_n^0(\phi_0)$ is asymptotically noncentral chi-square distributed under local alternatives. One could compute $100(1-\tau)\%$ confidence set for $\phi(\alpha_0)$ as

$$\left\{r\in\mathbb{R}:\ \widehat{QLR}_n^0(r)\le c_{\chi_1^2}(1-\tau)\right\},$$

where $c_{\chi_1^2}(1-\tau)$ is the $(1-\tau)$-th quantile of the χ_1^2 distribution.

Bootstrap sieve QLR statistic. Regardless of whether $\phi(\alpha_0)$ is $\sqrt{n}$ estimable or not, Theorems 4.3(1) and 4.4 establish that the possibly non-optimally weighted SQLR statistic $\widehat{QLR}_n(\phi_0)$

[8]See Theorems 5.2 and A.4 for properties of bootstrap sieve t statistics.

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 $\widehat{QLR}_n^B$ denote a bootstrap SQLR statistic:

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

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

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

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 a sequence of 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. 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.2 establish that under the null H_0, the fixed alternatives H_1 or the local alternatives,[9] 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

[9] See Section A.3 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.

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 7, 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:

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

By Theorem 4.1 $\sqrt{n}\frac{\phi(\widehat{h}_n)-\phi(h_0)}{||v_n^*||_{sd}} \Rightarrow N(0,1)$ with $||v_n^*||_{sd}^2 = \frac{d\phi(h_0)}{dh}[q^{k(n)}(\cdot)]' D_n^- \mho_n D_n^- \frac{d\phi(h_0)}{dh}[q^{k(n)}(\cdot)]$,

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

and $\frac{d\phi(h_0)}{dh}[q^{k(n)}(\cdot)] \equiv \frac{\partial\phi(h_0+\beta' q^{k(n)}(\cdot))}{\partial\beta'}|_{\beta=0}$. For example, for a functional $\phi(h) = h(\overline{y}_2)$, or $= \int w(y)\nabla h(y)dy$ or $= \int w(y)\left|h(y)\right|^2 dy$, 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 $||v_n^*||_{sd}^2 \asymp \frac{d\phi(h_0)}{dh}[q^{k(n)}(\cdot)]' D_n^- \frac{d\phi(h_0)}{dh}[q^{k(n)}(\cdot)]$. 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 $||v_n^*||_{sd,ex}^2 \asymp \frac{d\phi(h_0)}{dh}[q^{k(n)}(\cdot)]' D_{n,ex}^- \frac{d\phi(h_0)}{dh}[q^{k(n)}(\cdot)]$, $D_{n,ex} = E[\{q^{k(n)}(Y_2)\}\{q^{k(n)}(Y_2)\}']$. 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. Under mild conditions on the weight $w()$ 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 Section 6 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 4.2 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:

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

$\frac{d\phi(\widehat{h}_n)}{dh}[q^{k(n)}(\cdot)] \equiv \frac{\partial\phi(\widehat{h}_n+\beta' q^{k(n)}(\cdot))}{\partial\beta'}|_{\beta=0}$ and

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

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

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 7.1 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 4.1 $\sqrt{n}\frac{\phi(\widehat{h}_n)-\phi(h_0)}{||v_n^*||_{sd}} \Rightarrow N(0,1)$ with $||v_n^*||_{sd}^2 = \left(\frac{d\phi(h_0)}{dh}[q^{k(n)}(\cdot)]\right)' D_n^- \left(\frac{d\phi(h_0)}{dh}[q^{k(n)}(\cdot)]\right)$ 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

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

and the sieve variance becomes $||v_n^*||_{sd,ex}^2 = \left(\frac{d\phi(h_0)}{dh}[q^{k(n)}(\cdot)]\right)' D_{n,ex}^- \left(\frac{d\phi(h_0)}{dh}[q^{k(n)}(\cdot)]\right)$ 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 $D_n \le D_{n,ex}$ and $||v_n^*||_{sd}^2 \ge ||v_n^*||_{sd,ex}^2$. Under mild conditions (see, e.g., Chen and Pouzo (2012a), Chen et al. (2014)), $\lambda_{\min}(D_n) \to 0$ 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 Section 6 for details.

Applying Theorem 4.3(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 $\left\{r \in \mathbb{R}\colon\ \widehat{QLR}_n^0(r) \le c_{\chi_1^2}(\tau)\right\}$. See Subsection 7.2 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). See Koenker (2005) for the usefulness of quantile Engel curves. Instead of using the asymptotic critical values, we could also construct a confidence set using the bootstrap critical values as in (2.17).

3 Basic Regularity Conditions

Before we establish asymptotic properties of sieve t (Wald) and SQLR statistics, we need to present three sets of basic regularity conditions. The first set of assumptions allows us to establish the convergence rates of the PSMD estimator $\widehat{\alpha}_n$ to the true parameter value α_0 in both weak and strong metrics, which in turn allows us to concentrate on some shrinking neighborhood of α_0 in the semi/nonparametric model (1.1). The second and third regularity conditions are respectively about the local curvatures of the functional $\phi()$ and of the criterion function under these two metrics. The weak metric $||\cdot||$ is closely related to the variance of the linear approximation to $\phi(\widehat{\alpha}_n) - \phi(\alpha_0)$, while the strong metric $||\cdot||_s$ is used to control the nonlinearity (in α) of the functional $\phi()$ and of the conditional mean function $m(x,\alpha)$. This section is mostly technical and applied researchers could skip this and directly go to the subsequent sections on the asymptotic properties of sieve Wald and SQLR statistics.

3.1 A brief discussion on the convergence rate of the PSMD estimator

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 A.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 $\|\alpha\|_s \equiv \|\theta\|_e + \|h\|_{\mathbf{H}}$.

Assumption 3.1 (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 $\mathcal{H} = cl(\cup_k \mathcal{H}_k)$, 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;*[10] *(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 3.2 (Penalty). *(i) $\lambda_n > 0$, $Q(\Pi_n\alpha_0) + o(n^{-1}) = O(\lambda_n) = o(1)$; (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.*[11]

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

Assumption 3.3 (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) $\widehat{Q}_n(\alpha) \geq cQ(\alpha) - O_{P_{Z^\infty}}(\bar{\delta}^2_{m,n})$ uniformly over $\mathcal{A}^{M_0}_{k(n)}$ for some $\bar{\delta}^2_{m,n} = o(1)$ and a finite constant $c > 0$.*

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

Lemma 3.1. *Let $\widehat{\alpha}_n$ be the PSMD estimator defined in (2.2), and Assumptions 3.1, 3.2 and 3.3 hold. Then: $||\widehat{\alpha}_n - \alpha_0||_s = o_{P_{Z^\infty}}(1)$ and $Pen(\widehat{h}_n) = O_{P_{Z^\infty}}(1)$.*

Given the consistency result, we can restrict our attention to a convex, $||\cdot||_s$−neighborhood around α_0, denoted as $\mathcal{A}_{os}$ such that

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

[10] 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 $\mathcal{A}$.

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

for a positive finite constant M_0 (the existence of a convex $\mathcal{A}_{os}$ is implied by the convexity of $\mathcal{A}$ and quasi-convexity of $Pen(\cdot)$). 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 $\alpha_1,\ \alpha_2 \in \mathcal{A}_{os}$,

$$||\alpha_1-\alpha_2||^2 \equiv E\left[\left(\frac{dm(X,\alpha_0)}{d\alpha}[\alpha_1-\alpha_2]\right)'\Sigma(X)^{-1}\left(\frac{dm(X,\alpha_0)}{d\alpha}[\alpha_1-\alpha_2]\right)\right]; \tag{3.1}$$

$$||\alpha_1-\alpha_2||_0^2 \equiv E\left[\left(\frac{dm(X,\alpha_0)}{d\alpha}[\alpha_1-\alpha_2]\right)'\Sigma_0(X)^{-1}\left(\frac{dm(X,\alpha_0)}{d\alpha}[\alpha_1-\alpha_2]\right)\right]. \tag{3.2}$$

It is clear that, under Assumption 3.1(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 sufficient condition, Assumption 3.1(iv), throughout the paper.

Let $\mathcal{A}_{osn} = \mathcal{A}_{os} \cap \mathcal{A}_{k(n)}$. Let $\{\delta_n\}_{n=1}^{\infty}$ be a sequence of positive real values such that $\delta_n = o(1)$ and $\delta_n \leq \bar{\delta}_{m,n}$.

Assumption 3.4. *(i) There exists a $||\cdot||_s$–neighborhood of α_0, $\mathcal{A}_{os}$, such that* [12] *$\mathcal{A}_{os}$ 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) $Q(\alpha) \asymp ||\alpha-\alpha_0||^2$ for all $\alpha \in \mathcal{A}_{os}$; (iii) $\widehat{Q}_n(\alpha) \geq cQ(\alpha) - O_{P_{Z^\infty}}(\delta_n^2)$ uniformly over $\mathcal{A}_{osn}$, and* $\max\{\delta_n^2, Q(\Pi_n\alpha_0), \lambda_n, o(n^{-1})\} = \delta_n^2$*; (iv) $\lambda_n \times \sup_{\alpha,\alpha' \in \mathcal{A}_{os}} |Pen(h)-Pen(h')| = o(n^{-1})$ or $\lambda_n = o(n^{-1})$.*

Assumption 3.4(ii) is about the local curvature of the population criterion $Q(\alpha)$ at α_0. When $\widehat{Q}_n(\alpha)$ is computed using the series LS estimator (2.5), Lemma C.2 of Chen and Pouzo (2012a) shows that $\widehat{Q}_n(\alpha) \asymp Q(\alpha) - O_{P_{Z^\infty}}(\delta_n^2)$ uniformly over $\mathcal{A}_{osn}$ and hence Assumption 3.4(iii) is satisfied.

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

$$\tau_n \equiv \sup_{\alpha \in \mathcal{A}_{osn}: ||\alpha-\Pi_n\alpha_0|| \neq 0} \frac{||\alpha-\Pi_n\alpha_0||_s}{||\alpha-\Pi_n\alpha_0||}. \tag{3.3}$$

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 $\tau_n =$

[12] Given the consistency result, the PSMD estimator will belong to any $||\cdot||_s$–neighborhood around α_0 with probability approaching one.

$O(\exp\{\frac{a}{2}k(n)\})$ 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.

Lemma 3.2. *Let $\widehat{\alpha}_n$ be the PSMD estimator defined in (2.2), and Assumptions 3.1, 3.2(ii)(iii), 3.3 and 3.4(i)(ii)(iii) hold. Then:*

$$||\widehat{\alpha}_n - \alpha_0|| = O_{P_{Z^\infty}}(\delta_n) \quad \textit{and} \quad ||\widehat{\alpha}_n - \alpha_0||_s = O_{P_{Z^\infty}}(||\alpha_0 - \Pi_n\alpha_0||_s + \tau_n\delta_n).$$

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 δ_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 $m(X,\alpha)$.

Let $\{\delta_{s,n} : n \geq 1\}$ be a sequence of real positive numbers such that $\delta_{s,n} = ||h_0 - \Pi_n h_0||_s + \tau_n\delta_n = o(1)$. Lemma 3.2 implies that $\widehat{\alpha}_n \in \mathcal{N}_{osn} \subseteq \mathcal{N}_{os}$ wpa1-P_{Z^∞}, where

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

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. Assumption 3.4(iv) is not needed for establishing a convergence rate in Lemma 3.2. but, it will be imposed in the rest of the paper so that we can ignore penalty effect in the first order local asymptotic analysis.

3.2 (Sieve) Riesz representation and (sieve) variance

We first introduce a representation of the functional of interest $\phi()$ at α_0 that is crucial for all the subsequent local asymptotic theories. Let $\phi : \mathbb{R}^{d_\theta} \times \mathbf{H} \to \mathbb{R}$ be continuous in $||\cdot||_s$. We assume that $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot] : (\mathbb{R}^{d_\theta} \times \mathbf{H}, ||\cdot||_s) \to \mathbb{R}$ is a $||\cdot||_s-$bounded linear functional (i.e., $\left|\frac{d\phi(\alpha_0)}{d\alpha}[v]\right| \leq c||v||_s$ uniformly over $v \in \mathbb{R}^{d_\theta} \times \mathbf{H}$ for a finite positive constant c), which could be computed as a pathwise (directional) derivative of the functional $\phi(\cdot)$ at α_0 in the direction of $v = \alpha - \alpha_0 \in \mathbb{R}^{d_\theta} \times \mathbf{H}$:

$$\frac{d\phi(\alpha_0)}{d\alpha}[v] = \left.\frac{\partial\phi(\alpha_0 + \tau v)}{\partial\tau}\right|_{\tau=0}.$$

Let $\mathbf{V}$ be a linear span of $\mathcal{A}_{os} - \{\alpha_0\}$, which is endowed with both $||\cdot||_s$ and $||\cdot||$ (in equation (3.1)) norms, and $||v|| \leq C||v||_s$ for all $v \in \mathbf{V}$ (under Assumption 3.4(i)). Let $\overline{\mathbf{V}} \equiv clsp(\mathcal{A}_{os} - \{\alpha_0\})$, 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 $||\cdot||$). It is clear that $(\overline{\mathbf{V}}, ||\cdot||)$ is an infinite dimensional Hilbert space (under Assumptions 3.1(i)(iii)(iv) and 3.4(i)(ii)).

If the linear functional $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ is *bounded* on $(\mathbf{V}, ||\cdot||)$, i.e.

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

then there is a unique extension of $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ from $(\mathbf{V}, ||\cdot||)$ to $(\overline{\mathbf{V}}, ||\cdot||)$, and a unique Riesz representer $v^*\in\overline{\mathbf{V}}$ of $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ on $(\overline{\mathbf{V}}, ||\cdot||)$ such that[13]

$$\frac{d\phi(\alpha_0)}{d\alpha}[v]=\langle v^*, v\rangle \text{ for all } v\in\overline{\mathbf{V}} \quad\text{and}\quad \|v^*\|\equiv\sup_{v\in\overline{\mathbf{V}},v\neq 0}\frac{\left|\frac{d\phi(\alpha_0)}{d\alpha}[v]\right|}{\|v\|}=\sup_{v\in\mathbf{V},v\neq 0}\frac{\left|\frac{d\phi(\alpha_0)}{d\alpha}[v]\right|}{\|v\|}<\infty. \tag{3.4}$$

If $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ is *unbounded* on $(\mathbf{V}, ||\cdot||)$, i.e.

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

then there is no unique extension of the mapping $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ from $(\mathbf{V}, ||\cdot||)$ to $(\overline{\mathbf{V}}, ||\cdot||)$, and nor existing any Riesz representer of $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ on $(\overline{\mathbf{V}}, ||\cdot||)$.

Since $||v||\le C||v||_s$ for all $v\in\mathbf{V}$, it is clear that a $||\cdot||_s-$bounded linear functional $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ could be either bounded or unbounded on $(\mathbf{V}, ||\cdot||)$. As explained in Appendix A, in this paper we also call $\phi()$ *regular* (or *irregular*) at α_0 whenever $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ is bounded (or unbounded) on $(\mathbf{V}, ||\cdot||)$.

Sieve Riesz representation. Let $\alpha_{0,n}\in\mathbb{R}^{d_\theta}\times\mathcal{H}_{k(n)}$ be such that

$$||\alpha_{0,n}-\alpha_0||\equiv\min_{\alpha\in\mathbb{R}^{d_\theta}\times\mathcal{H}_{k(n)}}||\alpha-\alpha_0||. \tag{3.5}$$

Let $\overline{\mathbf{V}}_{k(n)}\equiv clsp\left(\mathcal{A}_{osn}-\{\alpha_{0,n}\}\right)$, where $clsp\,(.)$ denotes the closed linear span under $\|\cdot\|$. Then $\overline{\mathbf{V}}_{k(n)}$ is a finite dimensional Hilbert space under $\|\cdot\|$. Moreover, $\overline{\mathbf{V}}_{k(n)}$ is dense in $\overline{\mathbf{V}}$ under $\|\cdot\|$. To simplify the presentation, we assume that $\dim(\overline{\mathbf{V}}_{k(n)})=\dim(\mathcal{A}_{k(n)})\asymp k(n)$, all of which grow to

[13] See, e.g., page 206-207 and theorem 3.10.1 in Debnath and Mikusinski (1999).

infinity with n. By definition we have $\langle v_n, \alpha_{0,n} - \alpha_0 \rangle = 0$ for all $v_n \in \overline{\mathbf{V}}_{k(n)}$.

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

$$\frac{d\phi(\alpha_0)}{d\alpha}[v] = \langle v_n^*, v \rangle \text{ for all } v \in \overline{\mathbf{V}}_{k(n)} \quad \text{and} \quad \|v_n^*\| \equiv \sup_{v \in \overline{\mathbf{V}}_{k(n)}: \|v\| \neq 0} \frac{\left| \frac{d\phi(\alpha_0)}{d\alpha}[v] \right|}{\|v\|} < \infty. \tag{3.6}$$

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 space $(\mathbf{V}, ||\cdot||)$ or not. Moreover, $v_n^* \in \overline{\mathbf{V}}_{k(n)}$ and its norm $\|v_n^*\|$ can be computed in closed form (see Subsection 4.1.1). The next Lemma allows us to verify whether or not $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ is bounded on $(\mathbf{V}, ||\cdot||)$ (i.e., $\phi(\cdot)$ is regular at α_0) by checking whether or not $\lim_{k(n)\to\infty} \|v_n^*\| < \infty$.

Lemma 3.3. *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 (3.6). (1) If $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ is bounded on $(\mathbf{V}, ||\cdot||)$, then (3.4) holds, $v_n^* = \arg\min_{v \in \overline{\mathbf{V}}_{k(n)}} \|v^* - v\|$, $\|v^* - v_n^*\| \to 0$ and $\lim_{k(n)\to\infty} \|v_n^*\| = \|v^*\| < \infty$; (2) Let $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ be bounded on $(\mathbf{V}, ||\cdot||_s)$ and $\{\overline{\mathbf{V}}_k\}_{k=1}^{\infty}$ be dense in $(\mathbf{V}, \|\cdot\|_s)$. If $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ is unbounded on $(\mathbf{V}, ||\cdot||)$ then $\lim_{k(n)\to\infty} \|v_n^*\| = \infty$.*

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

$$S_{n,i}^* \equiv \left(\frac{dm(X_i, \alpha_0)}{d\alpha}[v_n^*] \right)' \Sigma(X_i)^{-1} \rho(Z_i, \alpha_0) \tag{3.7}$$

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

$$\|v_n^*\|_{sd}^2 = E[S_{n,i}^* S_{n,i}^{*\prime}] = E\left[\left(\frac{dm(X, \alpha_0)}{d\alpha}[v_n^*] \right)' \Sigma(X)^{-1} \Sigma_0(X) \Sigma(X)^{-1} \left(\frac{dm(X, \alpha_0)}{d\alpha}[v_n^*] \right) \right] \tag{3.8}$$

See Subsection 4.1.1 for closed form expressions of $\|v_n^*\|_{sd}^2$. Under Assumption 3.1(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 α_0). Moreover, if

$\phi(\cdot)$ is regular at α_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 3.3(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 A.1 in Appendix A for further discussion.

3.3 Two key local conditions

For all $k(n)$, let

$$u_n^* \equiv \frac{v_n^*}{\|v_n^*\|_{sd}} \tag{3.9}$$

be the "scaled sieve Riesz representer". Since $\|v_n^*\|_{sd}^2 \asymp \|v_n^*\|^2$ (under Assumption 3.1(iv)), we have: $\|u_n^*\| \asymp 1$ and $\|u_n^*\|_s \leq c\tau_n$ for τ_n defined in (3.3) 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 3.5 (Local behavior of ϕ)**.** *(i)* $v \mapsto \frac{d\phi(\alpha_0)}{d\alpha}[v]$ *is a non-zero linear functional mapping from* $\mathbf{V}$ *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* $\frac{\|v_n^*\|}{\sqrt{n}} = o(1)$*;*

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

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

Since $\|v_n^*\|_{sd}^2 \asymp \|v_n^*\|^2$ (under Assumption 3.1(iv)), we could rewrite Assumption 3.5 using $\|v_n^*\|_{sd}$ instead $\|v_n^*\|$. As it will become clear in Theorem 4.1 that $\frac{\|v_n^*\|_{sd}^2}{n}$ is the variance of $\phi(\widehat{\alpha}_n) - \phi(\alpha_0)$, Assumption 3.5(i) puts a restriction on how fast the sieve dimension $k(n)$ could grow with the sample size n.

Assumption 3.5(ii) controls the nonlinearity bias of $\phi(\cdot)$ (i.e., the linear approximation error of a possibly nonlinear functional $\phi(\cdot)$). It is automatically satisfied when $\phi(\cdot)$ is a linear functional. For a nonlinear functional $\phi(\cdot)$ (such as the quadratic functional), it can be verified using the smoothness of $\phi(\cdot)$ and the convergence rates in both $||\cdot||$ and $||\cdot||_s$ metrics (the definition of $\mathcal{N}_{osn}$). See Section 6 for verification.

Assumption 3.5(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 3.5(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 3.5(iii) is satisfied if*

$$||v^* - v_n^*|| \times ||\alpha_{0,n} - \alpha_0|| = o(n^{-1/2}). \tag{3.10}$$

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), Condition (3.10) 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.

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 $(\alpha, t) \in \mathcal{N}_{osn} \times \mathcal{T}_n$, we let $\widehat{\Lambda}_n(\alpha(t), \alpha) \equiv 0.5\{\widehat{Q}_n(\alpha(t)) - \widehat{Q}_n(\alpha)\}$ with $\alpha(t) \equiv \alpha + tu_n^*$. Denote

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

Assumption 3.6 (LQA). *(i) $\alpha(t) \in \mathcal{A}_{k(n)}$ for any $(\alpha, t) \in \mathcal{N}_{osn} \times \mathcal{T}_n$; 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 3.6(ii) is a standard one, and is implied by the following Lindeberg condition: For all $\epsilon > 0$,

$$\limsup_{n\to\infty} E\left[\left(\frac{S_{n,i}^*}{\|v_n^*\|_{sd}}\right)^2 1\left\{\left|\frac{S_{n,i}^*}{\epsilon\sqrt{n}\,\|v_n^*\|_{sd}}\right| > 1\right\}\right] = 0, \tag{3.12}$$

which, under Lemma 3.3(1) and Assumption 3.1(iv), is satisfied when the functional $\phi(\cdot)$ is regular ($\|v_n^*\|_{sd} \asymp \|v_n^*\| \to \|v^*\| < \infty$). This is why Assumption 3.6(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 3.6(i) implicitly imposes restrictions on the nonparametric estimator $\widehat{m}(x,\alpha)$ of $m(x,\alpha) = E[\rho(Z,\alpha)|X = x]$ 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 (such as the NPQIV model). 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. When $\widehat{m}(x,\alpha)$ is the series LS estimator (2.5), Lemma 5.1 in Section 5 shows that Assumption 3.6(i) is satisfied by a set of relatively low level sufficient conditions (Assumptions A.4 - A.7 in Appendix A). See Section 6 for verification of these sufficient conditions for functionals of the NPQIV model.

4 Asymptotic Properties of Sieve Wald and SQLR Statistics

In this section, we first establish the asymptotic normality of the plug-in PSMD estimator $\phi(\widehat{\alpha}_n)$ of $\phi(\alpha_0)$ for the model (1.1), regardless of it is root-n estimable or not. We then provide a simple consistent variance estimator and hence the asymptotic standard normality of the corresponding sieve t statistic for a real-valued functional $\phi : \mathbb{R}^{d_\theta} \times \mathbf{H} \to \mathbb{R}$. We finally derive the asymptotic properties of SQLR tests for the hypothesis $\phi(\alpha_0) = \phi_0$. See Appendix A for the case of a vector-valued functional $\phi : \mathbb{R}^{d_\theta} \times \mathbf{H} \to \mathbb{R}^{d_\phi}$ (where d_ϕ could grow slowly with n).

4.1 Asymptotic normality of the plug-in PSMD estimator

The next result allows for a (possibly) nonlinear irregular functional $\phi()$ of the general model (1.1).

Theorem 4.1. *Let $\widehat{\alpha}_n$ be the PSMD estimator (2.2) and Assumptions 3.1 - 3.4 hold. If Assumptions 3.5 and 3.6 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 4.1 implies that

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

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

Thus, Theorem 4.1 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 A.1 in Appendix A for further discussion.

4.1.1 Closed form expressions of sieve Riesz representer and sieve variance

To apply Theorem 4.1, one needs to know the sieve Riesz representer v^*_n defined in (3.6) and the sieve variance $\|v^*_n\|^2_{sd}$ given in (3.8). It turns out that both can be computed in closed form.

Lemma 4.1. *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 (3.1). Then: the sieve Riesz representer $v^*_n = (v^{*\prime}_{\theta,n}, v^*_{h,n}(\cdot))' \in \overline{\mathbf{V}}_{k(n)}$ of $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ has a closed form expression:*

$$v^*_n = (v^{*\prime}_{\theta,n}, \psi^{k(n)}(\cdot)'\beta^*_n)' = \overline{\psi}^{k(n)}(\cdot)'\gamma^*_n, \text{ and } \gamma^*_n = D_n^{-1}\digamma_n \tag{4.1}$$

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*

$$\|v^*_n\|^2 = \gamma^{*\prime}_n D_n \gamma^*_n = \digamma'_n D_n^{-1}\digamma_n. \tag{4.2}$$

The sieve variance (3.8) also has a closed form expression:

$$\|v^*_n\|^2_{sd} = \digamma'_n D_n^{-1}\mho_n D_n^{-1}\digamma_n, \tag{4.3}$$

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

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 $\overline{\psi}^{k(n)}(\cdot) = \overline{q}^{k(n)}(\cdot)$ in Lemma 4.1, and (4.3) becomes the sieve variance expression given in (2.6).

Lemmas 3.3 and 4.1 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 $= \infty$).*

According to Lemma 4.1 we could use different finite dimensional linear sieve basis $\psi^{k(n)}$ to compute sieve Riesz representer $v^*_n = (v^{*\prime}_{\theta,n}, v^*_{h,n}(\cdot))' \in \overline{\mathbf{V}}_{k(n)}$, $\|v^*_n\|^2$ and $\|v^*_n\|^2_{sd}$. 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 Sections 2.2, 4.2 and 6 for related discussions.

4.2 Consistent estimator of sieve variance of $\phi(\widehat{\alpha}_n)$

In order to apply the asymptotic normality Theorem 4.1, we need an estimator of the sieve variance $\|v_n^*\|_{sd}^2$ defined in (3.8). We now provide one simple consistent estimator of the sieve variance when the residual function $\rho()$ is pointwise smooth with respect to α_0. See Appendix B for additional consistent variance estimators.

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

$$\langle v_1, v_2 \rangle_{n,M} \equiv \frac{1}{n}\sum_{i=1}^{n}\left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n)}{d\alpha}[v_1]\right)' M_{n,i}\left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n)}{d\alpha}[v_2]\right), \tag{4.4}$$

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

$$\frac{d\phi(\widehat{\alpha}_n)}{d\alpha}[\widehat{v}_n^*] = \sup_{v\in\overline{\mathbf{V}}_{k(n)}, v\neq 0}\frac{|\frac{d\phi(\widehat{\alpha}_n)}{d\alpha}[v]|^2}{||v||^2_{n,\widehat{\Sigma}^{-1}}} < \infty \tag{4.5}$$

and

$$\frac{d\phi(\widehat{\alpha}_n)}{d\alpha}[v] = \langle \widehat{v}_n^*, v\rangle_{n,\widehat{\Sigma}^{-1}} \quad \text{for any } v \in \overline{\mathbf{V}}_{k(n)}. \tag{4.6}$$

For $\|v_n^*\|_{sd}^2 = E\left(S_{n,i}^* S_{n,i}^{*\prime}\right)$ given in (3.8) we can define a simple plug-in sieve variance estimator:

$$||\widehat{v}_n^*||^2_{n,sd} = \frac{1}{n}\sum_{i=1}^{n}\widehat{S}_{n,i}^*\widehat{S}_{n,i}^{*\prime} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n)}{d\alpha}[\widehat{v}_n^*]\right)'\widehat{\Sigma}_i^{-1}\left(\widehat{\rho}_i\widehat{\rho}_i'\right)\widehat{\Sigma}_i^{-1}\left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n)}{d\alpha}[\widehat{v}_n^*]\right) \tag{4.7}$$

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

Under condition stated in Lemma 4.1, $\widehat{v}_n^*$ defined in (4.5-4.6) also has a closed form solution:

$$\widehat{v}_n^* = \overline{\psi}^{k(n)}(\cdot)'\widehat{\gamma}_n^*, \quad \text{and} \quad \widehat{\gamma}_n^* = \widehat{D}_n^{-1}\widehat{\digamma}_n, \tag{4.8}$$

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 (4.7) now becomes

$$||\widehat{v}_n^*||_{n,sd}^2 = \widehat{V}_1 \equiv \widehat{F}_n' \widehat{D}_n^{-1} \widehat{\mho}_n \widehat{D}_n^{-1} \widehat{F}_n \quad \text{with} \tag{4.9}$$

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

In particular, with $\psi^{k(n)} = q^{k(n)}$ the sieve variance estimator $||\widehat{v}_n^*||_{n,sd}^2$ given in (4.9) becomes the one given in (2.10) in Subsection 2.2.

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 $\overline{\mathbf{V}}_{k(n)}^1 \equiv \{v \in \overline{\mathbf{V}}_{k(n)} : ||v|| = 1\}$.

Assumption 4.1. *(i)* $\sup_{\alpha\in\mathcal{N}_{osn}} \sup_{v\in\overline{\mathbf{V}}_{k(n)}^1} \left|\frac{d\phi(\alpha)}{d\alpha}[v] - \frac{d\phi(\alpha_0)}{d\alpha}[v]\right| = o(1)$*;*
(ii) for each $k(n)$ *and any* $\alpha \in \mathcal{N}_{osn}$*,* $v \in \overline{\mathbf{V}}_{k(n)} \mapsto \frac{d\widehat{m}(\cdot,\alpha)}{d\alpha}[v] \in L^2(f_X)$ *is a linear functional measurable with respect to* Z^n*; and* $\sup_{v_1,v_2\in\overline{\mathbf{V}}_{k(n)}^1} \left|\langle v_1, v_2\rangle_{n,\Sigma^{-1}} - \langle v_1, v_2\rangle_{\Sigma^{-1}}\right| = o_{P_{Z^\infty}}(1)$*;*
(iii) $\sup_{x\in\mathcal{X}} ||\widehat{\Sigma}(x) - \Sigma(x)||_e = o_{P_{Z^\infty}}(1)$*;*
(iv) $\sup_{x\in\mathcal{X}} E\left[\sup_{\alpha\in\mathcal{N}_{osn}} ||\rho(Z,\alpha)\rho(Z,\alpha)' - \rho(Z,\alpha_0)\rho(Z,\alpha_0)'||_e | X = x\right] = o(1)$*.*
(v) $\sup_{v\in\overline{\mathbf{V}}_{k(n)}^1} |\langle v, v\rangle_{n,M} - \langle v, v\rangle_M| = o_{P_{Z^\infty}}(1)$ *with* $M = \Sigma^{-1}\rho(Z,\alpha_0)\rho(Z,\alpha_0)'\Sigma^{-1}$*.*

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 $v \in \overline{\mathbf{V}}_{k(n)}^1$. 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 (see, e.g., Ai and Chen (2003) and Chen and Pouzo (2009))

Theorem 4.2. *Let Assumptions 3.1 - 3.4 hold. If Assumption 4.1 is satisfied, then:*
(1) $\left|\frac{||\widehat{v}_n^*||_{n,sd}}{||v_n^*||_{sd}} - 1\right| = o_{P_{Z^\infty}}(1)$ *for* $||\widehat{v}_n^*||_{n,sd}$ *given in (4.7).*
(2) If, in addition, Assumptions 3.5 and 3.6 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 4.2(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 4.2 (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.3 in Appendix A for asymptotic properties of $\mathcal{W}_n$ *under local alternatives.*

4.3 Sieve QLR statistics

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

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

$$\widehat{Q}_n(\widehat{\alpha}_n^R) + \lambda_n Pen(\widehat{h}_n^R) \leq \inf_{\alpha \in \mathcal{A}_{k(n)}^R} \left\{\widehat{Q}_n(\alpha) + \lambda_n Pen(h)\right\} + o_{P_{Z^\infty}}(n^{-1}). \tag{4.10}$$

Then:

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

Recall that $u_n^* \equiv v_n^*/\left\|v_n^*\right\|_{sd}$, and that $\widehat{QLR}_n^0(\phi_0)$ denotes the optimally weighted (i.e., $\Sigma = \Sigma_0$) SQLR statistic in Subsection 2.2. We note that $||u_n^*|| = 1$ for the optimally weighted case.

Theorem 4.3. *Let Assumptions 3.1 - 3.6 hold with* $\left|B_n - ||u_n^*||^2\right| = o_{P_{Z^\infty}}(1)$. *If* $\widehat{\alpha}_n^R \in \mathcal{N}_{osn}$ *wpa1-*P_{Z^∞}*, then: (1) under the null* $H_0 : \phi(\alpha_0) = \phi_0$,

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

(2) Further, let $\widehat{\alpha}_n$ *be the optimally weighted PSMD estimator (2.2) with* $\Sigma = \Sigma_0$. *Then: under* $H_0 : \phi(\alpha_0) = \phi_0$,

$$\widehat{QLR}_n^0(\phi_0) = \left(\sqrt{n}\mathbb{Z}_n\right)^2 + o_{P_{Z^\infty}}(1) \Rightarrow \chi_1^2.$$

See Theorem A.1 in Appendix A for the asymptotic behavior under local alternatives.

Compared to Theorem 4.1 on the asymptotic normality of $\phi(\widehat{\alpha}_n)$, Theorem 4.3 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 5.1 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 3.2 (i.e., $\widehat{\alpha}_n \in \mathcal{N}_{osn}$ wpa1-P_{Z^∞}).

Theorem 4.3(2) recommends to construct an asymptotic $100(1-\tau)\%$ confidence set for $\phi(\alpha)$ by inverting the optimally weighted SQLR statistic: $\left\{r \in \mathbb{R} : \ \widehat{QLR}_n^0(r) \leq c_{\chi_1^2}(1-\tau)\right\}$. This result extends that of Chen and Pouzo (2009) to allow for irregular functionals.

Next, we consider the asymptotic behavior of $\widehat{QLR}_n(\phi_0)$ under the fixed alternatives $H_1 : \phi(\alpha_0) \neq \phi_0$.

Theorem 4.4. *Let Assumptions 3.1, 3.2 and 3.3 hold. Suppose that* $\sup_{h\in\mathcal{H}} Pen(h) < \infty$ *and* ϕ *is continuous in* $||\cdot||_s$*. Then: under* $H_1 : \phi(\alpha_0) \neq \phi_0$*, there is a constant* $C > 0$ *such that*

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

5 Inference Based on Generalized Residual Bootstrap

The inference procedures described in Sections 4 and 4.3 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

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

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 5.2 for another simple bootstrap Wald statistic).

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

5.1 Bootstrap local quadratic approximation (LQA^B)

Lemma A.1 in Appendix A shows that the bootstrap PSMD estimator $\widehat{\alpha}_n^B \in \mathcal{N}_{osn}$ wpa1 under Assumptions 3.1 - 3.4 and A.1. In the following we introduce a condition that is a bootstrap version of the LQA Assumption 3.6. For any $\alpha \in \mathcal{N}_{osn}$, we let $\widehat{\Lambda}_n^B(\alpha(t_n),\alpha) \equiv 0.5\{\widehat{Q}_n^B(\alpha(t_n)) - \widehat{Q}_n^B(\alpha)\}$ with $\alpha(t_n) \equiv \alpha + t_n u_n^*$ for $t_n \in \mathcal{T}_n$,. For any sequence of non-negative weights $(b_i)_i$, let

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

Assumption Boot.3 (LQA^B). *(i) $\alpha(t) \in \mathcal{A}_{k(n)}$ for any $(\alpha,t) \in \mathcal{N}_{osn}\times\mathcal{T}_n$, and with $r_n(t_n) = \left(\max\{t_n^2, t_n n^{-1/2}, o(n^{-1})\}\right)^{-1}$,*

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

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

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

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

Assumption Boot.3(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 Boot.3(i) when $\widehat{m}^B(x,\alpha)$ is a bootstrap series LS estimator.

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

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 5.1. *Let Assumptions 3.1 - 3.4 and A.4 - A.7 hold.*

(1) Let $\widehat{m}$ be the series LS estimator (2.5). Then Assumption 3.6(i) is satisfied. Further, if Assumption B holds then $\left|B_n - ||u_n^*||^2\right| = o_{P_{Z^\infty}}(1)$.

(2) Let $\widehat{m}^B(\cdot,\alpha)$ be the bootstrap series LS estimator (2.16), Assumption A.1, and either Assumption Boot.1 or Boot.2 hold. Then Assumption Boot.3(i) holds with $B_n^\omega = B_n$. Further, if Assumption B holds then $\left|B_n^\omega - ||u_n^*||^2\right| = o_{P_{V^\infty|Z^\infty}}(1)$ wpa1(P_{Z^∞}).

Lemma 5.1 indicates that the low level Assumptions A.4 - A.7 are sufficient for both the original-sample LQA Assumption 3.6(i) and the bootstrap LQA Assumption Boot.3(i).

Assumption Boot.3(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 Boot.3(ii).

Lemma 5.2. *Let either Assumption Boot.1 or Assumption Boot.2 hold. If there is a positive real sequence $(b_n)_n$ such that $b_n = o(\sqrt{n})$ and*

$$\limsup_{n\to\infty} E\left[(g(X,u_n^*)\rho(Z,\alpha_0))^2\, 1\left\{\frac{(g(X,u_n^*)\rho(Z,\alpha_0))^2}{b_n} > 1\right\}\right] = 0. \tag{5.1}$$

Then: Assumptions Boot.3(ii) and 3.6(ii) hold.

5.2 Bootstrap sieve Student t statistic

Lemma A.1 shows that $\widehat{\alpha}_n^B \in \mathcal{N}_{osn}$ wpa1 under virtually the same conditions as those for the original-sample estimator $\widehat{\alpha}_n \in \mathcal{N}_{osn}$ wpa1. This would easily lead to the consistency of the simplest bootstrap sieve t statistic $\widehat{W}_{1,n}^B \equiv \sqrt{n}\frac{\phi(\widehat{\alpha}_n^B)-\phi(\widehat{\alpha}_n)}{\sigma_\omega ||\widehat{v}_n^*||_{n,sd}}$.

We now establish the consistency of another bootstrap sieve t statistic $\widehat{W}_{2,n}^B \equiv \sqrt{n}\frac{\phi(\widehat{\alpha}_n^B)-\phi(\widehat{\alpha}_n)}{||\widehat{v}_n^*||_{B,sd}}$, where $||\widehat{v}_n^*||_{B,sd}^2$ is a bootstrap sieve variance estimator:

$$||\widehat{v}_n^*||_{B,sd}^2 \equiv \frac{1}{n}\sum_{i=1}^n \left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n)}{d\alpha}[\widehat{v}_n^*]\right)' \widehat{\Sigma}_i^{-1} \varrho(V_i,\widehat{\alpha}_n)\varrho(V_i,\widehat{\alpha}_n)'\widehat{\Sigma}_i^{-1}\left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n)}{d\alpha}[\widehat{v}_n^*]\right) \tag{5.2}$$

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.

We note that $||\widehat{v}_n^*||_{B,sd}^2$ is an analog to $||\widehat{v}_n^*||_{n,sd}^2$ defined in (4.7) but using the bootstrapped generalized residual $\varrho(V_i,\widehat{\alpha}_n)$ instead of the original sample fitted residual $\rho(Z_i,\widehat{\alpha}_n)$. It also 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, $||\widehat{v}_n^*||^2_{B,sd}$ is computed in the same way as $||\widehat{v}_n^*||^2_{n,sd} = \widehat{F}'_n \widehat{D}_n^{-1} \widehat{\mho}_n \widehat{D}_n^{-1} \widehat{F}_n$ given in (4.9) except using $\widehat{\mho}_n^B$ instead of $\widehat{\mho}_n$.

Assumption Boot.4. $\sup_{v \in \overline{\mathbf{V}}^1_{k(n)}} |\langle v, v\rangle_{n,\hat{M}^B} - \sigma^2_\omega \langle v, v\rangle_{n,\hat{M}}| = o_{P_{V^\infty|Z^\infty}}(1)$ *wpa1*(P_{Z^∞}) *with* $\hat{M}_i^B = (\omega_{i,n}-1)^2 \hat{M}_i$ *and* $\hat{M}_i = \hat{\Sigma}_i^{-1}\rho(Z_i,\widehat{\alpha}_n)\rho(Z_i,\widehat{\alpha}_n)'\hat{\Sigma}_i^{-1}$.

This assumption can be verified given Assumptions Boot.1 or Boot.2. The following result is a bootstrap version of Theorem 4.2(1).

Theorem 5.1. *Let Assumptions 3.1 - 3.4, 4.1 and Boot.4 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 statistic $\widehat{W}^B_{2,n}$.

Theorem 5.2. *Let* $\widehat{\alpha}_n$ *be the PSMD estimator (2.2) and* $\widehat{\alpha}_n^B$ *the bootstrap PSMD estimator. Let Assumptions 3.1 - 3.4 and A.1 hold. Let Assumptions 3.5, 3.6 and Boot.3 hold.*

(1) Let Assumptions 4.1 and Boot.4 hold. Then:

$$\sup_{t\in\mathbb{R}} \left| P_{V^\infty|Z^\infty}\left(\widehat{W}^B_{2,n} \leq t \mid Z^n\right) - P_{Z^\infty}\left(\widehat{W}_n \leq t\right) \right| = o_{P_{V^\infty|Z^\infty}}(1)\ wpa1(P_{Z^\infty}).$$

(2) If ϕ *is regular, without imposing Assumptions 4.1 and Boot.4, 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 bootstrap sieve variance estimator $||\widehat{v}_n^*||^2_{B,sd}$ to apply Theorem 5.2(1). Luckily $||\widehat{v}_n^*||^2_{B,sd}$ is easy to compute when the residual function $\rho(Z_i,\alpha)$ is pointwise smooth in α_0. Moreover, since $E\left(||\widehat{v}_n^*||^2_{B,sd} \mid Z^n\right) = \sigma^2_\omega ||\widehat{v}_n^*||^2_{n,sd}$ we suspect that the bootstrap sieve t statistic $\widehat{W}^B_{2,n}$ might have second order refinement property by choices of bootstrap weights $\{\omega_{i,n}\}$. This will be a subject of future research.

The bootstrap sieve t statistic $\widehat{W}^B_{2,n}$ requires to compute the original sample PSMD estimator $\widehat{\alpha}_n$ and the bootstrap PSMD estimator $\widehat{\alpha}_n^B$. In online supplemental 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 5.1. *Theorems 4.2(2) and 5.2(1) imply that the bootstrap Wald test statistic* $\mathcal{W}^B_{2,n} \equiv \left(\widehat{W}^B_{2,n}\right)^2$ *always has the same limiting distribution* χ^2_1 *(conditional on the data) under the null and the alternatives. Let* $\widehat{c}_{2,n}(a)$ *be the* $a-th$ *quantile of the distribution of* $\mathcal{W}^B_{2,n}$ *(conditional on the data* $\{Z_i\}^n_{i=1}$*). 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 any* $\tau \in (0,1)$*,*

under $H_0: \phi(\alpha_0) = \phi_0$, $\lim_{n\to\infty} \Pr\left(\mathcal{W}_n \geq \widehat{c}_{2,n}(1-\tau)\right) = \tau$*;*

under $H_1: \phi(\alpha_0) \neq \phi_0$, $\lim_{n\to\infty} \Pr\left(\mathcal{W}_n \geq \widehat{c}_{2,n}(1-\tau)\right) = 1$.

See Theorem A.4 in Appendix A for properties under local alternatives.

See online supplemental Appendix B for consistency of $\mathcal{W}^B_{1,n} \equiv \left(\sqrt{n}\frac{\phi(\widehat{\alpha}^B_n)-\widehat{\phi}_n}{\sigma_\omega||\widehat{v}^*_n||_{n,sd}}\right)^2$ and other bootstrap sieve Wald (t) statistics based on different sieve variance estimators.

5.3 Bootstrap SQLR statistic

If $\Sigma \neq \Sigma_0$, the SQLR statistic $\widehat{QLR}_n(\phi_0) = n\left(\widehat{Q}_n(\widehat{\alpha}^R_n) - \widehat{Q}_n(\widehat{\alpha}_n)\right)$ is no longer asymptotically chi-square even under the null; Theorem 4.3(1), 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 $\widehat{QLR}^B_n(\widehat{\phi}_n)$. Recall that

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

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

$$\widehat{Q}^B_n(\widehat{\alpha}^{R,B}_n) + \lambda_n Pen(\widehat{h}^{R,B}_n) \leq \inf_{\alpha\in\mathcal{A}_{k(n)}:\phi(\alpha)=\widehat{\phi}_n} \left\{\widehat{Q}^B_n(\alpha) + \lambda_n Pen(h)\right\} + o_{P_{V^\infty|Z^\infty}}(\frac{1}{n})\ wpa1(P_{Z^\infty}). \tag{5.3}$$

Lemma A.1 in Appendix A implies that $\widehat{\alpha}^{R,B}_n$, $\widehat{\alpha}^B_n \in \mathcal{N}_{osn}$ 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 $\widehat{QLR}^B_n(\widehat{\phi}_n)$ 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 Assumptions 3.1 - 3.4 and A.1 hold. Let Assumptions 3.5, 3.6 and Boot.3 hold with* $\left|B^\omega_n - ||u^*_n||^2\right| = o_{P_{V^\infty|Z^\infty}}(1)\ wpa1(P_{Z^\infty})$*. Then:*

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

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

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 5.2. *Let $\widehat{c}_n(a)$ be the $a-th$ quantile of the distribution of $\frac{\widehat{QLR}_n^B(\widehat{\phi}_n)}{\sigma_\omega^2}$ (conditional on the data $\{Z_i\}_{i=1}^n$). Then Theorems 4.3, 4.4 and 5.3 immediately imply that for any $\tau\in(0,1)$,*

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

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

See Theorem A.2 in Appendix A for properties under local alternatives.

6 Verification of Assumptions 3.5 and 3.6(i)

In this section, we illustrate the verification of the two key regularity conditions, Assumption 3.5 and Assumption 3.6(i), via some functionals $\phi(h)$ of the (nonlinear) nonparametric IV regressions:

$$E[\rho(Y_1;h_0(Y_2))|X]=0\quad a.s.-X, \tag{6.1}$$

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 $\widehat{h}\in\mathcal{H}_{k(n)}$ 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)).[14] By definition, $\mathcal{H}\subset L^2(f_{Y_2})$ and we let $||\cdot||_s=||\cdot||_{L^2(f_{Y_2})}$. We assume that $\mathcal{H}_{k(n)}=clsp\{q_1,...,q_{k(n)}\}$ 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 3.1 (since there is no θ here).

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

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 $(TT^*)^{1/2}$:

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

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

$$h_{0,n} = \arg\min_{\{a_k\}} \left[\sum_{k=1}^{k(n)} \left(\langle h_0, \psi_k\rangle_s - a_k \right)^2 \mu_k^2 + \sum_{k=k(n)+1}^{\infty} \langle h_0, \psi_k\rangle_s^2 \mu_k^2 \right] = \sum_{k=1}^{k(n)} \langle h_0, \psi_k\rangle_s \psi_k = \Pi_n h_0. \quad (6.2)$$

The next remark specializes Theorem 4.1 to a general functional $\phi(h)$ of the model (6.1).

Remark 6.1. *Let $\widehat{m}$ be the series LS estimator (2.5) for the model (6.1) with $\widehat{\Sigma} = \Sigma = 1$, and Assumptions 3.1(i)(ii), 3.2(ii)(iii), and 3.4 hold with $\delta_n = O\left(\sqrt{\frac{k(n)}{n}}\right) = o(n^{-1/4})$ and $\delta_{s,n} = O\left(\{k(n)\}^{-\varsigma} + \mu^{-1}_{k(n)}\sqrt{\frac{k(n)}{n}}\right) = o(1)$. Let Assumption 3.5, equation (3.12) and Assumptions A.4 - A.7 hold. Then:*

$$\sqrt{n}\frac{\phi(\widehat{h}_n) - \phi(h_0)}{||v_n^*||_{sd}} \Rightarrow N(0,1) \text{ with } ||v_n^*||^2_{sd} = (\frac{d\phi(h_0)}{dh}[q^{k(n)}(\cdot)])' D_n^{-1} \mho_n D_n^{-1} (\frac{d\phi(h_0)}{dh}[q^{k(n)}(\cdot)]), \quad (6.3)$$

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

Remark 6.1 includes the NPIV and NPQIV examples in Subsection 2.2 as special cases. In particular, the sieve variance expression (6.3) 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 3.5(iii) for irregular functionals.

6.1 Verification of Assumption 3.5

Let $b_j \equiv \frac{d\phi(h_0)}{dh}[\psi_j(\cdot)]$ for all j. By Lemma 4.1 $D_n = E\left[\left(T[q^{k(n)}(\cdot)']\right)'\left(T[q^{k(n)}(\cdot)']\right)\right] = Diag\left\{\mu_1^2, ..., \mu_{k(n)}^2\right\}$ and

$$||v_n^*||^2 = \left(\frac{d\phi(h_0)}{dh}[q^{k(n)}(\cdot)]\right)' D_n^{-1}\left(\frac{d\phi(h_0)}{dh}[q^{k(n)}(\cdot)]\right) = \sum_{j=1}^{k(n)} \mu_j^{-2} b_j^2. \tag{6.4}$$

By Lemma 3.3, $\phi(h)$ of the model (6.1) 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 (6.5) without endogeneity:

$$E[\rho(Y_1; h_0(Y_2))|Y_2] = 0 \quad a.s. - Y_2, \tag{6.5}$$

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 (6.5) 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 (6.5) without endogeneity, then it is irregular for the model (6.1). But, even if a functional $\phi(h)$ is regular for the model (6.5) without endogeneity, it could still be irregular for the model (6.1) with endogeneity.

6.1.1 Linear functionals of the model (6.1)

For a linear functional $\phi(h)$ of the model (6.1), given relation (6.2), Assumption 3.5 is satisfied provided that the sieve dimension $k(n)$ satisfies (6.6):

$$\frac{||v_n^*||}{\sqrt{n}} = o(1) \quad \text{and} \quad \sqrt{n}\frac{\left|\frac{d\phi(h_0)}{dh}[\Pi_n h_0 - h_0]\right|}{||v_n^*||} = o(1). \tag{6.6}$$

When $\phi(h)$ of the model (6.1) is regular, Remark 3.1 implies that (6.6) is satisfied provided

$$\sum_{j=1}^{\infty} \mu_j^{-2} b_j^2 < \infty \quad \text{and} \quad n \times \sum_{j=k(n)+1}^{\infty} \mu_j^{-2} b_j^2 \times ||\Pi_n h_0 - h_0||^2 = o(1). \tag{6.7}$$

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

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

To provide concrete sufficient condition for (6.6), 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 (6.6) is satisfied provided that

$$\frac{||v_n^*||^2}{n} = \frac{\sum_{k=1}^{k(n)} \mu_k^{-2}}{n} = o(1) \quad \text{and} \quad \frac{\{k(n)\}^{-2\varsigma}}{\frac{1}{n}||v_n^*||^2} = \frac{\{k(n)\}^{-2\varsigma}}{\frac{1}{n}\sum_{k=1}^{k(n)} \mu_k^{-2}} = o(1). \tag{6.8}$$

Condition (6.8) 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 (6.8) 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 (6.8) 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 (6.8) 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 (6.5) 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 (6.1), Condition (6.7) 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 (6.1), Condition (6.6) is satisfied provided that

$$\frac{||v_n^*||^2}{n} = \frac{\sum_{j=1}^{k(n)} \mu_j^{-2} b_j^2}{n} = o(1) \quad \text{and} \quad \frac{\{k(n)\}^{-2\varsigma}}{\frac{1}{n}||v_n^*||^2} = \frac{\{k(n)\}^{-2\varsigma}}{\frac{1}{n}\sum_{j=1}^{k(n)} \mu_j^{-2} b_j^2} = o(1). \tag{6.9}$$

Condition (6.9) allows for both mildly and severely ill-posed cases. To provide concrete sufficient conditions for (6.9) 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 $0 < c \leq c' < \infty$. Condition (6.9) 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 $0 < c \leq c' < \infty$. Condition (6.9) 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 $\varsigma > 1$.

6.1.2 Nonlinear functionals

For a nonlinear functional $\phi(h)$ of the model (6.1), Assumption 3.5 is satisfied provided that the sieve dimension $k(n)$ satisfies (6.6) (or (6.7) if $\phi(h)$ is regular) and Assumption 3.5(ii), which is implied by the following condition:

Assumption 3.5(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 3.5(ii) or (ii)' controls the nonlinearity bias of $\phi(\cdot)$ (i.e., the linear approximation error of a nonlinear functional $\phi(\cdot)$). It 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 (6.5) without endogeneity. Also,

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

(i) When the weighted quadratic functional is regular (i.e., $\sum_{j=1}^{\infty} \mu_j^{-2} b_j^2 < \infty$) for the model (6.1), Condition (6.7) 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 3.5(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 (6.1), Condition (6.6) is satisfied provided that Condition (6.9) holds with $b_j = \int h_0(y)w(y)\psi_j(y)dy$ for Example 3. Assumption 3.5(ii)' is satisfied provided that

$$\sqrt{n}\frac{\delta_{s,n}^2}{||v_n^*||} = \frac{\sqrt{n} \times \left(\{k(n)\}^{-\varsigma} + \mu_{k(n)}^{-1}\sqrt{\frac{k(n)}{n}}\right)^2}{||v_n^*||} \le n^{-1/2}\frac{\mu_{k(n)}^{-2} k(n)}{\sqrt{\sum_{j=1}^{k(n)} \mu_j^{-2} b_j^2}} = o(1). \tag{6.10}$$

Any $k(n)$ satisfying Conditions (6.9) and (6.10) 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 $0 < c \le c' < \infty$. Conditions (6.9) and (6.10) 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 $0 < c \le c' < \infty$. Conditions (6.9) and (6.10) are satisfied with $k(n) \asymp a^{-1}[\ln(n) - \varpi\ln(\ln(n))]$ and $\varpi \in (3, 2\varsigma - 1)$ for $\varsigma > 2$.

6.2 Verification of Assumption 3.6(i)

By Lemma 5.1(1), to verify Assumption 3.6(i), it suffices to verify Assumptions A.4 - A.7 in Appendix A. Note that Assumptions A.4 and A.5 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 (6.1). Assumptions A.6 and A.7 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 (6.1). Here, we present verifications of Assumptions A.6 and A.7 for irregular functionals of the NPIV and NPQIV examples.

Condition 6.1. *(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 6.1. *Let all conditions for Remark 6.1 hold. Under Condition 6.1, Assumptions A.6 and A.7 hold for the NPIV model (2.18).*

Proposition 6.1 allows for irregular functionals of the NPIV model with severely ill-posed case.

Condition 6.2. *(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 6.3. $n(\log\log n)^4\delta_{s,n}^4 = o(1)$

Proposition 6.2. *Let all conditions for Remark 6.1 hold. Under conditions 6.1(ii) and 6.2-6.3, Assumptions A.6 and A.7 hold for the NPQIV model (2.21).*

It is clear that Condition 6.3 rules out severely ill-posed case, and hence Proposition 6.2 only allows for irregular functionals of the NPQIV model with mildly ill-posed case.

7 Simulation Studies and An Empirical Illustration

In this section, we first present simulation studies for SQLR and sieve t tests of linear and nonlinear hypotheses for the NPQIV and NPIV models respectively. We then provide an empirical illustration of the optimally weighted SQLR inferences for a NPQIV Engel curve. In this section, we use the series LS estimator (2.5) of $m(x,h)$ with $p^{J_n}(x)$ as its basis, and $q^{k(n)}$ as the basis approximating the unknown structure function h_0. We use $p^J = \mathrm{P-Spline}(r,k)$ to denote rth degree polynomial spline with k (quantile) equally spaced knots, hence $J = (r+1)+k$ is the total number of sieve terms. We use $p^J = \mathrm{Pol}(J)$ to denote power series up to $(J-1)$th degree. See Chen (2007) for definitions.

7.1 Simulation Studies

We run Monte Carlo (MC) studies to assess the finite sample performance of SQLR and sieve t procedures in two models: the NPQIV (2.21) and the NPIV (2.18). We also consider linear and non-linear functionals.

For all cases, our design is based on the MC design of Newey and Powell (2003) and Santos (2012) for a NPIV model, which we adapt to cover both NPIV and NPQIV models. Specifically, we generate i.i.d. draws of (Y_2, X, U^*) from

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

and $Y_2 = 2(\Phi(Y_2^*/3) - 0.5)$ and $X = 2(\Phi(X^*/3) - 0.5)$. The true function h_0 is given by $h_0(\cdot) = 2\sin(\pi\cdot)$. We consider 5,000 MC repetitions and $n = 750$ for each of the cases studied below. We use $Pen(h) = ||h||^2_{L^2} + ||\nabla h||^2_{L^2}$ in all the simulations, and have used a very small $\lambda_n = 10^{-5}$ in most cases (except for the cases we study the sensitivity to the choice of λ_n).

Summary of MC findings: For both NPQIV and NPIV, for both SQLR and sieve t tests, for both linear and nonlinear hypotheses, as long as $J_n > k(n) + 1$ with not too large $k(n)$, the MC sizes of the tests are good and are insensitive to the choices of basis $q^{k(n)}$ and p^{J_n} or the very small penalty λ_n. This is consistent with previous MC findings in Blundell et al. (2007) and Chen and Pouzo (2012a) for PSMD estimation of NPIV and NPQIV respectively.

NPQIV model: SQLR test for an irregular linear functional. We consider the NPQIV model $Y_1 = h_0(Y_2) + U = 2\sin(\pi Y_2) + U$ with $U = 2(\Phi(U^*) - \gamma)$. This last transformation is done to ensure that $E[1\{U \leq 0\}|X] = \gamma$. To save space we only present the case with $\gamma = 0.5$. The parameter of interest is $\phi(h_0) = h_0(0)$, hence ϕ is a irregular linear functional. We study the finite sample properties of the SQLR and bootstrap-SQLR tests. The SQLR-based confidence intervals are specially well-suited for models like NPQIV where the generalized residual function is non-smooth and also where the optimal weighting matrix is easy to compute.

Size. Table 7.1 reports the simulated size of the SQLR test of $H_0\colon \phi(h_0) = 0$ as a function of the nominal size (NS), for different choices of $q^{k(n)}$ and p^{J_n}, and different values of the tuning parameters $(\lambda_n, k(n), J_n)$.

Table 7.1 shows that for small value of $k(n)$, say in $(k(n), J_n) = (4, 7)$ (i.e., rows 1-3), the SQLR test performs well and is fairly insensitive to different choices of λ_n. For a fixed relatively small $J_n = 7$, rows 1-6 indicate that as $k(n)$ increases, the results become a bit more sensitive to the choice of λ_n. For a fixed very small penalty $\lambda_n = 10^{-5}$, rows 7-15 show that the results are fairly

$q^{k(n)}$	p^{J_n}	λ_n	10%	5%	1%
Pol(4)	Pol(7)	(1×10^{-3})	0.099	0.055	0.008
	Pol(7)	(2×10^{-4})	0.096	0.048	0.008
	Pol(7)	(4×10^{-5})	0.107	0.053	0.010
Pol(6)	Pol(7)	(1×10^{-3})	0.133	0.068	0.011
	Pol(7)	(2×10^{-4})	0.091	0.036	0.006
	Pol(7)	(4×10^{-5})	0.105	0.052	0.008
Pol(6)	Pol(9)	(1×10^{-5})	0.107	0.055	0.012
	Pol(15)	(1×10^{-5})	0.109	0.058	0.014
	Pol(21)	(1×10^{-5})	0.112	0.058	0.013
P-Spline(3,2)	Pol(9)	(1×10^{-5})	0.103	0.049	0.010
	Pol(15)	(1×10^{-5})	0.105	0.049	0.009
	Pol(21)	(1×10^{-5})	0.105	0.052	0.009
P-Spline(3,2)	P-Spline(5,3)	(1×10^{-5})	0.098	0.049	0.008
	P-Spline(5,9)	(1×10^{-5})	0.103	0.050	0.009
	P-Spline(5,18)	(1×10^{-5})	0.106	0.051	0.009

Table 7.1: Size of the SQLR test of $\phi(h_0) = 0$ for NPQIV model.

insensitive to different choices of J_n and basis for p^{J_n} and $q^{k(n)}$ as long as $J_n > k(n) + 1$.

Local power. The dashed blue line in Figure 7.1 shows the rejection probabilities at 5% level of the null hypothesis as a function of r where $r\colon \phi(h_0) = r$. We do this for the specification corresponding to Pol(4) for $q^{k(n)}$ and $\lambda_n = 2 \times 10^{-4}$. We note that since our functional $\phi(h) = h(0)$ is estimated at a slower than root-n rate, the deviations considered for r which are in the range of $\{0, 1/\sqrt{n}, ..., 8/\sqrt{n}\}$ are indeed "small". Finally, we study the finite sample behavior of the generalized residual bootstrap SQLR corresponding to Pol(4) for $q^{k(n)}$ and $\lambda_n = 2 \times 10^{-4}$, using multinomial bootstrap weights. We employ 250 bootstrap evaluations, and lower the number of MC repetitions to 500 to ease the computational burden. The solid red line in Figure 7.1 shows the rejection probabilities at 5% level of the null hypothesis as a function of r where $r\colon \phi(h_0) = r$. We can see from the figure that the bootstrap SQLR performance is similar to its non-bootstrapped counterpart. We expect that the performance will improve if we increase number of bootstrap runs.

NPIV model: sieve variance estimators for an irregular linear functional. We now consider the NPIV model: $Y_1 = h_0(Y_2) + 0.76U = 2\sin(\pi Y_2) + 0.76U$, with $U = U^*$ so the identifying condition of NPIV holds: $E[U|X] = 0$. The parameter of interest is $\phi(h_0) = h_0(0)$, and the null hypothesis is $H_0\colon \phi(h_0) = 0$. We focus on the finite sample performance of the sieve variance estimators for irregular linear functionals. 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 $\widehat{D}_n = n^{-1}\left(\widehat{C}_n(P'P)^{-}\widehat{C}_n'\right)$, $\widehat{C}_n \equiv \sum_{i=1}^n q^{k(n)}(Y_{2i})p^{J_n}(X_i)'$, $\widehat{\mho}_n$ is given in equation (2.20),

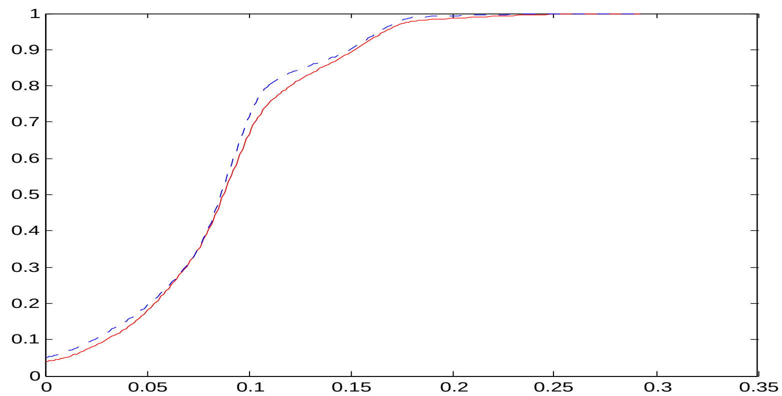

Figure 7.1: Rejection probabilities at 5% level of the null hypothesis as a function of $r = \phi(h_0)$ for the SQLR (dashed blue line) and for the bootstrap SQLR (solid red line).

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 $\widehat{\Sigma}_0(x) = \left(\sum_{j=1}^n \widehat{U}_j^2 p^{J_n}(X_j)'\right)(P'P)^- p^{J_n}(x)$. (See Theorem B.1 in Appendix B for the definition and consistency of $\widehat{V}_2$ as another sieve variance estimator for any plug-in PSMD $\phi(\widehat{\alpha})$.)

Table 7.2 reports the results for different choices of bases for $q^{k(n)}$ and p^{J_n}, and for different values of $k(n)$ and J_n; in all cases we use a very small $\lambda_n = 10^{-5}$. This table 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. It also 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 $j = 1, 2$.

We can see that the two sieve variance estimators have almost identical performance and the associated sieve t tests have good rejection probabilities. These results are fairly robust to different choices of basis for $q^{k(n)}$ and p^{J_n} and different values of $k(n)$ and J_n as long as $J_n > k(n) + 1$. Figure 7.2 (first row) 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 $j = 1, 2$ for the case Pol(4)-Pol(16) in the table; the right panel in the first row corresponds to $\hat{t}_1$ and the left panel in the first row to $\hat{t}_2$. Both sieve t tests are almost identical to each other and to the standard normal.

NPIV model: sieve variance estimators for an irregular *nonlinear* functional. This case is identical to the previous one for the NPIV model, 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 a nonlinear functional.

Table 7.3 shows Med_{MC} and rejection probabilities for this nonlinear case. By comparing the results with those in Table 7.2 we note that the results are very similar in both cases; the rejections probabilities being slightly higher for the nonlinear functional case. Overall, we think that these results suggest that our sieve t tests perform equally well for both functionals. Figure 7.2 (second row) shows the QQ-Plot for the two sieve t tests for the non-linear case; the right panel in the

$q^{k(n)}$	p^{J_n}	Med_{MC} $\widehat{V}_1$	Med_{MC} $\widehat{V}_2$	5% $\widehat{V}_1$	5% $\widehat{V}_2$	10% $\widehat{V}_1$	10% $\widehat{V}_2$
Pol(4)	Pol(6)	0.0946	0.0937	0.0512	0.0514	0.0980	0.0974
	Pol(10)	0.0922	0.0920	0.0536	0.0532	0.0992	0.0990
	Pol(12)	0.0918	0.0917	0.0538	0.0532	0.1002	0.0998
	Pol(16)	0.0911	0.0912	0.0540	0.0538	0.1000	0.0998
Pol(4)	P-Spline(3,2)	0.0939	0.0942	0.051	0.0516	0.0984	0.0986
	P-Spline(3,5)	0.0939	0.0920	0.053	0.0532	0.099	0.0984
	P-Spline(3,11)	0.0923	0.0925	0.055	0.0548	0.1014	0.1014
	P-Spline(3,17)	0.0922	0.0917	0.0542	0.0538	0.100	0.1008
P-Spline(3,2)	Pol(12)	0.0938	0.0930	0.0572	0.0564	0.1082	0.1074
	Pol(16)	0.0936	0.0936	0.0582	0.0578	0.1082	0.1082
	Pol(18)	0.0936	0.0935	0.0580	0.0578	0.1088	0.1086
	Pol(20)	0.0936	0.0937	0.0580	0.0574	0.1086	0.1092
P-Spline(3,2)	P-Spline(3,2)	0.1106	0.1116	0.0606	0.0598	0.1130	0.1120
	P-Spline(3,5)	0.1019	0.1023	0.0584	0.0574	0.1122	0.1116
	P-Spline(3,11)	0.0961	0.0960	0.0572	0.0566	0.1100	0.1094
	P-Spline(3,17)	0.0949	0.0944	0.0570	0.0566	0.1082	0.1080
P-Spline(3,2)	P-Spline(5,3)	0.1007	0.0998	0.0586	0.0576	0.1102	0.1088
	P-Spline(5,6)	0.1011	0.1009	0.0586	0.0578	0.1100	0.1092
	P-Spline(5,12)	0.1007	0.1009	0.0580	0.0572	0.1110	0.1096
	P-Spline(5,18)	0.1009	0.1010	0.0580	0.0570	0.1106	0.1092

Table 7.2: Relative performance of $\hat{V}_1$ and $\hat{V}_2$: $Med_{MC}\left[\left|\frac{\hat{V}_j}{||v_n^*||_{sd}^2} - 1\right|\right]$, and Nominal size and MC rejection frequencies for t tests $\hat{t}_j$ for $j = 1, 2$ for a linear functional of NPIV.

second row corresponds to $\hat{t}_1$ whereas the left panel in the second row corresponds to $\hat{t}_2$. Again both t tests are almost identical to each other, and to the standard normal.

Finally we wish to point out that we have tried other bases such as Hermite polynomials and cosine series and even larger J_n in these two NPIV MC studies, the results are all similar to the ones reported here and hence are not reported due to the lack of space.

7.2 An Empirical Application

We compute SQLR based confidence bands for nonparametric quantile IV Engel curves using the British FES data set from Blundell et al. (2007):

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

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

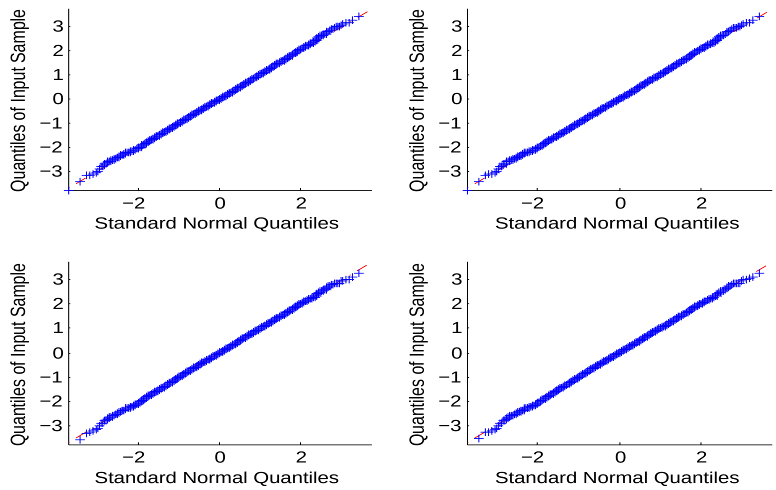

Figure 7.2: QQ-Plot for t tests $\hat{t}_j$ for $j = 1, 2$ for a linear functional (first row) and a nonlinear functional (second row) of NPIV, with $q^{k(n)} = Pol(4)$ and $p^{J_n} = Pol(16)$.

kids" sub-sample of the data set, which consists of $n = 628$ observations. Blundell et al. (2007) estimated NPIV Engel curves using this data set. But, as pointed in Koenker (2005) and others, quantile Engel curves are more informative.

We estimate $h_0(\cdot)$ for foot-in quantile Engel curve 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 $k(n) = 4$, $Pen(h) = ||h||^2_{L^2} + ||\nabla h||^2_{L^2}$ 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 as $p^{J_n}(X)$ and results remained essentially the same. See Chen and Pouzo (2009) for PSMD estimates of NPQIV Engel curves for other non-durable goods.

We use the fact that the optimally weighted SQLR of testing $\phi(h) = h(y_2)$ (for any fixed y_2) is asymptotically χ^2_1 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^2_1.[15] Figure 7.3 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 tighter in the middle.

To test whether the quantile Engel curve for food-in is linear or not, one can test whether $\phi(h_0) \equiv \int \left|\nabla^2 h(y_2)\right|^2 w(y_2)dy_2 = 0$ using our SQLR test. Let $w(\cdot) = (\sigma_{Y_2})^{-1} \exp\left(-\frac{1}{2}(\sigma^{-1}_{Y_2}(\cdot - \mu_{Y_2}))^2\right) 1\{t_{0.01} \leq$

[15]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^2_1. 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.

		Med_{MC}		5%		10%	
$q^{k(n)}$	p^{J_n}	$\widehat{V}_1$	$\widehat{V}_2$	$\widehat{V}_1$	$\widehat{V}_2$	$\widehat{V}_1$	$\widehat{V}_2$
Pol(4)	Pol(6)	0.0990	0.0985	0.0528	0.0530	0.0982	0.0988
	Pol(10)	0.0971	0.0958	0.0524	0.0522	0.1014	0.1012
	Pol(12)	0.0967	0.0959	0.0526	0.0526	0.1020	0.1018
	Pol(16)	0.0961	0.0958	0.0524	0.0528	0.1018	0.1014
Pol(4)	P-Spline(3,2)	0.0996	0.0983	0.0534	0.053	0.0978	0.0976
	P-Spline(3,5)	0.0982	0.0969	0.0538	0.0542	0.099	0.0992
	P-Spline(3,11)	0.0985	0.0984	0.0554	0.0552	0.1014	0.101
	P-Spline(3,17)	0.0982	0.0978	0.0544	0.0546	0.101	0.1008
P-Spline(3,2)	Pol(12)	0.1011	0.1009	0.0580	0.0568	0.1120	0.1122
	Pol(16)	0.1014	0.1005	0.0588	0.0574	0.1128	0.1126
	Pol(18)	0.1014	0.1007	0.0582	0.0568	0.1130	0.1122
	Pol(20)	0.1015	0.1006	0.0580	0.0568	0.1138	0.1128
P-Spline(3,2)	P-Spline(3,2)	0.1191	0.1192	0.0620	0.0612	0.1132	0.1120
	P-Spline(3,5)	0.1090	0.1103	0.0596	0.0594	0.1140	0.1134
	P-Spline(3,11)	0.1028	0.1032	0.0582	0.0572	0.1130	0.1126
	P-Spline(3,17)	0.1029	0.1029	0.0588	0.0580	0.1124	0.1112
P-Spline(3,2)	P-Spline(5,3)	0.1059	0.1064	0.0594	0.0592	0.1114	0.1104
	P-Spline(5,6)	0.1066	0.1076	0.0598	0.0586	0.1124	0.1118
	P-Spline(5,12)	0.1071	0.1079	0.0594	0.0586	0.1126	0.1120
	P-Spline(5,18)	0.1069	0.1079	0.0594	0.0586	0.1122	0.1120

Table 7.3: Relative performance of $\hat{V}_1$ and $\hat{V}_2$: $Med_{MC}\left[\left|\frac{\hat{V}_j}{||v_n^*||_{sd}^2}-1\right|\right]$, and Nominal size and MC rejection frequencies for t tests $\hat{t}_j$ for $j=1,2$ for a nonlinear functional of NPIV.

$\cdot \leq t_{0.99}\}$ where μ_{Y_2}, σ_{Y_2}, $t_{0.01}$ and $t_{0.99}$ are the sample mean, standard deviation and the 1% and 99% quantiles of Y_2. For this specification, the p-value is smaller than 0.0001 and we consequently reject the hypothesis of linearity.[16]

8 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 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

[16] We use the standard Riemann sum with 1000 terms to compute the integral. We also considered other choices of w such that $w(\cdot)=1\{t_{0.25}\leq\cdot\leq t_{0.75}\}$ and $w(\cdot)=1\{t_{0.01}\leq\cdot\leq t_{0.99}\}$. Although the numerical value of the SQLR test changes, all produce p-values below 0.0001.

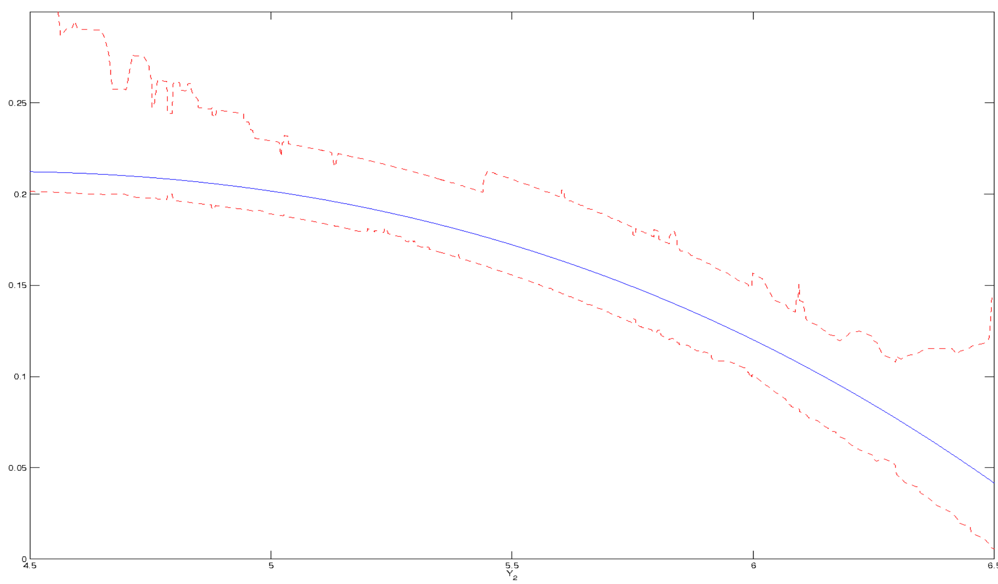

Figure 7.3: PSMD Estimate of the NPQIV food-in Engel curve (blue solid line), with the 95% pointwise confidence bands (red dash lines).

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 sieve variance estimators for $\phi(\widehat{\alpha}_n)$ (in the text and in online Appendix B), 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 (2011) proposed a root-n asymptotically normal estimation of a regular linear functional of h_0 in the NPIV model $E[Y_1 - h_0(Y_2)|X] = 0$, and Santos (2012) considered Bierens' type of test of the NPIV model 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. (2007). Large sample sieve estimation of semi-nonparametric models. *The Handbook of Econometrics, J.J. Heckman and E.E. Leamer (eds.), North-Holland, Amsterdam 6B.*

Chen, X., V. Chernozhukov, S. Lee, and W. Newey (2014). Local identification of nonparametric and semiparametric models. *Econometrica 82*, 785–809.

Chen, X. and T. Christensen (2013). Optimal uniform convergence rates for sieve nonparametric instrumental variables regression. *Cemmap working paper CWP56/13*.

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.

Debnath, L. and P. Mikusinski (1999). *Introduction to Hilbert Spaces with Applications* (2rd ed.), Volume I. Academic Press.

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.

Koenker, R. (2005). *Quantile Regression.* Econometric Society Monograph Series. Cambridge Univesity Press.

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

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

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. (2011). Instrumental variable methods for recovering continuous linear functionals. *Journal of Econometrics 161*, 129–146.

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 Additional Results and Sufficient Conditions

Appendix A consists of several subsections. Subsection A.1 presents additional results on (sieve) Riesz representation of the functional of interest. Subsection A.2 derives the convergence rates of the bootstrap PSMD estimator. Subsection A.3 presents asymptotic properties under local alternatives of the SQLR and the sieve Wald tests, and of their bootstrap versions. Subsection A.4 provides some inference results for functionals of increasing dimension. Subsection A.5 provides some low level sufficient conditions for the high level LQA Assumption 3.6(i) and the bootstrap LQA Assumption Boot.3(i) with series LS estimated conditional mean functions $m(\cdot,\alpha)$. Subsection A.6 states useful lemmas with series LS estimated conditional mean functions $m(\cdot,\alpha)$. See online supplemental Appendix C for the proofs of all the results in this Appendix.

A.1 Additional discussion on (sieve) Riesz representation

The discussion in Subsection 3.2 on Riesz representation seems to depend on the weighting matrix Σ, but, under Assumption 3.1(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 (3.2)), and the space $\mathbf{V}$ (or $\overline{\mathbf{V}}$) under $||\cdot||$ is equivalent to that under $||\cdot||_0$. Therefore, under Assumption 3.1(iv), $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ is bounded on $(\mathbf{V}, ||\cdot||)$ iff $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ is *bounded* on $(\mathbf{V}, ||\cdot||_0)$, i.e.,

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

which corresponds to *non-singular semiparametric efficiency bound*, and in this case we say that $\phi(\cdot)$ is *regular* (at $\alpha=\alpha_0$).[17] Likewise, $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ is unbounded on $(\mathbf{V}, ||\cdot||)$ iff $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ is *unbounded* on $(\mathbf{V}, ||\cdot||_0)$ i.e., $\sup_{v\in\mathbf{V},v\neq 0}\left\{\left|\frac{d\phi(\alpha_0)}{d\alpha}[v]\right|/\|v\|_0\right\}=\infty$, 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 $(\mathbf{V}, ||\cdot||_0)$) 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 $\sup_{v\in\mathbf{V},v\neq 0}\left\{\left|\frac{d\phi(\alpha_0)}{d\alpha}[v]\right|/\|v\|_0\right\}$ explicitly; and hence difficult to verify its root-n estimableness.

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 4.1 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 A.1. *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 $v_{\theta,n}^*=I_n^{11}\lambda$, $\beta_n^*=$*

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

$-I_{n,22}^{-1} I_{n,21}' v_{\theta,n}^{*}$, *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

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

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 (A.1). 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 3.3 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.*

A.2 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 $\widehat{\alpha}_n^B$ (and the restricted bootstrap PSMD estimator $\widehat{\alpha}_n^{R,B}$) under virtually the same conditions as those imposed for the consistency and the convergence rate of the original-sample PSMD estimator $\widehat{\alpha}_n$.

The next assumption is needed to control the difference of the bootstrap criterion function $\widehat{Q}_n^B(\alpha)$ and the original-sample criterion function $\widehat{Q}_n(\alpha)$. 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 $\overline{\delta}_{m,n}^* \geq \delta_n$. Let c_0^* and c^* be finite positive constants.

Assumption A.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^∞}*); (iii)* $\widehat{Q}_n^B(\alpha) \geq c^*\widehat{Q}_n(\alpha) - O_{P_{V^\infty|Z^\infty}}(\delta_n^2)$ *uniformly over* $\mathcal{A}_{osn}$ *wpa1(*P_{Z^∞}*).*

Assumption A.1(i)(ii) is analogous to Assumption 3.3 for the original sample, while Assumption A.1(iii) is analogous to Assumption 3.4(iii) for the original sample. 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 series LS estimator $\widehat{m}(x,\alpha)$, Assumption A.1 can be verified.[18]

Lemma A.1. *Let Assumption A.1(i)(ii) and conditions for Lemma 3.1 hold. Then:*

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

(2) In addition, let Assumption 3.4(i)(ii)(iii) and Assumption A.1(iii) hold, then:

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

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

[18] The verification is amounts to follow the proof of Lemma C.2 of Chen and Pouzo (2012a) except that the original-sample series LS estimator $\widehat{m}(x,\alpha)$ is replaced by its bootstrap version $\widehat{m}^B(x,\alpha)$.

Lemma A.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.

A.3 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 3.5 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.3.1 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.1. *Let conditions for Lemma 3.2 and Proposition B.1 and Assumption 3.6 (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 3.5 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 3.6 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 3.6 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.1 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 5.2.

Theorem A.2. *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.1 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.3.2 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.3. *Let* $\widehat{\alpha}_n$ *be the PSMD estimator (2.2), conditions for Lemma 3.2 and Theorem 4.2 and Assumption 3.6 hold under the local alternatives* $\boldsymbol{\alpha}_n$ *defined in (A.2). Let Assumption 3.5 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.2. *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 5.1.

Theorem A.4. *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.3 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.4 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., $d_\phi \equiv d(n) > 1$), and in fact $d(n)$ 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 3.5 but for vector-valued ϕ. Under Assumption 3.1(iv), we could use Ω_n instead of $\Omega_{sd,n}$ in Assumption A.2(ii)(iii) below.

Assumption A.2. *(i) for each* $j = 1,...,d(n)$, $\frac{d\phi_j(\alpha_0)}{d\alpha}$ *satisfies Assumption 3.5(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;*

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

(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 (3.11) corresponding to the $j-$th sieve Riesz representer.

The next assumption is analogous to Assumption 3.6(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.3. *(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) $s_n n d(n) = o(1)$, $b_n\sqrt{d(n)} = o(1)$, $\sqrt{nd(n)} \times a_n = o(1)$.

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 4.1 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.5. *Let Conditions for Lemma 3.2, Assumptions A.2 and A.3 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.5(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 3.1(iv), is bounded in the sense that $\mathbb{D}_n \asymp I_{d(n)}$ (see Lemma C.1 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 4.3 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 $\Sigma = \Sigma_0$.

Theorem A.6. *Let Conditions for Lemma 3.2 and Proposition B.1 (in Appendix B) hold. Let Assumptions A.2 and A.3 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) $\widehat{QLR}_n(\phi_0) = (\sqrt{n}\Omega_{sd,n}^{-1/2}\mathbf{Z}_n)'\mathbb{D}_n(\sqrt{n}\Omega_{sd,n}^{-1/2}\mathbf{Z}_n) + o_P(\sqrt{d(n)})$;

[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).

(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.6(2) is a multivariate version of Theorem 4.3(2). Theorem A.6(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.6(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.5 and A.6 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.2 and A.3 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.

A.5 Sufficient conditions for LQA(i) and LQA^B(i) with series LS estimator $\widehat{m}$

Assumption A.4. *(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* $j = 1, ..., J_n$ *and* $J_n \log(J_n) = o(n)$ *for* $p^{J_n}(X)$ *a polynomial spline or wavelet or trigonometric polynomial sieve; (iv) There is* $p^{J_n}(X)'\pi$ *such that* $\sup_x |g(x) - p^{J_n}(x)'\pi| = O(b_{m,J_n}) = o(1)$ *uniformly in* $g \in \{m(\cdot,\alpha) : \alpha \in \mathcal{A}_{k(n)}^{M_0}\}$.

Thanks to lemma 5.2 in Chen and Christensen (2013), Assumption A.4(iii) now allows $J_n \log(J_n) = o(n)$ for $p^{J_n}(X)$ being a (tensor-product) wavelet or a trigonometric polynomial in addition to a polynomial spline sieve. 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.5. *(i) There is a sequence* $\{\bar{\rho}_n(Z)\}_n$ *of measurable functions such that* $\sup_{\mathcal{A}_{k(n)}^{M_0}} |\rho(Z,\alpha)| \leq \bar{\rho}_n(Z)$ *a.s.-Z and* $E[|\bar{\rho}_n(Z)|^2|X] \leq const. < \infty$*; (ii) there exist some* $\kappa \in (0,1]$ *and* $K\colon \mathcal{X} \to \mathbb{R}$ *measurable with* $E[|K(X)|^2] \leq const.$ *such that* $\forall\delta > 0$,

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

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.6. *(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.7. *(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.4 and A.5 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.6 and A.7 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.6 Lemmas for series LS estimator $\widehat{m}(x,\alpha)$ and its bootstrap version

The next lemma (Lemma A.2) 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.2. *Let $\widehat{m}^B(\cdot,\alpha)$ be the bootstrap series LS estimator (2.16). Let Assumptions 3.1(iv), 3.4(i)(ii), 4.1(iii), A.4, A.5(i)(ii), and Boot.1 or Boot.2 hold. Then: (1) For all $\delta>0$, there is a $M(\delta)>0$ such that for all $M\geq M(\delta)$,*

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

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

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

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

eventually, with

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

(3) Let Assumption A.5(iii) hold. For all $\delta > 0$, there is $N(\delta)$ such that, for all $n \geq N(\delta)$,

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

with

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

where $\{L_n\}_{n=1}^\infty$ is a slowly divergent sequence of positive real numbers (such a choice of L_n exists under assumption A.5(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.3. *Let all of the conditions for Lemma A.2(2) hold. If Assumptions A.5(iv), A.6 and A.7(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.4. *Let all of the conditions for Lemma A.2(2) hold. If Assumption A.7(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.5. *Let Assumptions 3.1(iv), 3.4(i), 4.1(iii), A.4, A.6(i), A.7(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 $n \geq N(\delta)$,

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

B Supplement: Additional Results and Proofs of the Results in the Main Text

In Appendix B, we provide the proofs of all the lemmas, theorems and propositions stated in the main text. Additioal results on consistent sieve variance estimators and bootstrap sieve t statistics are also presented.

B.1 Proofs for Section 3 on basic conditions

Proof of Lemma 3.3: For **Result (1).** Observe that $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ is bounded on $(\mathbf{V}, ||\cdot||)$; and in this case equation (3.4) holds. By definitions of v_n^* and v^*, we have: $\frac{d\phi(\alpha_0)}{d\alpha}[v] = \langle v_n^*, v\rangle$ and $\frac{d\phi(\alpha_0)}{d\alpha}[v] = \langle v^*, v\rangle$ for all $v \in \overline{\mathbf{V}}_{k(n)}$. Thus

$$\langle v^* - v_n^*, v\rangle = 0 \text{ for all } v \in \overline{\mathbf{V}}_{k(n)} \quad \text{and} \quad \|v^*\|^2 = \|v^* - v_n^*\|^2 + \|v_n^*\|^2 .$$

Since $\overline{\mathbf{V}}_{k(n)}$ is a finite dimensional Hilbert space we have $v_n^* = \arg\min_{v\in\overline{\mathbf{V}}_{k(n)}} \|v^* - v\|$. Since $\overline{\mathbf{V}}_{k(n)}$ is dense in $(\overline{\mathbf{V}}, ||\cdot||)$ we have $\|v^* - v_n^*\| \to 0$ and $\|v_n^*\| \to \|v^*\| < \infty$ as $k(n) \to \infty$.

For **Result (2).** We show this part by contradiction. That is, assume that $\lim_{k(n)\to\infty} \|v_n^*\| = C^* < \infty$. Since $\frac{d\phi(\alpha_0)}{d\alpha}$ is unbounded under $||\cdot||$ in $\mathbf{V}$, we have: for any $M > 0$, there exists a $v_M \in \mathbf{V}$ such that $\left|\frac{d\phi(\alpha_0)}{d\alpha}[v_M]\right| > M||v_M||$.

Since $v_M \in \mathbf{V}$, and $\{\mathbf{V}_k\}_k$ is dense (under $||\cdot||_s$) in $\mathbf{V}$, there exists a sequence $(v_{n,M})_n$ such that $v_{n,M} \in \mathbf{V}_{k(n)}$ and $\lim_{n\to\infty} ||v_{n,M} - v_M||_s = 0$. This result and the fact that $||\cdot|| \le C||\cdot||_s$ for some finite $C > 0$, imply that $\lim_{n\to\infty} ||v_{n,M}|| = ||v_M||$. Also, since $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ is continuous or bounded on $(\mathbf{V}, ||\cdot||_s)$, we have:

$$\lim_{n\to\infty} \left|\frac{d\phi(\alpha_0)}{d\alpha}[v_{n,M} - v_M]\right| = 0.$$

Hence, there exists a $N(M)$ such that

$$\left|\frac{d\phi(\alpha_0)}{d\alpha}[v_{n,M}]\right| \ge M||v_{n,M}||$$

for all $n \ge N(M)$. Since $v_{n,M} \in \mathbf{V}_{k(n)}$, the previous inequality implies that

$$||v_n^*|| = \sup_{v\in\overline{\mathbf{V}}_{k(n)}:\|v\|\neq 0} \frac{\left|\frac{d\phi(\alpha_0)}{d\alpha}[v]\right|}{\|v\|} \ge M$$

for all $n \ge N(M)$. Since M is arbitrary we have $\lim_{k(n)\to\infty} \|v_n^*\| = \infty$. A contradiction. *Q.E.D.*

B.2 Proofs for Section 4 on sieve t (Wald) and SQLR

Lemma B.1. *Let $\widehat{\alpha}_n$ be the PSMD estimator (2.2) and conditions for Lemma 3.2 hold. Let Assumptions 3.5(i) and 3.6(i) hold. Then:*

$$\sqrt{n}\langle u_n^*, \widehat{\alpha}_n - \alpha_0\rangle = -\sqrt{n}\mathbb{Z}_n + o_{P_{Z^\infty}}(1).$$

Proof of Lemma B.1: We note that $n^{-1}\sum_{i=1}^{n}\|\widehat{m}(X_i,\alpha)\|^2_{\widehat{\Sigma}^{-1}}=\widehat{Q}_n(\alpha)$. By Assumption 3.6(i), we have: for any $\epsilon_n\in\mathcal{T}_n$,

$$
\begin{aligned}
&n^{-1}\sum_{i=1}^{n}\|\widehat{m}(X_i,\widehat{\alpha}_n+\epsilon_n u_n^*)\|^2_{\widehat{\Sigma}^{-1}}-n^{-1}\sum_{i=1}^{n}\|\widehat{m}(X_i,\widehat{\alpha}_n)\|^2_{\widehat{\Sigma}^{-1}}\\
=&2\epsilon_n\{\mathbb{Z}_n+\langle u_n^*,\widehat{\alpha}_n-\alpha_0\rangle\}+\epsilon_n^2B_n+o_{P_{Z^\infty}}(r_n^{-1}),
\end{aligned}
\tag{B.1}
$$

where $r_n^{-1}=\max\{\epsilon_n^2,\epsilon_n n^{-1/2},s_n^{-1}\}$ with $s_n^{-1}=o(n^{-1})$, and

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

By adding

$$
E_n(\widehat{\alpha}_n,\epsilon_n)\equiv o(n^{-1})+\lambda_n\left(Pen\left(\widehat{h}_n+\epsilon_n\frac{v_{h,n}^*}{\|v_n^*\|_{sd}}\right)-Pen\left(\widehat{h}_n\right)\right)
$$

to both sides of equation (B.1), we have, by the definition of the approximate minimizer $\widehat{\alpha}_n$ and the fact $\widehat{\alpha}_n+\epsilon_n u_n^*\in\mathcal{A}_{k(n)}$ that, for all $\epsilon_n\in\mathcal{T}_n$

$$
2\epsilon_n\{\mathbb{Z}_n+\langle u_n^*,\widehat{\alpha}_n-\alpha_0\rangle\}+\epsilon_n^2B_n+E_n(\widehat{\alpha}_n,\epsilon_n)+o_{P_{Z^\infty}}(r_n^{-1})\geq 0.
$$

Or, equivalently, for any $\delta>0$ and some $N(\delta)$

$$
P_{Z^\infty}\left(\forall\epsilon_n\ :\ \widehat{\alpha}_n+\epsilon_n u_n^*\in\mathcal{N}_{osn},\ 2\epsilon_n\{\mathbb{Z}_n+\langle u_n^*,\widehat{\alpha}_n-\alpha_0\rangle\}+\epsilon_n^2B_n+E_n(\widehat{\alpha}_n,\epsilon_n)\geq-\delta r_n^{-1}\right)\geq 1-\delta \tag{B.2}
$$

for all $n\geq N(\delta)$. In particular, this holds for $\epsilon_n\equiv\pm\{s_n^{-1/2}+o(n^{-1/2})\}=\pm o(n^{-1/2})$ since $s_n^{-1/2}=o(n^{-1/2})$. Under this choice of ϵ_n, $r_n^{-1}=\max\{s_n^{-1},s_n^{-1/2}n^{-1/2}\}$. Moreover Assumptions 3.2(i)(ii) and 3.4(iv) imply that $E(\widehat{\alpha}_n,\epsilon_n)=o_{P_{Z^\infty}}(n^{-1})$. Thus $\sqrt{n}\epsilon_n^{-1}E(\widehat{\alpha}_n,\epsilon_n)=o_{P_{Z^\infty}}(\sqrt{n}\epsilon_n^{-1}n^{-1})=o_{P_{Z^\infty}}(1)$. Thus, from equation (B.2), it follows,

$$
P_{Z^\infty}\left(A_{n,\delta}\geq\sqrt{n}\{\mathbb{Z}_n+\langle u_n^*,\widehat{\alpha}_n-\alpha_0\rangle\}\geq B_{n,\delta}\right)\geq 1-\delta
$$

eventually, where

$$
A_{n,\delta}\equiv-0.5\sqrt{n}\epsilon_nB_n-\delta\sqrt{n}\epsilon_n^{-1}r_n^{-1}+0.5\delta
$$

and

$$
B_{n,\delta}\equiv-0.5\sqrt{n}\epsilon_nB_n-0.5\sqrt{n}\delta\epsilon_n^{-1}r_n^{-1}-0.5\delta
$$

(here the 0.5δ follows from the previous algebra regarding $\sqrt{n}\epsilon_n^{-1}E(\widehat{\alpha}_n,\epsilon_n)$). Note that $\sqrt{n}\epsilon_n=o(1)$, $B_n=O_{P_{Z^\infty}}(1)$, and $\sqrt{n}\epsilon_n^{-1}r_n^{-1}=\pm\max\{s_n^{-1/2}\sqrt{n},1\}\asymp\pm 1$. Thus

$$
P_{Z^\infty}\left(2\delta\geq\sqrt{n}\{\mathbb{Z}_n+\langle u_n^*,\widehat{\alpha}_n-\alpha_0\rangle\}\geq-2\delta\right)\geq 1-\delta,\ \textit{eventually}.
$$

Hence we have established $\sqrt{n}\langle u_n^*,\widehat{\alpha}_n-\alpha_0\rangle=-\sqrt{n}\mathbb{Z}_n+o_{P_{Z^\infty}}(1)$. *Q.E.D.*

Proof of Theorem 4.1: By Lemma B.1 and Assumption 3.6(ii), we immediately obtain: $\sqrt{n}\langle u_n^*,\widehat{\alpha}_n-\alpha_0\rangle\Rightarrow N(0,1)$. Hence, in order to show the result, it suffices to prove that

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

By Riesz representation Theorem and the orthogonality property of $\alpha_{0,n}$, it follows

$$\frac{d\phi(\alpha_0)}{d\alpha}[\widehat{\alpha}_n-\alpha_{0,n}]=\langle v_n^*,\widehat{\alpha}_n-\alpha_{0,n}\rangle=\langle v_n^*,\widehat{\alpha}_n-\alpha_0\rangle\,.$$

By Assumptions 3.1(iv) and 3.5(i) we have $||v_n^*||_{sd}\asymp||v_n^*||$. This and Assumption 3.5 (ii)(iii) imply

$$\begin{aligned}\sqrt{n}\frac{\phi(\widehat{\alpha}_n)-\phi(\alpha_0)}{||v_n^*||_{sd}}&=\sqrt{n}||v_n^*||_{sd}^{-1}\frac{d\phi(\alpha_0)}{d\alpha}[\widehat{\alpha}_n-\alpha_0]+o_{P_{Z^\infty}}(1)\\&=\sqrt{n}||v_n^*||_{sd}^{-1}\frac{d\phi(\alpha_0)}{d\alpha}[\widehat{\alpha}_n-\alpha_{0,n}]+\sqrt{n}||v_n^*||_{sd}^{-1}\frac{d\phi(\alpha_0)}{d\alpha}[\alpha_{0,n}-\alpha_0]+o_{P_{Z^\infty}}(1)\\&=\sqrt{n}||v_n^*||_{sd}^{-1}\frac{d\phi(\alpha_0)}{d\alpha}[\widehat{\alpha}_n-\alpha_{0,n}]+o_{P_{Z^\infty}}(1)\\&=\sqrt{n}||v_n^*||_{sd}^{-1}\langle v_n^*,\widehat{\alpha}_n-\alpha_0\rangle+o_{P_{Z^\infty}}(1).\end{aligned}$$

Thus

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

and the claimed result now follows from Lemma B.1 and Assumption 3.6(ii). *Q.E.D.*

Proof of Lemma 4.1: By the definitions of $\overline{\mathbf{V}}_{k(n)}$ and the sieve Riesz representer $v_n^*\in\overline{\mathbf{V}}_{k(n)}$ of $\frac{d\phi(\alpha_0)}{d\alpha}[\cdot]$ given in (3.6), we know that $v_n^*=(v_{\theta,n}^{*\prime},v_{h,n}^*(\cdot))'=(v_{\theta,n}^{*\prime},\psi^{k(n)}(\cdot)'\beta_n^*)'\in\overline{\mathbf{V}}_{k(n)}$ solves the following optimization problem:

$$\begin{aligned}\frac{d\phi(\alpha_0)}{d\alpha}[v_n^*]=\|v_n^*\|^2&=\sup_{v=(v_\theta',v_h)'\in\overline{\mathbf{V}}_{k(n)},v\neq0}\frac{\left|\frac{\partial\phi(\alpha_0)}{\partial\theta'}v_\theta+\frac{\partial\phi(\alpha_0)}{\partial h}[v_h(\cdot)]\right|^2}{E\left[\left(\frac{dm(X,\alpha_0)}{d\alpha}[v]\right)'\Sigma(X)^{-1}\left(\frac{dm(X,\alpha_0)}{d\alpha}[v]\right)\right]}\\&=\sup_{\gamma=(v_\theta',\beta')'\in\mathbb{R}^{d_\theta+k(n)},\gamma\neq0}\frac{\gamma'F_nF_n'\gamma}{\gamma'D_n\gamma},\end{aligned}\tag{B.3}$$

where $D_n=E\left[\left(\frac{dm(X,\alpha_0)}{d\alpha}[\overline{\psi}^{k(n)}(\cdot)']\right)'\Sigma(X)^{-1}\left(\frac{dm(X,\alpha_0)}{d\alpha}[\overline{\psi}^{k(n)}(\cdot)']\right)\right]$ is a $(d_\theta+k(n))\times(d_\theta+k(n))$ positive definite matrix such that

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

and $F_n\equiv\left(\frac{\partial\phi(\alpha_0)}{\partial\theta'},\frac{\partial\phi(\alpha_0)}{\partial h}[\psi^{k(n)}(\cdot)']\right)'=\frac{d\phi(\alpha_0)}{d\alpha}[\overline{\psi}^{k(n)}(\cdot)]$ is a $(d_\theta+k(n))\times1$ vector.

The sieve Riesz representation (3.6) becomes: for all $v=\left(v_\theta',\psi^{k(n)}(\cdot)'\beta\right)'\in\overline{\mathbf{V}}_{k(n)}$,

$$\frac{d\phi(\alpha_0)}{d\alpha}[v]=F_n'\gamma=\langle v_n^*,v\rangle=\gamma_n^{*\prime}D_n\gamma\quad\text{for all }\gamma=(v_\theta',\beta')'\in\mathbb{R}^{d_\theta+k(n)}.\tag{B.4}$$

It is obvious that the optimal solution of γ in (B.3) or in (B.4) has a closed-form expression:

$$\gamma_n^*=\left(v_{\theta,n}^{*\prime},\beta_n^{*\prime}\right)'=D_n^{-1}F_n.$$

The sieve Riesz representer is then given by

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

Consequently, $\|v_n^*\|^2 = \gamma_n^{*\prime} D_n \gamma_n^* = F_n' D_n^{-1} F_n$. *Q.E.D.*

Another consistent variance estimator. For $\|v_n^*\|_{sd}^2 = E\left(S_{n,i}^* S_{n,i}^{*\prime}\right)$ given in (3.8) and (4.3), by Lemma 4.1, it has an alternative closed form expression:

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

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

Therefore, in addition to the sieve variance estimator $||\widehat{v}_n^*||_{n,sd}$ given in (4.7), we can define another simple plug-in sieve variance estimator:

$$||\widehat{v}_n^*||_{n,sd}^2 = ||\widehat{v}_n^*||_{n,\widehat{\Sigma}^{-1}\widehat{\Sigma}_0\widehat{\Sigma}^{-1}}^2 = \frac{1}{n}\sum_{i=1}^n \left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n)}{d\alpha}[\widehat{v}_n^*]\right)' \widehat{\Sigma}_i^{-1}\widehat{\Sigma}_{0i}\widehat{\Sigma}_i^{-1}\left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n)}{d\alpha}[\widehat{v}_n^*]\right) \tag{B.5}$$

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.

The sieve variance estimator given in (B.5) can also be expressed as

$$||\widehat{v}_n^*||_{n,sd}^2 = \widehat{V}_2 \equiv \widehat{F}_n' \widehat{D}_n^{-1}\widehat{\Omega}_n\widehat{D}_n^{-1}\widehat{F}_n \quad \text{with} \tag{B.6}$$

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

Assumption B.1. *(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*
(ii) $\sup_{\alpha\in\mathcal{N}_{osn}}\sup_{x\in\mathcal{X}} ||\widehat{E}_n[\rho(z,\alpha)\rho(z,\alpha)'|X=x] - E[\rho(z,\alpha)\rho(z,\alpha)'|X=x]||_e = o_{P_{Z^\infty}}(1)$.

Theorem B.1. *Let Assumption 4.1(i)-(iv), Assumption B.1 and assumptions for Lemma 3.2 hold. Then: Results (1) and (2) of Theorem 4.2 hold with* $||\widehat{v}_n^*||_{n,sd}^2$ *given in (B.5).*

Monte Carlo studies indicate that both sieve variance estimators perform well and similiarly in finite samples.

Proof of Theorems 4.2 and B.1: In the proof we use simplified notation $o_{P_{Z^\infty}}(1) = o_P(1)$. Also, Result (2) trivially follows from Result (1) and Theorem 4.1. So we only show Result (1). For **Result (1),** by the triangle inequality, we have: that

$$\begin{aligned}\left|\frac{||\widehat{v}_n^*||_{n,sd} - ||v_n^*||_{sd}}{||v_n^*||_{sd}}\right| &\le \left|\frac{||\widehat{v}_n^*||_{n,sd} - ||\widehat{v}_n^*||_{sd}}{||v_n^*||_{sd}}\right| + \left|\frac{||\widehat{v}_n^*||_{sd} - ||v_n^*||_{sd}}{||v_n^*||_{sd}}\right| \\ &\le \left|\frac{||\widehat{v}_n^*||_{n,sd} - ||\widehat{v}_n^*||_{sd}}{||v_n^*||_{sd}}\right| + \frac{||\widehat{v}_n^* - v_n^*||_{sd}}{||v_n^*||_{sd}}.\end{aligned}$$

This and the fact $\frac{||\widehat{v}_n^*-v_n^*||_{sd}}{||v_n^*||_{sd}} \asymp \frac{||\widehat{v}_n^*-v_n^*||}{||v_n^*||}$ (under Assumption 3.1(iv)) imply that Result (1) follows from:

$$\frac{||\widehat{v}_n^* - v_n^*||}{||v_n^*||} = o_P(1), \tag{B.7}$$

and

$$\left| \frac{||\widehat{v}_n^*||_{n,sd} - ||\widehat{v}_n^*||_{sd}}{||v_n^*||_{sd}} \right| = o_P(1). \tag{B.8}$$

We will establish results (B.7) and (B.8) in Step 1 and Step 2 below.

Step 1. Observe that result (B.7) is about the consistency of the empirical sieve Riesz representer $\widehat{v}_n^*$ in $||\cdot||$ norm, which is the same whether we use $\hat{\rho}_i\hat{\rho}_i'$ or $\hat{\Sigma}_{0i}$ to compute the sieve variance estimators (4.7) or (B.5). By the Riesz representation theorem, we have for all $v \in \overline{\mathbf{V}}_{k(n)}$,

$$\frac{d\phi(\widehat{\alpha}_n)}{d\alpha}[v] = \langle \widehat{v}_n^*, v\rangle_{n,\widehat{\Sigma}^{-1}} \quad \text{and} \quad \frac{d\phi(\alpha_0)}{d\alpha}[v] = \langle v_n^*, v\rangle = \langle v_n^*, v\rangle_{\Sigma^{-1}}.$$

Hence, by Assumption 4.1(i), we have:

$$\begin{aligned}
o_P(1) &= \sup_{v\in\overline{\mathbf{V}}_{k(n)}} \left| \frac{\langle \widehat{v}_n^*, v\rangle_{n,\widehat{\Sigma}^{-1}} - \langle v_n^*, v\rangle}{||v||} \right| \\
&= \sup_{v\in\overline{\mathbf{V}}_{k(n)}} \left| \frac{\langle \widehat{v}_n^*, v\rangle_{n,\widehat{\Sigma}^{-1}} - \langle \widehat{v}_n^*, v\rangle}{||\widehat{v}_n^*|| \times ||v||} ||\widehat{v}_n^*|| + \frac{\langle \widehat{v}_n^*, v\rangle - \langle v_n^*, v\rangle}{||v||} \right| \\
&\geq \sup_{v\in\overline{\mathbf{V}}_{k(n)}} \left| \frac{\langle \widehat{v}_n^* - v_n^*, v\rangle}{||v||} \right| - \sup_{\varpi\in\overline{\mathbf{V}}_{k(n)}:||\varpi||=1} \left| \langle \widehat{\varpi}_n^*, \varpi\rangle_{n,\widehat{\Sigma}^{-1}} - \langle \widehat{\varpi}_n^*, \varpi\rangle \right| \times ||\widehat{v}_n^*||,
\end{aligned}$$

where $\varpi \equiv v/||v||$ and $\widehat{\varpi}_n^* \equiv \widehat{v}_n^*/||\widehat{v}_n^*||$. First note that

$$\begin{aligned}
\left| \langle \widehat{\varpi}_n^*, \varpi\rangle_{n,\widehat{\Sigma}^{-1}} - \langle \widehat{\varpi}_n^*, \varpi\rangle \right| &\leq \left| \langle \widehat{\varpi}_n^*, \varpi\rangle_{n,\widehat{\Sigma}^{-1}} - \langle \widehat{\varpi}_n^*, \varpi\rangle_{n,\Sigma^{-1}} \right| + \left| \langle \widehat{\varpi}_n^*, \varpi\rangle_{n,\Sigma^{-1}} - \langle \widehat{\varpi}_n^*, \varpi\rangle_{\Sigma^{-1}} \right| \\
&\equiv |T_{1n}(\varpi)| + |T_{2n}(\varpi)|.
\end{aligned}$$

By Assumption 4.1(ii), we have: $\sup_{\varpi\in\overline{\mathbf{V}}_{k(n)}:||\varpi||=1} |T_{2n}(\varpi)| = o_P(1)$. Note that

$$T_{1n}(\varpi) = n^{-1}\sum_{i=1}^{n} \left(\frac{d\widehat{m}(X_i, \widehat{\alpha}_n)}{d\alpha}[\widehat{\varpi}_n^*] \right)' \{\widehat{\Sigma}^{-1}(X_i) - \Sigma^{-1}(X_i)\} \left(\frac{d\widehat{m}(X_i, \widehat{\alpha}_n)}{d\alpha}[\varpi] \right).$$

By the triangle inequality, Assumptions 3.1(iv) and 4.1(ii)(iii), we obtain

$$\begin{aligned}
|T_{1n}(\varpi)| &\leq \sup_{x\in\mathcal{X}} ||\widehat{\Sigma}^{-1}(x) - \Sigma^{-1}(x)||_e \sqrt{n^{-1}\sum_{i=1}^{n} \left\| \frac{d\widehat{m}(X_i, \widehat{\alpha}_n)}{d\alpha}[\widehat{\varpi}_n^*] \right\|_e^2} \sqrt{n^{-1}\sum_{i=1}^{n} \left\| \frac{d\widehat{m}(X_i, \widehat{\alpha}_n)}{d\alpha}[\varpi] \right\|_e^2} \\
&\leq o_P(1) \times O_P\left(\sqrt{\langle \widehat{\varpi}_n^*, \widehat{\varpi}_n^*\rangle_{n,\Sigma^{-1}}} \times \sqrt{\langle \varpi, \varpi\rangle_{n,\Sigma^{-1}}} \right) = o_P(1) \times O_P(1) = o_P(1).
\end{aligned}$$

Hence

$$\sup_{v\in\overline{\mathbf{V}}_{k(n)}} \left| \frac{\langle \widehat{v}_n^* - v_n^*, v\rangle}{||v||} \right| = o_P(1) \times ||\widehat{v}_n^*||.$$

In particular, for $v = \widehat{v}_n^* - v_n^*$, this implies

$$\frac{||\widehat{v}_n^* - v_n^*||}{||v_n^*||} = o_P(1) \times \frac{||\widehat{v}_n^*||}{||v_n^*||}.$$

Note that $\frac{||\widehat{v}_n^*||}{||v_n^*||} \leq \frac{||\widehat{v}_n^* - v_n^*||}{||v_n^*||} + 1$, and thus, the previous equation implies

$$\frac{||\widehat{v}_n^* - v_n^*||}{||v_n^*||}(1 - o_P(1)) = o_P(1) \quad \text{and} \quad \frac{||\widehat{v}_n^*||}{||v_n^*||} = O_P(1).$$

Step 2. We now show that result (B.8) holds for the sieve variance estimators $||\hat{v}_n^*||^2_{n,sd}$ defined in (4.7) and (B.5). By Assumption 3.1(iv), we have:

$$\begin{aligned}
\left|\frac{||\widehat{v}_n^*||_{n,sd} - ||\widehat{v}_n^*||_{sd}}{||v_n^*||_{sd}}\right| &= \left|\frac{||\widehat{v}_n^*||_{n,sd} - ||\widehat{v}_n^*||_{sd}}{||\widehat{v}_n^*||_{sd}}\right| \times \frac{||\widehat{v}_n^*||_{sd}}{||v_n^*||_{sd}} \asymp \left|\frac{||\widehat{v}_n^*||_{n,sd}}{||\widehat{v}_n^*||_{sd}} - 1\right| \times \frac{||\widehat{v}_n^*||}{||v_n^*||} \\
&\leq \left(\frac{||\widehat{v}_n^*||_{n,sd}}{||\widehat{v}_n^*||_{sd}} + 1\right)\left|\frac{||\widehat{v}_n^*||_{n,sd}}{||\widehat{v}_n^*||_{sd}} - 1\right| \times \frac{||\widehat{v}_n^*||}{||v_n^*||} = \left|\frac{||\widehat{v}_n^*||^2_{n,sd}}{||\widehat{v}_n^*||^2_{sd}} - 1\right| \times \frac{||\widehat{v}_n^*||}{||v_n^*||} \\
&= \left|||\widehat{\varpi}_n^*||^2_{n,sd} - ||\widehat{\varpi}_n^*||^2_{sd}\right| \times \frac{||\widehat{v}_n^*||^2}{||\widehat{v}_n^*||^2_{sd}} \times \frac{||\widehat{v}_n^*||}{||v_n^*||} \\
&= \left|||\widehat{\varpi}_n^*||^2_{n,sd} - ||\widehat{\varpi}_n^*||^2_{sd}\right| \times O_P(1),
\end{aligned}$$

where $\widehat{\varpi}_n^* \equiv \widehat{v}_n^*/||\widehat{v}_n^*||$, $\frac{||\widehat{v}_n^*||}{||v_n^*||} = O_P(1)$ (by Step 1), and $\frac{||\widehat{v}_n^*||^2}{||\widehat{v}_n^*||^2_{sd}} = O_P(1)$ (by Assumption 3.1(iv) and i.i.d. data). Thus, it suffices to show that

$$\left|||\widehat{\varpi}_n^*||^2_{n,sd} - ||\widehat{\varpi}_n^*||^2_{sd}\right| = o_P(1). \tag{B.9}$$

Step 2a for the Estimator $||\hat{v}_n^*||^2_{n,sd}$ defined in (4.7). We now establish the result (B.9) when the sieve variance estimator is defined in (4.7).

Let $\widehat{M}(Z_i, \alpha) = \hat{\Sigma}_i^{-1}\rho(Z_i, \alpha)\rho(Z_i, \alpha)'\hat{\Sigma}_i^{-1}$ and $M(z, \alpha_0) \equiv \Sigma^{-1}(x)\rho(z, \alpha_0)\rho(z, \alpha_0)'\Sigma^{-1}(x)$ and $M_i = M(Z_i, \alpha_0)$. Also let $\widehat{T}_i[v_n] \equiv \frac{d\hat{m}(X_i,\hat{\alpha}_n)}{d\alpha}[v_n]$, $T_i[v_n] \equiv \frac{dm(X_i,\alpha_0)}{d\alpha}[v_n]$ and $\Sigma(x, \alpha) \equiv E[\rho(Z, \alpha)\rho(Z, \alpha)'|x]$.

It turns out that $\left|||\widehat{\varpi}_n^*||^2_{n,sd} - ||\widehat{\varpi}_n^*||^2_{sd}\right|$ can be bounded above by

$$\begin{aligned}
&\sup_{v_n \in \overline{\mathbf{V}}^1_{k(n)}} \left|n^{-1}\sum_{i=1}^n \widehat{T}_i[v_n]'\widehat{M}(Z_i, \hat{\alpha}_n)\widehat{T}_i[v_n] - n^{-1}\sum_{i=1}^n \widehat{T}_i[v_n]'M_i\widehat{T}_i[v_n]\right| \\
&+ \sup_{v_n \in \overline{\mathbf{V}}^1_{k(n)}} \left|n^{-1}\sum_{i=1}^n \widehat{T}_i[v_n]'M_i\widehat{T}_i[v_n] - E[T_i[v_n]'M_iT_i[v_n]]\right| \\
&+ \sup_{v_n \in \overline{\mathbf{V}}^1_{k(n)}} \left|E[T_i[v_n]'M_iT_i[v_n]] - E[T_i[v_n]'\Sigma^{-1}(X_i)\Sigma(X_i, \alpha_0)\Sigma^{-1}(X_i)T_i[v_n]]\right| \\
&\equiv A_{1n} + A_{2n} + A_{3n}.
\end{aligned}$$

Note that $A_{3n} = 0$ by the fact that $E[M_i|X_i] = \Sigma^{-1}(X_i)\Sigma(X_i, \alpha_0)\Sigma^{-1}(X_i)$, and that $A_{2n} = o_P(1)$ by Assumption 4.1(v). Thus it remains to show that $A_{1n} = o_P(1)$. We note that

$$\begin{aligned}
A_{1n} \leq &\sup_z \sup_{\alpha \in \mathcal{N}_{osn}} ||\widehat{M}(z, \alpha) - M(z, \alpha_0)||_e \sup_{v_n \in \overline{\mathbf{V}}^1_n} \left|n^{-1}\sum_{i=1}^n \widehat{T}_i[v_n]'\widehat{T}_i[v_n]\right| \\
\leq &Const. \times \sup_z \sup_{\alpha \in \mathcal{N}_{osn}} ||\widehat{M}(z, \alpha) - M(z, \alpha_0)||_e \sup_{v_n \in \overline{\mathbf{V}}^1_n} \left|n^{-1}\sum_{i=1}^n \widehat{T}_i[v_n]'M(Z_i, \alpha_0)\widehat{T}_i[v_n]\right|
\end{aligned}$$

where the first inequality follows from the fact that for matrices A and B, $|A'BA| \leq ||A||_e||B||_e$ and Assumption 3.1(iv). Observe that by Assumptions 4.1(iii)(iv) and 3.1(iv),

$$\begin{aligned}
&\sup_z \sup_{\alpha\in\mathcal{N}_{osn}} ||\widehat{M}(z,\alpha) - M(z,\alpha_0)||_e \\
&\leq \sup_z \sup_{\alpha\in\mathcal{N}_{osn}} ||\hat{\Sigma}^{-1}(x)\{\rho(z,\alpha)\rho(z,\alpha)' - \rho(z,\alpha_0)\rho(z,\alpha_0)'\}\hat{\Sigma}^{-1}(x)||_e \\
&+ \sup_z ||\hat{\Sigma}^{-1}(x)\rho(z,\alpha_0)\rho(z,\alpha_0)'\hat{\Sigma}^{-1}(x) - \Sigma^{-1}(x)\rho(z,\alpha_0)\rho(z,\alpha_0)'\Sigma^{-1}(x)||_e.
\end{aligned}$$

The first term in the RHS is $o_P(1)$ by Assumptions 4.1(iii)(iv) and 3.1(iv); the second term in the RHS is also of order $o_P(1)$ by Assumptions 4.1(iii) and 3.1(iv) and the fact that $\rho(Z,\alpha_0)\rho(Z,\alpha_0)' = O_P(1)$. By Assumption 4.1(v), $\sup_{v_n\in\overline{\mathbf{V}}_n^1} \left|n^{-1}\sum_{i=1}^n \widehat{T}_i[v_n]'M(Z_i,\alpha_0)\widehat{T}_i[v_n]\right| = O_P(1)$. Hence $A_{1n} = o_P(1)$ and result (B.9) holds.

Step 2b for the Estimator $||\hat{v}_n^*||^2_{n,sd}$ defined in (B.5). Since we already provide a detailed proof for result (B.9) in Step 2a for the case of (4.7), here we present a more succinct proof for the case of (B.5).

By the triangle inequality,

$$\left|||\widehat{\varpi}_n^*||^2_{n,sd} - ||\widehat{\varpi}_n^*||^2_{sd}\right| \leq \left|||\widehat{\varpi}_n^*||^2_{n,sd} - ||\widehat{\varpi}_n^*||^2_{n,\Sigma^{-1}\Sigma_0\Sigma^{-1}}\right| + \left|||\widehat{\varpi}_n^*||^2_{n,\Sigma^{-1}\Sigma_0\Sigma^{-1}} - ||\widehat{\varpi}_n^*||^2_{sd}\right| \equiv R_{1n} + R_{2n}.$$

By Assumptions 3.1(iv), 4.1(iii)(iv) and B.1, we have:

$$\sup_{x\in\mathcal{X}} ||\widehat{\Sigma}^{-1}(x)\widehat{\Sigma}_0(x)\widehat{\Sigma}^{-1}(x) - \Sigma^{-1}(x)\Sigma_0(x)\Sigma^{-1}(x)||_e = o_P(1),$$

where $\widehat{\Sigma}_0(x) = \hat{E}_n[\rho(Z,\hat{\alpha}_n)\rho(Z,\hat{\alpha}_n)'|x]$. Therefore, by Assumptions 3.1(iv) and 4.1(ii) and similar algebra to the one used to bound $T_{1n}(\varpi)$, we have:

$$R_{1n} \leq o_P(1) \times n^{-1}\sum_{i=1}^n \left\|\frac{d\widehat{m}(X_i,\widehat{\alpha}_n)}{d\alpha}[\widehat{\varpi}_n^*]\right\|_e^2 = o_P(1)\times O_P(1) = o_P(1).$$

Also by Assumption B.1, $R_{2n} = o_P(1)$. Thus result (B.9) holds. *Q.E.D.*

Before we prove Theorem 4.3, we introduce some notation that will simplify the presentation of the proofs. For any $\bar{\phi}\in\mathbb{R}$ let $\mathcal{A}(\bar{\phi}) \equiv \{\alpha\in\mathcal{A} : \phi(\alpha) = \bar{\phi}\}$, and $\mathcal{A}_{k(n)}(\bar{\phi}) \equiv \mathcal{A}(\bar{\phi})\cap\mathcal{A}_{k(n)}$. In particular, let $\mathcal{A}^0 \equiv \mathcal{A}(\phi(\alpha_0))$ and $\mathcal{A}^0_{k(n)} \equiv \mathcal{A}_{k(n)}(\phi(\alpha_0))$.

Also, we need to show that for any deviation of α of the type $\alpha + tu_n^*$, there exists a t such that $\phi(\alpha+tu_n^*)$ is "close" to $\phi(\alpha_0)$. Formally,

Lemma B.2. *Let Assumption 3.5 hold. For any $n\in\{1,2,....\}$, any $r\in\{|r|\leq 2M_n||v_n^*||\delta_n\}$, and any $\alpha\in\mathcal{N}_{osn}$, there exists a $t\in\mathcal{T}_n$ such that $\phi(\alpha+tu_n^*)-\phi(\alpha_0) = r$ and $\alpha+tu_n^*\in\mathcal{A}_{k(n)}$.*

Proof of Lemma B.2: We first show that there exists a $t\in\mathcal{T}_n$ such that $\phi(\alpha+tu_n^*)-\phi(\alpha_0) = r$. By Assumption 3.5, there exists a $(F_n)_n$ such that $F_n > 0$ and $F_n = o(n^{-1/2}||v_n^*||)$ and, for any $\alpha\in\mathcal{N}_{osn}$ and $t\in\mathcal{T}_n$,

$$\left|\phi(\alpha+tu_n^*) - \phi(\alpha_0) - \langle v_n^*, \alpha-\alpha_0\rangle - t\frac{||v_n^*||^2}{||v_n^*||_{sd}}\right| \leq F_n. \tag{B.10}$$

(note that by assumption 3.5, F_n does not depend on α nor t).

For any $r \in \{|r| \leq 2M_n||v_n^*||\delta_n\}$, we define $(t_l)_{l=1,2}$ as

$$t_l||u_n^*||^2 = -\langle u_n^*, \alpha - \alpha_0\rangle + a_{l,n}F_n||v_n^*||_{sd}^{-1} + r||v_n^*||_{sd}^{-1}.$$

where $a_l = (-1)^l 2$. Note that, by assumption 3.5(i) (the second part), $||u_n^*||^{-2} \leq c^{-2}$, and thus

$$|t_l| \leq c^{-2}\left(||u_n^*|| \times ||\alpha - \alpha_0|| + 2|F_n| \times ||v_n^*||_{sd}^{-1} + |r| \times ||v_n^*||_{sd}^{-1}\right).$$

Without loss of generality, we can re-normalize M_n so that $c^{-2}C < M_n$ and $C \geq 1$. Hence,

$$\begin{aligned}|t_l| &\leq c^{-2}\left(||u_n^*|| \times ||\alpha - \alpha_0|| + 2|F_n| \times ||v_n^*||_{sd}^{-1} + |r| \times ||v_n^*||_{sd}^{-1}\right)\\ &= c^{-2}\left(||u_n^*|| \times ||\alpha - \alpha_0|| + 2|F_n| \times ||v_n^*||_{sd}^{-1} + |r| \times ||v_n^*||^{-1}||u_n^*||\right)\\ &\leq c^{-2}C\left(||u_n^*|| \times ||\alpha - \alpha_0|| + 2|F_n| \times ||v_n^*||_{sd}^{-1} + |r| \times ||v_n^*||^{-1}\right) \leq 4M_n^2\delta_n,\end{aligned}$$

where the third inequality follows from Assumption 3.5(i) (the second part), and the last inequality follows from the facts that $\alpha \in \mathcal{N}_{osn}$, $c^{-2}C2|F_n| \times ||v_n^*||_{sd}^{-1} = o(n^{-1/2}) \leq M_n^2\delta_n$, $r \in \{|r| \leq 2M_n||v_n^*||\delta_n\}$. Thus, t_l is a valid choice in the sense that $t_l \in \mathcal{T}_n$ for $l = 1, 2$.

Thus, this result and equation (B.10) imply

$$\begin{aligned}\phi(\alpha + t_1u_n^*) - \phi(\alpha_0) \leq& \langle v_n^*, \alpha - \alpha_0\rangle + t_1\frac{||v_n^*||^2}{||v_n^*||_{sd}} + F_n\\ =& ||v_n^*||_{sd}\left(\langle u_n^*, \alpha - \alpha_0\rangle + t_1||u_n^*||^2 + F_n||v_n^*||_{sd}^{-1}\right)\\ =& r - F_n < r.\end{aligned}$$

Hence, $\phi(\alpha + t_1u_n^*) - \phi(\alpha_0) < r$. Similarly,

$$\begin{aligned}\phi(\alpha + t_2u_n^*) - \phi(\alpha_0) \geq& \langle v_n^*, \alpha - \alpha_0\rangle + t_2\frac{||v_n^*||^2}{||v_n^*||_{sd}} - F_n\\ =& ||v_n^*||_{sd}\left(\langle u_n^*, \alpha - \alpha_0\rangle + t_2||u_n^*||^2 - F_n||v_n^*||_{sd}^{-1}\right)\\ =& r + F_n > r\end{aligned}$$

and thus $\phi(\alpha + t_2u_n^*) - \phi(\alpha_0) > r$. Since $t \mapsto \phi(\alpha + tu_n^*)$ is continuous, there exists a $t \in [t_1, t_2]$ such that $\phi(\alpha + tu_n^*) - \phi(\alpha_0) = r$. Clearly, $t \in \mathcal{T}_n$.

The fact that $\alpha(t) \equiv \alpha + tu_n^* \in \mathcal{A}_{k(n)}$ for $\alpha \in \mathcal{N}_{osn}$ and $t \in \mathcal{T}_n$ follows from the fact that the sieve space $\mathcal{A}_{k(n)}$ is assumed to be linear. *Q.E.D.*

Proof of Theorem 4.3: Result (2) directly follows from Result (1) with $\Sigma = \Sigma_0$ and $||u_n^*|| = 1$. The proof of Result (1) consists of several steps.

Step 1. For any $t_n \in \mathcal{T}_n$ wpa1., by Assumption 3.6 and Lemma B.1, we have:

$$\begin{aligned}0.5\left(\widehat{Q}_n(\widehat{\alpha}_n(-t_n)) - \widehat{Q}_n(\widehat{\alpha}_n)\right) =& -t_n\{\mathbb{Z}_n + \langle u_n^*, \widehat{\alpha}_n - \alpha_0\rangle\} + \frac{B_n}{2}t_n^2 + o_{P_{Z^\infty}}(r_n^{-1})\\ =& \frac{B_n}{2}t_n^2 + o_{P_{Z^\infty}}(r_n^{-1}),\end{aligned} \tag{B.11}$$

where $r_n^{-1} = \max\{t_n^2, t_nn^{-1/2}, s_n^{-1}\}$ and $s_n^{-1} = o(n^{-1})$.

And under the null hypothesis, $\widehat{\alpha}_n^R \in \mathcal{N}_{osn} \cap \mathcal{A}^0_{k(n)}$ wpa1,

$$
\begin{aligned}
0.5\left(\widehat{Q}_n(\widehat{\alpha}_n^R(t_n)) - \widehat{Q}_n(\widehat{\alpha}_n^R)\right) =& t_n\{\mathbb{Z}_n + \langle u_n^*, \widehat{\alpha}_n^R - \alpha_0\rangle\} + \frac{B_n}{2}t_n^2 + o_{P_{Z^\infty}}(r_n^{-1}) \\
=& t_n\mathbb{Z}_n + \frac{B_n}{2}t_n^2 + o_{P_{Z^\infty}}(r_n^{-1}),
\end{aligned} \tag{B.12}
$$

where the last line follows from the fact that $t_n\langle u_n^*, \widehat{\alpha}_n^R - \alpha_0\rangle = o_{P_{Z^\infty}}(r_n^{-1})$. To show this, note that under the null hypothesis, $\widehat{\alpha}_n^R \in \mathcal{N}_{osn} \cap \mathcal{A}^0_{k(n)}$ wpa1. This and Assumption 3.5(ii) imply that

$$
\left|\underbrace{\phi(\widehat{\alpha}_n^R) - \phi(\alpha_0)}_{=0} - \frac{d\phi(\alpha_0)}{d\alpha}[\widehat{\alpha}_n^R - \alpha_0]\right| = o_{P_{Z^\infty}}(n^{-1/2}||v_n^*||).
$$

Thus

$$
P_{Z^\infty}\left(\frac{\sqrt{n}}{||v_n^*||}\left|\frac{d\phi(\alpha_0)}{d\alpha}[\widehat{\alpha}_n^R - \alpha_0]\right| < \delta\right) \geq 1 - \delta
$$

eventually. By similar calculations to those in the proof of Theorem 4.1, we have

$$
P_{Z^\infty}\left(\sqrt{n}\left|\langle u_n^*, \widehat{\alpha}_n^R - \alpha_0\rangle\right| < \delta\right) \geq 1 - \delta, \ \textit{eventually}.
$$

Hence, $\langle u_n^*, \widehat{\alpha}_n^R - \alpha_0\rangle = o_{P_{Z^\infty}}(n^{-1/2})$, and thus $t_n\langle u_n^*, \widehat{\alpha}_n^R - \alpha_0\rangle = o_{P_{Z^\infty}}(n^{-1/2}t_n) = o_{P_{Z^\infty}}(r_n^{-1})$.

Step 2. We choose $t_n = -\mathbb{Z}_n B_n^{-1}$. Note that under assumption 3.6, $t_n \in \mathcal{T}_n$ wpa1. By the definition of $\widehat{\alpha}_n$, we have, under the null hypothesis,

$$
\begin{aligned}
0.5\left(\widehat{Q}_n(\widehat{\alpha}_n^R) - \widehat{Q}_n(\widehat{\alpha}_n)\right) \geq& 0.5\left(\widehat{Q}_n(\widehat{\alpha}_n^R) - \widehat{Q}_n(\widehat{\alpha}_n^R(t_n))\right) - o_{P_{Z^\infty}}(n^{-1}) \\
=& \frac{1}{2}\mathbb{Z}_n^2 B_n^{-1} - o_{P_{Z^\infty}}(\max\{B_n^{-2}\mathbb{Z}_n^2, -B_n^{-1}\mathbb{Z}_n n^{-1/2}, s_n^{-1}\}) - o_{P_{Z^\infty}}(n^{-1}) \\
=& \frac{1}{2}\mathbb{Z}_n^2 B_n^{-1} + o_{P_{Z^\infty}}(n^{-1}),
\end{aligned}
$$

where the first inequality follows from the fact that, since $t_n \in \mathcal{T}_n$ and $\widehat{\alpha}_n^R \in \mathcal{N}_{osn}$ wpa1, then $\widehat{\alpha}_n^R(t_n) \in \mathcal{A}_{k(n)}$ wpa1; and the second line follows from equation (B.12) with $t_n = -\mathbb{Z}_n B_n^{-1}$.

Step 3. We choose $t_n^* \in \mathcal{T}_n$ wpa1 such that (a) $\phi(\widehat{\alpha}_n(t_n^*)) = \phi(\alpha_0)$, $\widehat{\alpha}_n(t_n^*) \in \mathcal{A}_{k(n)}$, and (b) $t_n^* = \mathbb{Z}_n \frac{||v_n^*||_{sd}^2}{||v_n^*||^2} + o_{P_{Z^\infty}}(n^{-1/2}) = O_{P_{Z^\infty}}(n^{-1/2})$.

Suppose such a t_n^* exists, then $[r_n(t_n^*)]^{-1} = \max\{(t_n^*)^2, t_n^* n^{-1/2}, o(n^{-1})\} = O_{P_{Z^\infty}}(n^{-1})$. By the definition of $\widehat{\alpha}_n^R$, we have, under the null hypothesis,

$$
\begin{aligned}
0.5\left(\widehat{Q}_n(\widehat{\alpha}_n^R) - \widehat{Q}_n(\widehat{\alpha}_n)\right) \leq& 0.5\left(\widehat{Q}_n(\widehat{\alpha}_n(t_n^*)) - \widehat{Q}_n(\widehat{\alpha}_n)\right) + o_{P_{Z^\infty}}(n^{-1}) \\
=& t_n^*\{\mathbb{Z}_n + \langle u_n^*, \widehat{\alpha}_n - \alpha_0\rangle\} + \frac{B_n}{2}(t_n^*)^2 + o_{P_{Z^\infty}}(n^{-1}) \\
=& \frac{B_n}{2}\left(\mathbb{Z}_n \frac{||v_n^*||_{sd}^2}{||v_n^*||^2} + o_{P_{Z^\infty}}(n^{-1/2})\right)^2 + o_{P_{Z^\infty}}(n^{-1}) \\
=& \frac{1}{2}\mathbb{Z}_n^2 B_n^{-1} + o_{P_{Z^\infty}}(n^{-1}) = \frac{1}{2}\mathbb{Z}_n^2 \frac{||v_n^*||_{sd}^2}{||v_n^*||^2} + o_{P_{Z^\infty}}(n^{-1}),
\end{aligned}
$$

where the second line follows from Assumption 3.6(i) and the fact that t_n^* satisfying (b), $[r_n(t_n^*)]^{-1} = O_{P_{Z^\infty}}(n^{-1})$; the third line follows from equation (B.11) and the fact that t_n^* satisfying (b); and the last line follows from Assumptions 3.5(i) and 3.6(ii), $\left|B_n - ||u_n^*||^2\right| = o_{P_{Z^\infty}}(1)$ and $u_n^* = v_n^* / \left\|v_n^*\right\|_{sd}$.

We now show that there is a $t_n^* \in \mathcal{T}_n$ wpa1 such that (a) and (b) hold. Denote $r \equiv \phi(\widehat{\alpha}_n) - \phi(\alpha_0)$. Since $\widehat{\alpha}_n \in \mathcal{N}_{osn}$ wpa1 and $\phi(\widehat{\alpha}_n) - \phi(\alpha_0) = O_{P_{Z^\infty}}(||v_n^*||/\sqrt{n})$ (see the proof of Theorem 4.1), we have $|r| \leq 2M_n||v_n^*||\delta_n$. Thus, by Lemma B.2, there is a $t_n^* \in \mathcal{T}_n$ wpa1 such that $\widehat{\alpha}_n(t_n^*) = \widehat{\alpha}_n + t_n^* u_n^* \in \mathcal{A}_{k(n)}$ and $\phi(\widehat{\alpha}_n(t_n^*)) = \phi(\alpha_0)$, so (a) holds. Moreover, by Assumption 3.5(ii), such a choice of t_n^* also satisfies

$$\left| \underbrace{\phi(\widehat{\alpha}_n(t_n^*)) - \phi(\alpha_0)}_{=0} - \frac{d\phi(\alpha_0)}{d\alpha}[\widehat{\alpha}_n - \alpha_0 + t_n^* u_n^*] \right| = o_{P_{Z^\infty}}(||v_n^*||/\sqrt{n}).$$

By Assumption 3.5(i) and the definition of $u_n^* = v_n^* / \left\|v_n^*\right\|_{sd}$ we have: $\frac{d\phi(\alpha_0)}{d\alpha}[t_n^* u_n^*] = t_n^* \frac{||v_n^*||^2}{||v_n^*||_{sd}}$. Thus

$$P_{Z^\infty}\left(\frac{\sqrt{n}}{||v_n^*||} \left| \frac{d\phi(\alpha_0)}{d\alpha}[\widehat{\alpha}_n - \alpha_0] + t_n^* \frac{||v_n^*||^2}{||v_n^*||_{sd}} \right| < \delta \right) \geq 1 - \delta$$

eventually. By similar algebra to that in the proof of Theorem 4.1 it follows that the LHS of the equation above is majorized by

$$\begin{aligned} & P_{Z^\infty}\left(\frac{\sqrt{n}}{||v_n^*||} \left| \langle v_n^*, \widehat{\alpha}_n - \alpha_0 \rangle + t_n^* \frac{||v_n^*||^2}{||v_n^*||_{sd}} \right| < \delta \right) + \delta \\ = \; & P_{Z^\infty}\left(\frac{\sqrt{n}}{||v_n^*||} \left| -\mathbb{Z}_n ||v_n^*||_{sd} + t_n^* \frac{||v_n^*||^2}{||v_n^*||_{sd}} \right| < \delta \right) + \delta \\ = \; & P_{Z^\infty}\left(\sqrt{n} \frac{||v_n^*||_{sd}}{||v_n^*||} \left| -\mathbb{Z}_n + t_n^* \frac{||v_n^*||^2}{||v_n^*||_{sd}^2} \right| < \delta \right) + \delta, \end{aligned}$$

where the second line follows from the proof of Lemma B.1. Since $\frac{||v_n^*||_{sd}}{||v_n^*||} \asymp const.$ (by Assumption 3.5(i)), we obtain:

$$P_{Z^\infty}\left(\sqrt{n} \left| t_n^* - \mathbb{Z}_n \frac{||v_n^*||_{sd}^2}{||v_n^*||^2} \right| < \delta \right) \geq 1 - \delta, \; \textit{eventually}.$$

Since $\sqrt{n}\mathbb{Z}_n = O_{P_{Z^\infty}}(1)$ (Assumption 3.6(ii)), we have: $t_n^* = O_{P_{Z^\infty}}(n^{-1/2})$, and in fact, $\sqrt{n} t_n^* = \sqrt{n}\mathbb{Z}_n \frac{||v_n^*||_{sd}^2}{||v_n^*||^2} + o_{P_{Z^\infty}}(1)$ and hence (b) holds. *Q.E.D.*

Let $\mathcal{A}^R \equiv \{\alpha \in \mathcal{A} : \phi(\alpha) = \phi_0\}$ be the restricted parameter space. Then $\alpha_0 \in \mathcal{A}^R$ iff the null hypothesis $H_0 : \phi(\alpha_0) = \phi_0$ holds. Also, $\mathcal{A}_{k(n)}^R \equiv \{\alpha \in \mathcal{A}_{k(n)} \colon \phi(\alpha) = \phi_0\}$ is a sieve space for $\mathcal{A}^R$. Let $\{\bar{\alpha}_{0,n} \in \mathcal{A}_{k(n)}^R\}$ be a sequence such that $||\bar{\alpha}_{0,n} - \alpha_0||_s \leq \inf_{\alpha \in \mathcal{A}_{k(n)}^R} ||\alpha - \alpha_0||_s + o(n^{-1})$.[20]

Assumption B.2. *(i)* $|Pen(\bar{h}_{0,n}) - Pen(h_0)| = O(1)$ *and* $Pen(h_0) < \infty$*; (ii)* $\widehat{Q}_n(\bar{\alpha}_{0,n}) \leq c_0 Q(\bar{\alpha}_{0,n}) + o_{P_{Z^\infty}}(n^{-1})$.

This assumption on $\bar{\alpha}_{0,n} \in \mathcal{A}_{k(n)}^R$ is the same as Assumptions 3.2(ii) and 3.3(i) imposed on $\Pi_n \alpha_0 \in \mathcal{A}_{k(n)}$, and can be verified in the same way provided that $\alpha_0 \in \mathcal{A}^R$.

[20]Sufficient conditions for $\bar{\alpha}_{0,n} \in \mathcal{A}_{k(n)}^R$ to solve $\inf_{\alpha \in \mathcal{A}_{k(n)}^R} \left\|\alpha - \alpha_0\right\|_s$ under the null include either (a) $\mathcal{A}_{k(n)}$ is compact (in $|| \cdot ||_s$) and ϕ is continuous (in $|| \cdot ||_s$), or (b) $\mathcal{A}_{k(n)}$ is convex and ϕ is linear.

Proposition B.1. *Let $\widehat{\alpha}_n^R \in \mathcal{A}_{k(n)}^R$ be the restricted PSMD estimator (4.10) and $\alpha_0 \in \mathcal{A}^R$. Let Assumptions 3.1, 3.2(iii), 3.3(ii), B.2 and $Q(\bar{\alpha}_{0,n}) + o(n^{-1}) = O(\lambda_n) = o(1)$ hold. Then:*

(1) $Pen(\widehat{h}_n^R) = O_{P_{Z^\infty}}(1)$ and $||\widehat{\alpha}_n^R - \alpha_0||_s = o_{P_{Z^\infty}}(1)$;

(2) Further, let $Q(\bar{\alpha}_{0,n}) \asymp Q(\Pi_n\alpha_0)$ and Assumptions 3.2(ii), 3.3(i) and 3.4(i)(ii)(iii) hold. Then: $||\widehat{\alpha}_n^R - \alpha_0|| = O_{P_{Z^\infty}}(\delta_n)$ and $||\widehat{\alpha}_n^R - \alpha_0||_s = O_{P_{Z^\infty}}(||\alpha_0 - \Pi_n\alpha_0||_s + \tau_n\delta_n)$.

Proof of Proposition B.1. The proof is very similar to those for theorem 3.2 and remark 4.1 in Chen and Pouzo (2012a) by recognizing that $\mathcal{A}_{k(n)}^R$ is a sieve for $\alpha_0 \in \mathcal{A}^R$.

For **Result (1)**, we first want to show that $\widehat{\alpha}_n^R \in \mathcal{A}_{k(n)}^R \cap \{Pen(h) \leq M\}$ for some $M > 0$ wpa1-P_{Z^∞}. By definitions of $\widehat{\alpha}_n^R$ and $\bar{\alpha}_{0,n}$, Assumption B.2(i)(ii) and the condition that $Q(\bar{\alpha}_{0,n})+o(n^{-1}) = O(\lambda_n)$, we have:

$$Pen(\widehat{h}_n^R) \leq \frac{\widehat{Q}_n(\bar{\alpha}_{0,n})}{\lambda_n} + Pen(\bar{h}_{0,n}) + \frac{o(n^{-1})}{\lambda_n} \leq \frac{Q(\bar{\alpha}_{0,n}) + o(n^{-1})}{\lambda_n} + O_{P_{Z^\infty}}(1) = O_{P_{Z^\infty}}(1).$$

Therefore, for any $\epsilon > 0$, $\Pr(Pen(\widehat{h}_n^R) \geq M) < \epsilon$ for some M, eventually.

We now show that $\Pr(||\widehat{\alpha}_n^R - \alpha_0||_s \geq \epsilon) = o(1)$ for any $\epsilon > 0$. Let $\mathcal{A}_{k(n)}^{R,M} \equiv \mathcal{A}_{k(n)}^R \cap \{Pen(h) \leq M\}$ and $\mathcal{A}^{R,M} \equiv \mathcal{A}^R \cap \{Pen(h) \leq M\}$. These sets are compact under $||\cdot||_s$ (by Assumption 3.2(iii)). Assumptions 3.1(i)(iv) and B.2(i) imply that $\alpha_0 \in \mathcal{A}^{R,M}$ and $\bar{\alpha}_{0,n} \in \mathcal{A}_{k(n)}^{R,M}$. Under assumption 3.1(ii), $cl\left(\cup_k \mathcal{A}_k\right) \supseteq \mathcal{A}$ and thus $cl\left(\cup_k \mathcal{A}_k^{R,M}\right) \supseteq \mathcal{A}^{R,M}$. Therefore $||\bar{\alpha}_{0,n} - \alpha_0||_s = o(1)$ by the definition of $\bar{\alpha}_{0,n}$ and the fact that $\mathcal{A}_{k(n)}^{R,M}$ being dense in $\mathcal{A}^{R,M}$.

By standard calculations, it follows that, for any $\epsilon > 0$,

$$\Pr(||\widehat{\alpha}_n^R - \alpha_0||_s \geq \epsilon) \leq \Pr\left(\inf_{\mathcal{A}_{k(n)}^{R,M}: ||\alpha-\alpha_0||_s \geq \epsilon} \{\widehat{Q}_n(\alpha) + \lambda_n Pen(h)\} \leq \widehat{Q}_n(\bar{\alpha}_{0,n}) + \lambda_n Pen(\bar{h}_{0,n}) + o_P(n^{-1})\right) + 0.5\epsilon$$

Moreover (up to omitted constants)

$$\begin{aligned}
&\Pr\left(||\widehat{\alpha}_n^R - \alpha_0||_s \geq \epsilon\right) \\
&\leq \Pr\left(\inf_{\mathcal{A}_{k(n)}^{R,M}: ||\alpha-\alpha_0||_s \geq \epsilon} \{Q(\alpha) + \lambda_n Pen(h)\} \leq Q(\bar{\alpha}_{0,n}) + \lambda_n Pen(\bar{h}_{0,n}) + O_P(\bar{\delta}_{m,n}^2) + o_P(n^{-1})\right) + \epsilon \\
&\leq \Pr\left(\inf_{\mathcal{A}^{R,M}: ||\alpha-\alpha_0||_s \geq \epsilon} \{Q(\alpha) + \lambda_n Pen(h)\} \leq Q(\bar{\alpha}_{0,n}) + \lambda_n Pen(\bar{h}_{0,n}) + O_P(\bar{\delta}_{m,n}^2) + o_P(n^{-1})\right) + \epsilon,
\end{aligned}$$

where the first line follows by Assumptions 3.3(ii) and B.2 and the second by $\mathcal{A}_{k(n)}^{R,M} \subseteq \mathcal{A}^{R,M}$. Since $\mathcal{A}^{R,M}$ is compact under $||\cdot||_s$, $\alpha_0 \in \mathcal{A}^{R,M}$ is unique and Q is continuous (Assumption 3.1), then $\inf_{\mathcal{A}^{R,M}: ||\alpha-\alpha_0||_s \geq \epsilon} \{Q(\alpha) + \lambda_n Pen(h)\} \geq c(\epsilon) > 0$; however, the term $Q(\bar{\alpha}_{0,n}) + \lambda_n Pen(\bar{h}_{0,n}) + O_P(\bar{\delta}_{m,n}^2) + o_P(n^{-1}) = o_P(1)$ and thus the desired result follows.

For **Result (2)**, we now show that $||\widehat{\alpha}_n^R - \alpha_0|| = O_{P_{Z^\infty}}(\kappa_n)$ where $\kappa_n^2 \equiv \max\{\delta_n^2, ||\bar{\alpha}_{0,n} - \alpha_0||^2, \lambda_n, o(n^{-1})\}$. Let $\mathcal{A}_{osn}^R = \{\alpha \in \mathcal{A}_{osn} : \phi(\alpha) = \phi(\alpha_0)\}$ and $\mathcal{A}_{os}^R = \{\alpha \in \mathcal{A}_{os} : \phi(\alpha) = \phi(\alpha_0)\}$. Result (1) implies that $\widehat{\alpha}_n^R \in \mathcal{A}_{osn}^R$ wpa1. To show Result (2), we employ analogous arguments to those for Result (2) and obtain that for all large $K > 0$,

$$\begin{aligned}
&\Pr\left(||\widehat{\alpha}_n^R - \alpha_0|| \geq K\kappa_n\right) \\
&\leq \Pr\left(\inf_{\mathcal{A}_{osn}^R:||\alpha-\alpha_0||\geq K\kappa_n} Q(\alpha) + \lambda_n Pen(h) \leq Q(\bar{\alpha}_{0,n}) + \lambda_n Pen(\bar{h}_{0,n}) + O_P(\delta_n^2) + o_P(n^{-1})\right) + \epsilon \\
&\leq \Pr\left(\inf_{\mathcal{A}_{os}^R:||\alpha-\alpha_0||\geq K\kappa_n} ||\alpha - \alpha_0||^2 \leq Const.\{||\bar{\alpha}_{0,n} - \alpha_0||^2 + \lambda_n Pen(\bar{h}_{0,n}) + O_P(\delta_n^2) + o_P(n^{-1})\right) + \epsilon \\
&\leq \Pr\left(K^2\kappa_n^2 \leq Const.||\bar{\alpha}_{0,n} - \alpha_0||^2 + O(\lambda_n) + O_P(\delta_n^2) + o_P(n^{-1})\right) + \epsilon,
\end{aligned}$$

where the first inequality is due to Assumption B.2(ii) and the assumption that $\widehat{Q}_n(\alpha) \geq cQ(\alpha) - O_{P_{Z^\infty}}(\delta_n^2)$ uniformly over $\mathcal{A}_{osn}$; the second inequality is due to Assumption 3.4. By our choice of κ_n the first term in the RHS is zero for large K. So the desired result follows. The fact that κ_n coincides with δ_n follows from the fact that $||\bar{\alpha}_{0,n} - \alpha_0||^2 \asymp Q(\bar{\alpha}_{0,n}) \asymp Q(\Pi_n\alpha_0)$ by assumption in the Proposition.

Finally, the convergence rate under $||\cdot||_s$ is obtain by applying the previous result and the definition of τ_n. *Q.E.D.*

Proof of Theorem 4.4: Since $\sup_{h\in\mathcal{H}} Pen(h) < \infty$, the relevant parameter set is $\mathcal{A}^M \equiv \{\alpha \in \mathcal{A} : Pen(h) \leq M\}$ with $M = \sup_{h\in\mathcal{H}} Pen(h)$, which is non-empty and compact (in $||\cdot||_s$) under Assumptions 3.1(i)(ii) and 3.2(iii). Let $\mathcal{A}^{R,M} = \mathcal{A}^M \cap \{\alpha \in \mathcal{A} : \phi(\alpha) = \phi_0\}$. Since ϕ is continuous in $||\cdot||_s$, $\mathcal{A}^{R,M}$ is also compact (in $||\cdot||_s$). Note that $\alpha_0 \in \mathcal{A}^{R,M}$ iff the null $H_0 : \phi(\alpha_0) = \phi_0$ holds.

If $\mathcal{A}^{R,M}$ is empty, then there does not exist any $\alpha \in \mathcal{A}^M$ such that $\phi(\alpha) = \phi_0$, and hence it holds trivially that $\widehat{QLR}_n(\phi_0) \geq nC$ for some $C > 0$ wpa1.

If $\mathcal{A}^{R,M}$ is non-empty, under Assumption 3.1(iii) we have: $\min_{\alpha\in\mathcal{A}^{R,M}} Q(\alpha)$ is achieved at some point within $\mathcal{A}^{R,M}$, say, $\overline{\alpha} \in \mathcal{A}^{R,M}$. This and Assumption 3.1(i)(iv) imply that $Q(\overline{\alpha}) = \min_{\alpha\in\mathcal{A}^{R,M}} Q(\alpha) > 0 = Q(\alpha_0)$ under the fixed alternatives $H_1 : \phi(\alpha_0) \neq \phi_0$.

By definitions of $\widehat{\alpha}_n$ and $\Pi_n\alpha_0$ and Assumption 3.3(i), we have:

$$\widehat{Q}_n(\widehat{\alpha}_n) \leq \widehat{Q}_n(\Pi_n\alpha_0) \leq c_0 Q(\Pi_n\alpha_0) + o_{P_{Z^\infty}}(n^{-1}).$$

Since $M = \sup_{h\in\mathcal{H}} Pen(h) < \infty$, we also have that $\widehat{\alpha}_n^R \in \mathcal{A}_{k(n)}^{R,M} \subseteq \mathcal{A}_{k(n)}^M$ wpa1, so by Assumption 3.3(ii), we have:

$$\widehat{Q}_n(\widehat{\alpha}_n^R) \geq cQ(\widehat{\alpha}_n^R) - O_{P_{Z^\infty}}(\bar{\delta}_{m,n}^2) \geq c \times \min_{\alpha\in\mathcal{A}^{R,M}} Q(\alpha) - O_{P_{Z^\infty}}(\bar{\delta}_{m,n}^2).$$

Thus

$$\widehat{Q}_n(\widehat{\alpha}_n^R) - \widehat{Q}_n(\widehat{\alpha}_n) \geq c \times \min_{\alpha\in\mathcal{A}^{R,M}} Q(\alpha) - c_0 Q(\Pi_n\alpha_0) - o_{P_{Z^\infty}}(n^{-1}) - O_{P_{Z^\infty}}(\bar{\delta}_{m,n}^2) = cQ(\overline{\alpha}) + o_{P_{Z^\infty}}(1).$$

Thus under the fixed alternatives $H_1 : \phi(\alpha_0) \neq \phi_0$,

$$\liminf_{n\to\infty} \frac{\widehat{QLR}_n(\phi_0)}{n} \geq \liminf_{n\to\infty} [cQ(\overline{\alpha}) + o_{P_{Z^\infty}}(1)] = cQ(\overline{\alpha}) > 0 \quad \text{in probability.}$$

Q.E.D.

A consistent variance estimator for optimally weighted PSMD estimator. To stress the fact that we consider 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 $||v_n^0||_{sd}^2 = Var(S_{n,i}^0) = ||v_n^0||_0^2$. By Theorem 4.1, $||v_n^0||_{sd}^2 = ||v_n^0||_0^2$ 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 criterion $\widehat{Q}_n^0$:

$$\widehat{||v_n^0||_0^2} \equiv \left(\frac{\widehat{Q}_n^0(\widetilde{\alpha}_n) - \widehat{Q}_n^0(\widehat{\alpha}_n)}{\varepsilon_n^2}\right)^{-1}, \tag{B.13}$$

where $\widetilde{\alpha}_n$ is an approximate minimizer of $\widehat{Q}_n^0(\alpha)$ over $\{\alpha \in \mathcal{A}_{k(n)} \colon \phi(\alpha) = \phi(\widehat{\alpha}_n) - \varepsilon_n\}$.

Theorem B.2. *Let $\widehat{\alpha}_n$ be the optimally weighted PSMD estimator (2.2) with $\Sigma = \Sigma_0$, and conditions for Lemma 3.2, Assumptions 3.5 and 3.6 hold with $||v_n^0||_{sd} = ||v_n^0||_0$ and $|B_n - 1| = o_{P_{Z^\infty}}(1)$. Let $cn^{-1/2} \le \frac{\varepsilon_n}{||v_n^0||_0} \le C\delta_n$ for finite constants $c, C > 0$. Then: $\widetilde{\alpha}_n \in \mathcal{N}_{osn}$ wpa1-P_{Z^∞}, and*

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

When $\widehat{\alpha}_n$ is the optimally weighted PSMD estimator of α_0, Theorem B.2 suggests $\widehat{||v_n^0||_0^2}$ defined in (B.13) as an alternative consistent variance estimator for $\phi(\widehat{\alpha}_n)$. Compared to Theorems 4.2 and B.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.

Proof of Theorem B.2 Recall that for the optimally weighted criterion case $u_n^* = v_n^0/||v_n^0||_0$, and hence $||u_n^*|| = 1$, $B_n = 1 + o_{P_{Z^\infty}}(1)$.

We first show that $\widetilde{\alpha}_n \in \mathcal{N}_{osn}$ **wpa1**. Recall that $\widetilde{\alpha}_n$ is defined as an approximate optimally weighted PSMD estimator constrained to $\{\alpha \in \mathcal{A}_{k(n)} \colon \phi(\alpha) = \phi(\widehat{\alpha}_n) - \varepsilon_n\}$. In the following since there is no risk of confusion, we use $_P$ instead of $_{P_{Z^\infty}}$. Define

$$\bar{\alpha}_n \equiv \widehat{\alpha}_n + \frac{\vartheta_n}{||v_n^0||_0}u_n^*$$

where $\vartheta_n \equiv -\varepsilon_n - r_n \asymp \delta_n||v_n^0||_0$ and $r_n = o(n^{-1/2}||v^0||_0)$ (to be determined below). We first show that (a) $\bar{\alpha}_n \in \mathcal{N}_{osn}$ wpa1.; and $\vartheta_n||v_n^0||_0^{-1} \in \mathcal{T}_n$, and (b) $\bar{\alpha}_n \in \{\alpha \in \mathcal{A}_{k(n)} \colon \phi(\alpha) = \phi(\widehat{\alpha}_n) - \varepsilon_n\}$. We note that the definitions of $\bar{\alpha}_n$ and ϑ_n imply

$$||\bar{\alpha}_n - \alpha_0|| \le \delta_n + \vartheta_n||v^0||_0^{-1} \le 2\delta_n,$$

and

$$||\bar{\alpha}_n - \alpha_0||_s \le ||\widehat{\alpha}_n - \alpha_0||_s + \vartheta_n||v^0||_0^{-1}||u_n^*||_s \le \delta_{s,n} + \vartheta_n||v^0||_0^{-1}\tau_n$$

which is of order $\delta_{s,n}$. It is easy to see that $\vartheta_n||v_n^0||_0^{-1} \in \mathcal{T}_n$. Hence (a) $\bar{\alpha}_n \in \mathcal{N}_{osn}$ wpa1. is shown. Regarding (b), by assumption 3.5 and (a),

$$\phi(\bar{\alpha}_n) - \phi(\alpha_0) = \langle v_n^0, \widehat{\alpha}_n - \alpha_0\rangle_0 + \vartheta_n + r_n = \phi(\widehat{\alpha}_n) - \phi(\alpha_0) + \vartheta_n + r_n$$

with $r_n = o(n^{-1/2}||v_n^0||)$. Thus, $\phi(\bar{\alpha}_n) - \phi(\widehat{\alpha}_n) = \vartheta_n + r_n = -\varepsilon_n$, and hence (b) follows.

We now establish the consistency of $\widetilde{\alpha}_n$ using the properties of $\bar{\alpha}_n$. We observe that, for any $\epsilon > 0$,

$$\Pr(||\widetilde{\alpha}_n - \alpha_0||_s \geq \epsilon) \leq \Pr\left(\inf_{\mathcal{B}_n:||\alpha-\alpha_0||_s\geq\epsilon} \widehat{Q}_n(\alpha) \leq \widehat{Q}_n(\bar{\alpha}_n) + o(n^{-1}) + \lambda_n Pen(\bar{h}_n)\right)$$

where $\mathcal{B}_n \equiv \{\alpha \in \mathcal{A}_{k(n)}^{M_0} : \phi(\alpha) = \phi(\widehat{\alpha}_n) - \varepsilon_n\}$ and the inequality is valid because $\bar{\alpha}_n \in \mathcal{B}_n$ by (a) and (b). Under (a) and Lemma 3.2, $\lambda_n Pen(\bar{h}_n) = O_P(\lambda_n) = o(n^{-1})$.

By (a), under assumption 3.6(i)

$$\begin{aligned}\widehat{Q}_n(\bar{\alpha}_n) =& \widehat{Q}_n(\widehat{\alpha}_n) + \vartheta_n ||v_n^0||_0^{-1}\{\mathbb{Z}_n + \langle u_n^*, \widehat{\alpha}_n - \alpha_0\rangle\} + 0.5(\vartheta_n ||v_n^0||_0^{-1})^2 \\ &+ o_P(\vartheta_n ||v_n^0||_0^{-1} n^{-1/2} + (\vartheta_n ||v_n^0||_0^{-1})^2 + o(n^{-1})).\end{aligned}$$

By Lemma B.1, $\mathbb{Z}_n + \langle u_n^*, \widehat{\alpha}_n - \alpha_0\rangle = o_P(n^{-1/2})$ and thus, given that $\vartheta_n = -\varepsilon_n - r_n \asymp \delta_n ||v_n^0||_0$, the previous display implies that

$$\widehat{Q}_n(\bar{\alpha}_n) \leq \widehat{Q}_n(\widehat{\alpha}_n) + o_P(n^{-1/2}\delta_n + \delta_n^2 + o(n^{-1})) \leq O_P(\delta_n^2)$$

Therefore,

$$\Pr(||\widetilde{\alpha}_n - \alpha_0||_s \geq \epsilon) \leq \Pr\left(\inf_{\mathcal{B}_n:||\alpha-\alpha_0||_s\geq\epsilon} \widehat{Q}_n(\alpha) \leq \widehat{Q}_n(\widehat{\alpha}_n) + O(\lambda_n + \delta_n^2)\right).$$

Since $\widehat{Q}_n(\widehat{\alpha}_n) \leq \widehat{Q}_n(\Pi_n\alpha_0) + O(\lambda_n)$ by definition of $\widehat{\alpha}_n$ and from the fact that $\mathcal{B}_n \subseteq \mathcal{A}_{k(n)}^{M_0}$, it follows that

$$\Pr(||\widetilde{\alpha}_n - \alpha_0||_s \geq \epsilon) \leq \Pr\left(\inf_{\mathcal{A}_n^{M_0}:||\alpha-\alpha_0||_s\geq\epsilon} \widehat{Q}_n(\alpha) \leq \widehat{Q}_n(\Pi_n\alpha_0) + O(\lambda_n + \delta_n^2)\right).$$

The rest of the consistency proof follows from identical steps to the standard one; see Chen and Pouzo (2009).

In order to show the rate, by similar arguments to the previous ones

$$\widehat{Q}_n(\widetilde{\alpha}_n) \leq \widehat{Q}_n(\Pi_n\alpha_0) + O(\lambda_n + \delta_n^2),$$

under our assumptions $\widehat{Q}_n(\widetilde{\alpha}_n) \geq c||\widetilde{\alpha}_n - \alpha_0||^2 - O_P(\delta_n^2)$ and $\widehat{Q}_n(\Pi_n\alpha_0) \leq c_0 Q(\Pi_n\alpha_0) + o_P(n^{-1})$, so the desired rate under $||\cdot||$ follows. The rate under $||\cdot||_s$ immediately follows using the definition of sieve measure of local ill-posedness τ_n. Thus $\widetilde{\alpha}_n \in \mathcal{N}_{osn}$ wpa1.

We now show that $\frac{\widehat{||v_n^0||_0^2}}{||v_n^0||_0^2} = 1 + o_{P_{Z^\infty}}(1)$. This part of proof consists of several steps that are similar to those in the proof of Theorem 4.3, and hence we omit some details. We first provide an asymptotic expansion for $n(\widehat{Q}_n(\widetilde{\alpha}_n) - \widehat{Q}_n(\widehat{\alpha}_n))$ using Assumption 3.6(i) (with $B_n = 1 + o_{P_{Z^\infty}}(1)$), and then show that this is enough to establish the desired result.

In the following we let $t_n \equiv \varepsilon_n / ||v_n^0||_0$. By the assumption on ε_n we have: $cn^{-1/2} \leq t_n \leq C\delta_n$. Therefore, $t_n \in \mathcal{T}_n$, $t_n = o_{P_{Z^\infty}}(1)$ and $o_{P_{Z^\infty}}\left(\frac{1}{t_n}n^{-1/2}\right) = o_{P_{Z^\infty}}(1)$.

Step 1: First, we note that $\widehat{\alpha}_n \in \mathcal{N}_{osn}$ wpa1, that $-t_n \in \mathcal{T}_n$ and $\widehat{\alpha}_n(-t_n) \in \mathcal{A}_{k(n)}$. So we can apply Assumption 3.6(i) with $\alpha = \widehat{\alpha}_n$ and $-t_n$ as the direction, and obtain:

$$\begin{aligned}\frac{(\widehat{Q}_n(\widehat{\alpha}_n(-t_n)) - \widehat{Q}_n(\widehat{\alpha}_n))}{t_n^2} &= \frac{-2}{t_n}\{\mathbb{Z}_n + \langle u_n^*, \widehat{\alpha}_n - \alpha_0\rangle\} + 1 + o_P\left(\max\left\{1, \frac{n^{-1/2}}{t_n}, \frac{o(n^{-1})}{t_n^2}\right\}\right) \\ &= 1 + o_{P_{Z^\infty}}(1), \end{aligned} \tag{B.14}$$

where the last equality follows from the fact that $\langle u_n^*, \widehat{\alpha}_n - \alpha_0\rangle + \mathbb{Z}_n = o_{P_{Z^\infty}}(n^{-1/2})$ (by Lemma B.1), and that $o_{P_{Z^\infty}}\left(\frac{1}{t_n}n^{-1/2}\right) = o_{P_{Z^\infty}}(1)$ (by our choice of t_n).

STEP 2: Since $\widetilde{\alpha}_n \in \mathcal{N}_{osn}$ wpa1, $t_n \in \mathcal{T}_n$ and $\widetilde{\alpha}_n(t_n) \in \mathcal{A}_{k(n)}$, we can apply Assumption 3.6(i) with $\alpha = \widetilde{\alpha}_n$ and t_n as the direction, and obtain:

$$\begin{aligned}\frac{(\widehat{Q}_n(\widetilde{\alpha}_n(t_n)) - \widehat{Q}_n(\widetilde{\alpha}_n))}{t_n^2} &= \frac{2}{t_n}\{\mathbb{Z}_n + \langle u_n^*, \widetilde{\alpha}_n - \alpha_0\rangle\} + 1 + o_P\left(\max\left\{1, \frac{n^{-1/2}}{t_n}, \frac{o(n^{-1})}{t_n^2}\right\}\right) \\ &= -1 + o_{P_{Z^\infty}}(1), \end{aligned} \tag{B.15}$$

where the last line follows from the definition of the restricted estimator $\widetilde{\alpha}_n$. This is because $\phi(\widetilde{\alpha}_n) = \phi(\widehat{\alpha}_n) - \varepsilon_n$, by Assumptions 3.5(i)(ii),

$$\left|-\varepsilon_n - \frac{d\phi(\alpha_0)}{d\alpha}[\widehat{\alpha}_n - \widetilde{\alpha}_n]\right| = o_{P_{Z^\infty}}(||v_n^0||_0/\sqrt{n}).$$

Hence $\langle v_n^0, \widetilde{\alpha}_n - \alpha_0\rangle = \langle v_n^0, \widehat{\alpha}_n - \alpha_0\rangle - \varepsilon_n + o_{P_{Z^\infty}}(||v_n^0||_0/\sqrt{n})$. This implies that $\mathbb{Z}_n + \langle u_n^*, \widetilde{\alpha}_n - \alpha_0\rangle = -\frac{\varepsilon_n}{||v_n^0||_0} + o_{P_{Z^\infty}}(n^{-1/2}) = -t_n + o_{P_{Z^\infty}}(n^{-1/2})$.

STEP 3: It is easy to see that, from equation (B.15) and by the definition of $\hat{\alpha}_n$,

$$\frac{(\widehat{Q}_n(\widetilde{\alpha}_n) - \widehat{Q}_n(\widehat{\alpha}_n))}{t_n^2} \geq \frac{(\widehat{Q}_n(\widetilde{\alpha}_n)) - \widehat{Q}_n(\widetilde{\alpha}_n(t_n))}{t_n^2} - o_{P_{Z^\infty}}(1) = 1 + o_{P_{Z^\infty}}(1).$$

Also, from equation (B.14), Assumption 3.6(i) and by the definition of $\tilde{\alpha}_n$,

$$\begin{aligned}\frac{(\widehat{Q}_n(\widetilde{\alpha}_n) - \widehat{Q}_n(\widehat{\alpha}_n))}{t_n^2} &\leq \frac{(\widehat{Q}_n(\widehat{\alpha}_n(t_n^*)) - \widehat{Q}_n(\widehat{\alpha}_n))}{t_n^2} + o_{P_{Z^\infty}}(1) \\ &= \frac{2t_n^*\{\mathbb{Z}_n + \langle u_n^*, \widehat{\alpha}_n - \alpha_0\rangle\} + (t_n^*)^2 + o_P\left(\max\left\{(t_n^*)^2, t_n^* n^{-\frac{1}{2}}, o(n^{-1})\right\}\right)}{t_n^2} + o_P(1) \\ &= \frac{-2}{t_n}\{\mathbb{Z}_n + \langle u_n^*, \widehat{\alpha}_n - \alpha_0\rangle\} + 1 + o_{P_{Z^\infty}}(1) \\ &= 1 + o_{P_{Z^\infty}}(1),\end{aligned}$$

provided that there is a $t_n^* \in \mathcal{T}_n$ such that (3a) $\phi(\widehat{\alpha}_n(t_n^*)) = \phi(\widehat{\alpha}_n) - \varepsilon_n$ and (3b) $t_n^* = -t_n \times (1 + o_{P_{Z^\infty}}(1))$. In Step 5 we verify that such a t_n^* exists.

By putting these inequalities together, it follows

$$||v_n^0||_0^2 \frac{\widehat{Q}_n(\widetilde{\alpha}_n) - \widehat{Q}_n(\widehat{\alpha}_n)}{\varepsilon_n^2} = \frac{(\widehat{Q}_n(\widetilde{\alpha}_n) - \widehat{Q}_n(\widehat{\alpha}_n))}{t_n^2} = 1 + o_{P_{Z^\infty}}(1). \tag{B.16}$$

STEP 4: By equation (B.16) we have:

$$\frac{||v_n^0||_0^2}{\widehat{||v_n^0||_0^2}} = 1 + o_{P_{Z^\infty}}(1), \quad \text{with} \quad \widehat{||v_n^0||_0^2} \equiv \left(\frac{\widehat{Q}_n(\widetilde{\alpha}_n) - \widehat{Q}_n(\widehat{\alpha}_n)}{\varepsilon_n^2} \right)^{-1},$$

which implies that $0.5 \le \frac{||v_n^0||_0^2}{\widehat{||v_n^0||_0^2}} \le 1.5$ with probability P_{Z^∞} approaching one. By continuous mapping theorem, we obtain:

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

STEP 5: We now show that there is a $t_n^* \in \mathcal{T}_n$ such that (3a) and (3b) in Step 3 hold. Denote $r \equiv \phi(\widehat{\alpha}_n) - \phi(\alpha_0) - \varepsilon_n$. Since $\varepsilon_n \le C||v_n^0||_0\delta_n$, and $\widehat{\alpha}_n \in \mathcal{N}_{osn}$ wpa1, $\phi(\widehat{\alpha}_n) - \phi(\alpha_0) = O_P(||v_n^0||_0/\sqrt{n})$ (by Theorem 4.1), we have $|r| \le ||v_n^0||_0\delta_n(M_n + C) \le 2M_n||v_n^0||_0\delta_n$ (since $C < M_n$ eventually). Thus, by Lemma B.2, there exists a $t_n^* \in \mathcal{T}_n$ such that $\phi(\widehat{\alpha}_n(t_n^*)) = \phi(\widehat{\alpha}_n) - \varepsilon_n$ and $\widehat{\alpha}_n(t_n^*) = \widehat{\alpha}_n + t_n^* u_n^* \in \mathcal{A}_{k(n)}$, and hence (3a) holds. Moreover, by Assumption 3.5(i)(ii), such a choice of t_n^* also satisfies

$$\left| \underbrace{\phi(\widehat{\alpha}_n(t_n^*)) - \phi(\widehat{\alpha}_n)}_{=-\varepsilon_n} - \frac{d\phi(\alpha_0)}{d\alpha}[t_n^* u_n^*] \right| = o_{P_{Z^\infty}}(||v_n^0||_0 n^{-1/2}).$$

Since $u_n^* = v_n^0/||v_n^0||_0$ for optimally weighted criterion case, we have: $\frac{d\phi(\alpha_0)}{d\alpha}[u_n^*] = ||v_n^0||_0$. Thus

$$\left| -\varepsilon_n - t_n^* ||v_n^0||_0 \right| = o_{P_{Z^\infty}}(||v_n^0||_0 n^{-1/2}).$$

Since $t_n \equiv \varepsilon_n/||v_n^0||_0$, we obtain: $|-t_n - t_n^*| = o_{P_{Z^\infty}}(n^{-1/2})$, and hence

$$t_n^* = -t_n + o_{P_{Z^\infty}}(n^{-1/2}) = -t_n \times (1 + o_{P_{Z^\infty}}(1)) = o_{P_{Z^\infty}}(1)$$

where the second and third equal signs are due to the fact that $cn^{-1/2} \le t_n \le C\delta_n$. Thus (3b) holds. *Q.E.D.*

B.3 Proofs for Section 5 on bootstrap inference

Throughout the Appendices, we sometimes use the simplified term "wpa1" in the bootstrap world while its precise meaning is given in Section 5.

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

Lemma B.3. *Let $\widehat{\alpha}_n^B$ be the bootstrap PSMD estimator and conditions for Lemma 3.2 and Lemma A.1 hold. Let Assumption Boot.3(i) hold. Then: (1) for all $\delta > 0$, there exists a $N(\delta)$ such that for all $n \ge N(\delta)$,*

$$P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\sqrt{n}\left| \langle u_n^*, \widehat{\alpha}_n^B - \alpha_0 \rangle + \mathbb{Z}_n^\omega \right| \ge \delta | Z^n \right) < \delta \right) \ge 1 - \delta.$$

(2) If, in addition, assumptions of Lemma B.1 hold, then

$$\sqrt{n}\langle u_n^*, \widehat{\alpha}_n^B - \widehat{\alpha}_n \rangle = -\sqrt{n}\mathbb{Z}_n^{\omega-1} + o_{P_{V^\infty|Z^\infty}}(1) \ wpa1(P_{Z^\infty}).$$

Proof of Lemma B.3: The proof is very similar to that of Lemma B.1, so we only present the main steps.

For **Result (1)**. Under Assumption Boot.3(i) and using the fact that $\widehat{\alpha}_n^B$ is an approximate minimizer of $\widehat{Q}_n^B(\alpha)+\lambda_n Pen(h)$ on $\mathcal{A}_{k(n)}$, it follows (see the proof of Lemma B.1 for details), for sufficiently large n,

$$P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(2\epsilon_n\{\mathbb{Z}_n^\omega+\langle u_n^*,\widehat{\alpha}_n^B-\alpha_0\rangle\}+\epsilon_n^2 B_n^\omega+E_n(\widehat{\alpha}_n^B,\epsilon_n)\geq-\delta r_n^{-1}|Z^n\right)\geq 1-\delta\right)>1-\delta,$$

where r_n and E_n are defined as in the proof of Lemma B.1, and $\epsilon_n=\pm\{s_n^{-1/2}+o(n^{-1/2})\}$. Dividing by $2\epsilon_n$ and multiplying by $\sqrt{n}$, it follows that

$$P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(A_{n,\delta}^\omega\geq\sqrt{n}\{\mathbb{Z}_n^\omega+\langle u_n^*,\widehat{\alpha}_n^B-\alpha_0\rangle\}\geq B_{n,\delta}^\omega|Z^n\right)\geq 1-\delta\right)>1-\delta$$

eventually, where

$$A_{n,\delta}^\omega\equiv-0.5\sqrt{n}\epsilon_n B_n^\omega-\delta\sqrt{n}\epsilon_n^{-1}r_n^{-1}+0.5\delta$$
$$B_{n,\delta}^\omega\equiv-0.5\sqrt{n}\epsilon_n B_n^\omega-\delta\sqrt{n}\epsilon_n^{-1}r_n^{-1}-0.5\delta.$$

Since $\sqrt{n}\epsilon_n=o(1)$ and $B_n^\omega=O_{P_{V^\infty|Z^\infty}}(1)$ wpa1(P_{Z^∞}) and $|\sqrt{n}\epsilon_n^{-1}r_n^{-1}|\asymp 1$, it follows, for sufficiently large n,

$$P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(2\delta\geq\sqrt{n}\{\mathbb{Z}_n^\omega+\langle u_n^*,\widehat{\alpha}_n^B-\alpha_0\rangle\}\geq-2\delta|Z^n\right)\geq 1-\delta\right)>1-\delta.$$

Or equivalently, for sufficiently large n,

$$P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\left|\sqrt{n}\{\mathbb{Z}_n^\omega+\langle u_n^*,\widehat{\alpha}_n^B-\alpha_0\rangle\}\right|\geq 2\delta|Z^n\right)<\delta\right)\geq 1-\delta.$$

Result (2) directly follows from Result (1) and Lemma B.1. *Q.E.D.*

Proof of Theorem 5.1 We note that Assumption Boot.4 implies that $|n^{-1}\sum_{i=1}^n\widehat{T}_i[v_n]'\hat{M}_i^B\widehat{T}_i[v_n]-\sigma_\omega^2 n^{-1}\sum_{i=1}^n\hat{T}_i[v_n]'\hat{M}_i\hat{T}_i[v_n]|=o_{P_{V^\infty|Z^\infty}}(1)$, uniformly over $v_n\in\overline{\mathbf{V}}_{k(n)}^1$ with $\hat{M}_i=\widehat{M}(Z_i,\widehat{\alpha}_n)$ and $\widehat{T}_i[v_n]\equiv\frac{d\hat{m}(X_i,\hat{\alpha}_n)}{d\alpha}[v_n]$. The rest of the proof follows directly from that of Theorem 4.2(1) for the sieve variance defined in (4.7) case. *Q.E.D.*

Proof of Theorem 5.2 By Lemma B.3 and steps analogous to those used to show Theorem 4.1, it follows

$$\sqrt{n}\frac{\phi(\widehat{\alpha}_n^B)-\phi(\widehat{\alpha}_n)}{\sigma_\omega||v_n^*||_{sd}}=-\sqrt{n}\frac{\mathbb{Z}_n^{\omega-1}}{\sigma_\omega}+o_{P_{V^\infty|Z^\infty}}(1)\ wpa1(P_{Z^\infty}).\tag{B.17}$$

For **Result (1)**, we note that the result for $\widehat{W}_{2,n}^B$ follows directly from Theorem 5.1 and the proof of the Result (1) for $\widehat{W}_{1,n}^B\equiv\sqrt{n}\frac{\phi(\widehat{\alpha}_n^B)-\phi(\widehat{\alpha}_n)}{\sigma_\omega||\widehat{v}_n^*||_{n,sd}}$.

We now focus on establishing Result (1) for $\widehat{W}_{1,n}^B$. Theorem 4.2(1) and equation (B.17) imply that

$$\widehat{W}_{1,n}^B=-\sqrt{n}\frac{\mathbb{Z}_n^{\omega-1}}{\sigma_\omega}+o_{P_{V^\infty|Z^\infty}}(1)\ wpa1(P_{Z^\infty});\tag{B.18}$$

Equation (B.18) and Assumptions 3.6(ii) and Boot.3(ii) imply that:

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

Result (1) now follows from the following two equations:

$$\sup_{t\in\mathbb{R}}|P_{V^\infty|Z^\infty}(\widehat{W}^B_{1,n}\leq t|Z^n)-\Phi(t)|=o_{P_{V^\infty|Z^\infty}}(1)\ wpa1(P_{Z^\infty}),\tag{B.19}$$

and

$$\sup_{t\in\mathbb{R}}|P_{Z^\infty}(\widehat{W}_n\leq t)-\Phi(t)|=o_{P_{Z^\infty}}(1),\tag{B.20}$$

where $\Phi()$ is the cdf of a standard normal. Equation (B.20) follows directly from Theorem 4.2(2) and Polya's theorem (see e.g., Kosorok (2008)). Equation (B.19) follows by the same arguments in Lemma 10.11 in Kosorok (2008) (which are in turn analogous to those used in the proof of Polya's theorem).

Result (2) follows from equation (B.17) and the fact that $||v_n^*||_{sd}\to||v^*||_{sd}\in(0,\infty)$ for regular functionals. *Q.E.D.*

Proof of Theorem 5.3: For **Result (1),** denote

$$\mathcal{F}_n\equiv n\frac{\inf_{\mathcal{A}_{k(n)}(\widehat{\phi}_n)}\widehat{Q}^B_n(\alpha)-\widehat{Q}^B_n(\widehat{\alpha}^B_n)}{\sigma^2_\omega}=\frac{\widehat{QLR}^B_n(\widehat{\phi}_n)}{\sigma^2_\omega}=n\frac{\widehat{Q}^B_n(\widehat{\alpha}^{R,B}_n)-\widehat{Q}^B_n(\widehat{\alpha}^B_n)}{\sigma^2_\omega}+o_{P_{V^\infty|Z^\infty}}(1)\ wpa1(P_{Z^\infty})$$

where $\mathcal{A}_{k(n)}(\widehat{\phi}_n)\equiv\{\alpha\in\mathcal{A}_{k(n)}\colon\phi(\alpha)=\phi(\widehat{\alpha}_n)\}$. Since $o_{P_{V^\infty|Z^\infty}}(1)\ wpa1(P_{Z^\infty})$ will not affect the asymptotics we omit it from the rest of the proof to ease the notational burden. We want to show that for all $\delta>0$, there exists a $N(\delta)$ such that

$$P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\left|\mathcal{F}_n-\left(\sqrt{n}\frac{\mathbb{Z}^{\omega-1}_n}{\sigma_\omega||u^*_n||}\right)^2\right|\geq\delta\mid Z^n\right)<\delta\right)\geq1-\delta$$

for all $n\geq N(\delta)$. We divide the proof in several steps.

Step 1. By assumption $|B^\omega_n-||u^*_n||^2|=o_{P_{V^\infty|Z^\infty}}(1)\ wpa1(P_{Z^\infty})$ and $||u^*_n||\in(c,C)$, we have: $\left|\frac{||u^*_n||^2}{B^\omega_n}-1\right|=o_{P_{V^\infty|Z^\infty}}(1)\ wpa1(P_{Z^\infty})$. Therefore, it suffices to show that

$$P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\left|\mathcal{F}_n-\left(\sqrt{n}\frac{\mathbb{Z}^{\omega-1}_n}{\sigma_\omega\sqrt{B^\omega_n}}\right)^2\right|\geq\delta\mid Z^n\right)<\delta\right)\geq1-\delta\tag{B.21}$$

eventually.

Step 2. By Assumption Boot.3(i), for all $\delta>0$, there is a $M>0$ such that

$$P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\sqrt{n}|\mathbb{Z}^{\omega-1}_n/B^\omega_n|\geq M\mid Z^n\right)<\delta\right)\geq1-\delta$$

eventually. Thus $t_n=-\mathbb{Z}^{\omega-1}_n/B^\omega_n\in\mathcal{T}_n$ wpa1. By the definition of $\widehat{\alpha}^B_n$, and the fact that $\widehat{\alpha}^{R,B}_n\in\mathcal{N}_{osn}$ wpa1 (by Lemma A.1(3)),

$$\mathcal{F}_n\geq n\frac{\widehat{Q}^B_n(\widehat{\alpha}^{R,B}_n)-\widehat{Q}^B_n(\widehat{\alpha}^{R,B}_n(t_n))}{\sigma^2_\omega}-o_{P_{V^\infty|Z^\infty}}(1)\ wpa1(P_{Z^\infty}).$$

By specializing Assumption Boot.3(i) to $\alpha = \widehat{\alpha}_n^{R,B}$ and $t_n = -\mathbb{Z}_n^{\omega-1}/B_n^{\omega}$, it follows

$$\begin{aligned} & 0.5(\widehat{Q}_n^B(\widehat{\alpha}_n^{R,B}(-\frac{\mathbb{Z}_n^{\omega-1}}{B_n^{\omega}})) - \widehat{Q}_n^B(\widehat{\alpha}_n^{R,B})) \\ = \; & -\frac{\mathbb{Z}_n^{\omega-1}}{B_n^{\omega}}\{\mathbb{Z}_n^{\omega} + \langle u_n^*, \widehat{\alpha}_n^{R,B} - \alpha_0\rangle\} + \frac{(\mathbb{Z}_n^{\omega-1})^2}{2B_n^{\omega}} + o_{P_{V^\infty|Z^\infty}}(r_n^{-1}) \; wpa1(P_{Z^\infty}). \end{aligned} \tag{B.22}$$

By Assumption 3.5(i)(ii), and the fact that $\widehat{\alpha}_n^{R,B} \in \mathcal{N}_{osn}$ wpa1,

$$P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\frac{\sqrt{n}}{||v_n^*||}\left|\underbrace{\phi(\widehat{\alpha}_n^{R,B}) - \phi(\widehat{\alpha}_n)}_{=0} - \frac{d\phi(\alpha_0)}{d\alpha}[\widehat{\alpha}_n^{R,B} - \widehat{\alpha}_n]\right| \geq \delta \mid Z^n\right) < \delta\right) \geq 1 - \delta$$

eventually. Also by definition $\frac{d\phi(\alpha_0)}{d\alpha}[\widehat{\alpha}_n^{R,B} - \widehat{\alpha}_n] = \langle v_n^*, \widehat{\alpha}_n^{R,B} - \widehat{\alpha}_n\rangle$. This and Assumption 3.5(i) imply that

$$\sqrt{n}\langle u_n^*, \widehat{\alpha}_n^{R,B} - \widehat{\alpha}_n\rangle = o_{P_{V^\infty|Z^\infty}}(1) \; wpa1(P_{Z^\infty}). \tag{B.23}$$

Equation (B.23) and $\sqrt{n}\langle u_n^*, \widehat{\alpha}_n - \alpha_0\rangle = -\sqrt{n}\mathbb{Z}_n + o_{P_{Z^\infty}}(1)$ (Lemma B.1) imply that

$$\sqrt{n}\langle u_n^*, \widehat{\alpha}_n^{R,B} - \alpha_0\rangle = -\sqrt{n}\mathbb{Z}_n + o_{P_{V^\infty|Z^\infty}}(1) \; wpa1(P_{Z^\infty}).$$

Thus we can infer from equation (B.22) that

$$0.5(\widehat{Q}_n^B(\widehat{\alpha}_n^{R,B}(-\frac{\mathbb{Z}_n^{\omega-1}}{B_n^{\omega}})) - \widehat{Q}_n^B(\widehat{\alpha}_n^{R,B})) = -\frac{(\mathbb{Z}_n^{\omega-1})^2}{2B_n^{\omega}} + o_{P_{V^\infty|Z^\infty}}(r_n^{-1}) \; wpa1(P_{Z^\infty}). \tag{B.24}$$

Since $r_n \leq n$, multiplying both sides by $-2n\sigma_\omega^{-2}$, we obtain:

$$\mathcal{F}_n \geq \left(\sqrt{n}\frac{\mathbb{Z}_n^{\omega-1}}{\sigma_\omega\sqrt{B_n^{\omega}}}\right)^2 - o_{P_{V^\infty|Z^\infty}}(1) \; wpa1(P_{Z^\infty}).$$

Step 3. In order to show

$$\mathcal{F}_n \leq \left(\sqrt{n}\frac{\mathbb{Z}_n^{\omega-1}}{\sigma_\omega\sqrt{B_n^{\omega}}}\right)^2 + o_{P_{V^\infty|Z^\infty}}(1) \; wpa1(P_{Z^\infty}), \tag{B.25}$$

we can repeat the same calculations as in Step 2, provided there exists a $t_n^* \in \mathcal{T}_n$ wpa1 such that **(a)** $\phi(\widehat{\alpha}_n^B(t_n^*)) = \phi(\widehat{\alpha}_n)$ with $\widehat{\alpha}_n^B(t_n^*) \in \mathcal{A}_{k(n)}$, and **(b)** $t_n^* = Z_n^{\omega-1}/||u_n^*||^2 + o_{P_{V^\infty|Z^\infty}}(n^{-1/2}) = O_{P_{V^\infty|Z^\infty}}(n^{-1/2}) \; wpa1(P_{Z^\infty})$.

Because, by (a) and the definition of $\widehat{\alpha}_n^{R,B}$,

$$n\frac{\widehat{Q}_n^B(\widehat{\alpha}_n^{R,B}) - \widehat{Q}_n^B(\widehat{\alpha}_n^B)}{\sigma_\omega^2} \leq n\frac{\widehat{Q}_n^B(\widehat{\alpha}_n^B(t_n^*)) - \widehat{Q}_n^B(\widehat{\alpha}_n^B)}{\sigma_\omega^2} + o_{P_{V^\infty|Z^\infty}}(1) \; wpa1(P_{Z^\infty}).$$

By specializing Assumption Boot.3(i) to $\alpha = \widehat{\alpha}_n^B \in \mathcal{N}_{osn}$ wpa1 (by Lemma A.1(2)), and t_n^* as the direction, it follows

$$
\begin{aligned}
& 0.5(\widehat{Q}_n^B(\widehat{\alpha}_n^B(t_n^*)) - \widehat{Q}_n^B(\widehat{\alpha}_n^B)) \\
= \quad & t_n^*\{\mathbb{Z}_n^\omega + \langle u_n^*, \widehat{\alpha}_n^B - \alpha_0\rangle\} + \frac{B_n^\omega}{2}(t_n^*)^2 + o_{P_{V^\infty|Z^\infty}}(r_n^{-1}) \ wpa1(P_{Z^\infty}) \\
= \quad & \frac{B_n^\omega}{2}\left(\frac{Z_n^{\omega-1}}{||u_n^*||^2} + o_{P_{V^\infty|Z^\infty}}(n^{-1/2})\right)^2 + o_{P_{V^\infty|Z^\infty}}(r_n^{-1}) \ wpa1(P_{Z^\infty}) \\
= \quad & \frac{1}{2}\left(\frac{\mathbb{Z}_n^{\omega-1}}{\sqrt{B_n^\omega}}\right)^2 + o_{P_{V^\infty|Z^\infty}}(r_n^{-1}) \ wpa1(P_{Z^\infty}),
\end{aligned}
$$

where the second equality is due to Lemma B.3(2) and (b), the third equality is due to the assumption $|B_n^\omega - ||u_n^*||^2| = o_{P_{V^\infty|Z^\infty}}(1) \ wpa1(P_{Z^\infty})$ and $||u_n^*|| \in (c, C)$. Thus equation (B.25) holds.

Step 4. We now show that there exists a t_n^* such that **(a)** and **(b)** hold in Step 3.

Let $r \equiv \phi(\hat{\alpha}_n) - \phi(\alpha_0)$. Since $\widehat{\alpha}_n^B \in \mathcal{N}_{osn}$ wpa1, and $\phi(\hat{\alpha}_n) - \phi(\alpha_0) = O_{P_{Z^\infty}}(||v_n^*||/\sqrt{n})$, by Lemma B.2, there is a $t_n^* \in \mathcal{T}_n$ wpa1 satisfying (a) with $\widehat{\alpha}_n^B(t_n^*) = \widehat{\alpha}_n^B + t_n^* u_n^* \in \mathcal{A}_{k(n)}$ and $\phi(\widehat{\alpha}_n^B(t_n^*)) - \phi(\alpha_0) = r$. Moreover, by assumption 3.5(i)(ii), such a choice of t_n^* also satisfies

$$
\left| \underbrace{\phi(\widehat{\alpha}_n^B(t_n^*)) - \phi(\widehat{\alpha}_n)}_{=0} - \frac{d\phi(\alpha_0)}{d\alpha}[\widehat{\alpha}_n^B - \widehat{\alpha}_n + t_n^* u_n^*] \right| = o_{P_{V^\infty|Z^\infty}}(||v_n^*||/\sqrt{n}) \ wpa1(P_{Z^\infty}).
$$

Thus, for sufficiently large n,

$$
P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\frac{\sqrt{n}}{||v_n^*||}\left|\frac{d\phi(\alpha_0)}{d\alpha}[\widehat{\alpha}_n^B - \widehat{\alpha}_n] + t_n^*\frac{||v_n^*||^2}{||v_n^*||_{sd}}\right| \geq \delta \mid Z^n\right) < \delta\right) \geq 1 - \delta.
$$

By Assumption 3.5(i) and Lemma B.3(2), it follows that the LHS of the above equation is majorized by

$$
\begin{aligned}
& P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\frac{\sqrt{n}}{||v_n^*||}\left|\langle v_n^*, \widehat{\alpha}_n^B - \widehat{\alpha}_n\rangle + t_n^*\frac{||v_n^*||^2}{||v_n^*||_{sd}}\right| \geq 2\delta \mid Z^n\right) < \delta\right) + \delta \\
& = P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\frac{\sqrt{n}}{||v_n^*||}\left|-\mathbb{Z}_n^{\omega-1}||v_n^*||_{sd} + t_n^*\frac{||v_n^*||^2}{||v_n^*||_{sd}}\right| \geq 2\delta \mid Z^n\right) < \delta\right) + \delta,
\end{aligned}
$$

Therefore,

$$
\sqrt{n}t_n^* = \sqrt{n}Z_n^{\omega-1}/||u_n^*||^2 + o_{P_{V^\infty|Z^\infty}}(1) \ wpa1(P_{Z^\infty}).
$$

Since $\sqrt{n}Z_n^{\omega-1} = O_{P_{V^\infty|Z^\infty}}(1)$ with probability P_{Z^∞} approaching one (assumption Boot.3(ii)) and $||u_n^*||^2 = O(1)$, we have $t_n^* = O_{P_{V^\infty|Z^\infty}}(n^{-1/2})$ with probability P_{Z^∞} approaching one. Thus (b) holds.

Before we prove **Result (2),** we wish to establish the following **equation (B.26)**:

$$
\left|\mathcal{L}_{V^\infty|Z^\infty}\left(\frac{\widehat{QLR}_n^B(\widehat{\phi}_n)}{\sigma_\omega^2} \mid Z^n\right) - \mathcal{L}\left(\widehat{QLR}_n(\phi_0) \mid H_0\right)\right| = o_{P_{Z^\infty}}(1), \tag{B.26}
$$

where $\mathcal{L}\left(\widehat{QLR}_n(\phi_0) \mid H_0\right)$ denotes the law of $\widehat{QLR}_n(\phi_0)$ under the null $H_0 : \phi(\alpha) = \phi_0$, which will be simply denoted as $\mathcal{L}\left(\widehat{QLR}_n(\phi_0)\right)$ in the rest of the proof. By Result (1), it suffices to show that for any $\delta > 0$, there exists a $N(\delta)$ such that

$$P_{Z^\infty}\left(\sup_{f\in BL_1}\left|E\left[f\left(\left[\frac{\sqrt{n}\mathbb{Z}_n^{\omega-1}}{\sigma_\omega||u_n^*||}\right]^2\right) \mid Z^n\right] - E[f(\widehat{QLR}_n(\phi_0))]\right| \leq \delta\right) \geq 1-\delta$$

for all $n \geq N(\delta)$. Let $\mathbb{Z}$ denote a standard normal random variable (i.e., $\mathbb{Z} \sim N(0,1)$). If the following equation (B.27) holds, which will be shown at the end of the proof of equation (B.26),

$$T_n \equiv \sup_{f\in BL_1}\left|E\left[f\left(\left[\frac{\mathbb{Z}}{||u_n^*||}\right]^2\right)\right] - E[f(\widehat{QLR}_n(\phi_0))]\right| = o(1), \tag{B.27}$$

then, it suffices to show that

$$P_{Z^\infty}\left(\sup_{f\in BL_1}\left|E\left[f\left(\left[\frac{\sqrt{n}\mathbb{Z}_n^{\omega-1}}{\sigma_\omega||u_n^*||}\right]^2\right) \mid Z^n\right] - E\left[f\left(\left[\frac{\mathbb{Z}}{||u_n^*||}\right]^2\right)\right]\right| \leq \delta\right) \geq 1-\delta \tag{B.28}$$

for all $n \geq N(\delta)$.

Suppose we could show that

$$\sup_{f\in BL_1}\left|E\left[f\left(\sqrt{n}\frac{\mathbb{Z}_n^{\omega-1}}{\sigma_\omega||u_n^*||}\right) \mid Z^n\right] - E\left[f\left(\mathbb{Z}||u_n^*||^{-1}\right)\right]\right| \to 0,\ wpa1(P_{Z^\infty}), \tag{B.29}$$

or equivalently,

$$P_{Z^\infty}\left(\left|\mathcal{L}_{V^\infty|Z^\infty}\left(\sqrt{n}\frac{\mathbb{Z}_n^{\omega-1}}{\sigma_\omega||u_n^*||}|Z^n\right) - \mathcal{L}(\mathbb{Z}||u_n^*||^{-1})\right| \leq \delta\right) \geq 1-\delta,\ eventually.$$

Then, by the continuous mapping theorem (see Kosorok (2008) Theorem 10.8 and the discussion in section 10.1.4), we have:

$$P_{Z^\infty}\left(\left|\mathcal{L}_{V^\infty|Z^\infty}\left(\left(\sqrt{n}\frac{\mathbb{Z}_n^{\omega-1}}{\sigma_\omega||u_n^*||}\right)^2 \mid Z^n\right) - \mathcal{L}\left(\left(\mathbb{Z}||u_n^*||^{-1}\right)^2\right)\right| \leq \delta\right) \geq 1-\delta,\ eventually,$$

and hence equation (B.28) follows.

It remains to show equation (B.29). By Assumption Boot.3(ii), and the fact that if a sequence converges in probability, for all subsequence, there exists a subsubsequence that converges almost surely, it follows for all subsequence $(n_k)_k$, there exists a subsubsequence $(n_{k(j)})_j$ such that

$$\left|\mathcal{L}_{V^\infty|Z^\infty}\left(\sqrt{n_{k(j)}}\frac{\mathbb{Z}_{n_{k(j)}}^{\omega-1}}{\sigma_\omega} \mid Z^{n_{k(j)}}\right) - \mathcal{L}(\mathbb{Z})\right| \to 0,\ a.s.-P_{Z^\infty}.$$

Since $||u_{n_{k(j)}}^*|| \in (c, C)$, then there exists a further subsequence (which we still denote as $n_{k(j)}$), such that $\lim_{j\to\infty}||u_{n_{k(j)}}^*|| = d_\infty \in [c, C]$. Also, since $\sqrt{n}\frac{\mathbb{Z}_n^{\omega-1}}{\sigma_\omega}$ is a real valued sequence, by Helly's theorem, convergence in distribution also holds for $(n_{k(j)})_j$. Therefore, by Slutsky theorem,

$$\mathcal{L}_{V^\infty|Z^\infty}\left(\sqrt{n_{k(j)}}\frac{\mathbb{Z}^{\omega-1}_{n_{k(j)}}}{\sigma_\omega||u^*_{n_{k(j)}}||}\mid Z^{n_{k(j)}}\right)-\mathcal{L}\left(\mathbb{Z}d_\infty^{-1}\right)\to 0,\ a.s.-P_{Z^\infty}.$$

Since $\lim_{j\to\infty}||u^*_{n_{k(j)}}||=d_\infty\in[c,C]$ and $\mathbb{Z}$ is bounded in probability, this readily implies

$$\mathcal{L}_{V^\infty|Z^\infty}\left(\sqrt{n_{k(j)}}\frac{\mathbb{Z}^{\omega-1}_{n_{k(j)}}}{\sigma_\omega||u^*_{n_{k(j)}}||}\mid Z^{n_{k(j)}}\right)-\mathcal{L}\left(\mathbb{Z}||u^*_{n_{k(j)}}||^{-1}\right)\to 0,\ a.s.-P_{Z^\infty}.$$

Therefore, it follows that

$$\sup_{f\in BL_1}\left|E\left[f\left(\sqrt{n_{k(j)}}\frac{\mathbb{Z}^{\omega-1}_{n_{k(j)}}}{\sigma_\omega||u^*_{n_{k(j)}}||}\right)\mid Z^{n_{k(j)}}\right]-E\left[f\left(\mathbb{Z}||u^*_{n_{k(j)}}||^{-1}\right)\right]\right|\to 0,\ a.s.-P_{Z^\infty}.$$

Since the argument started with an arbitrary subsequence n_k, equation (B.29) holds.

To conclude the proof of equation (B.26), we now show that equation (B.27) in fact holds (i.e., $T_n=o(1)$). Again, it suffices to show that for any sub-sequence, there exists a sub-sub-sequence such that $T_{n(j)}=o(1)$. For any sub-sequence, since $(||u^*_n||)_n$ is a bounded sequence (under Assumption 3.1(iv)), there exists a further sub-sub-sequence (which we denote as $(n(j))_j$) such that $\lim_{j\to\infty}||u^*_{n(j)}||=d_\infty\in[c,C]$ for finite $c,C>0$. Observe that

$$\begin{aligned}
T_{n(j)}\leq&\sup_{f\in BL_1}\left|E\left[f\left(\left[\frac{\mathbb{Z}}{||u^*_{n(j)}||}\right]^2\right)\right]-E\left[f\left(\left[\frac{\mathbb{Z}}{d_\infty}\right]^2\right)\right]\right|\\
&+\sup_{f\in BL_1}\left|E\left[f\left(\left[\frac{\mathbb{Z}}{d_\infty}\right]^2\right)\right]-E\left[f\left(\left(\frac{||u^*_{n(j)}||}{d_\infty}\right)^2\widehat{QLR}_{n(j)}(\phi_0)\right)\right]\right|\\
&+\sup_{f\in BL_1}\left|E\left[f\left(\widehat{QLR}_{n(j)}(\phi_0)\right)\right]-E\left[f\left(\left(\frac{||u^*_{n(j)}||}{d_\infty}\right)^2\widehat{QLR}_{n(j)}(\phi_0)\right)\right]\right|.
\end{aligned}$$

The first term vanishes because $\mathbb{Z}$ is bounded in probability and $\lim_{j\to\infty}||u^*_{n(j)}||=d_\infty>0$; the third term follows by the same reason (by Theorem 4.3 and Assumption 3.6(ii), $\widehat{QLR}_n(\phi_0)$ is bounded in probability).

Finally, for any $f\in BL_1$, let $f(d_\infty^{-1}\cdot)\equiv f\circ d_\infty^{-2}(\cdot)$. Since $f\circ d_\infty^{-2}$ is bounded and $|f\circ d_\infty^{-2}(t)-f\circ d_\infty^{-2}(s)|\leq d_\infty^{-2}|t-s|\leq c^{-2}|t-s|$, we have $\{f\circ d_\infty^{-2}:f\in BL_1\}\subseteq BL_{c^{-2}}$. Therefore, the second term in the previous display is majorized by $\sup_{f\in BL_{c^{-2}}}\left|E\left[f\left([\mathbb{Z}]^2\right)\right]-E\left[f\left(||u^*_{n(j)}||^2\times\widehat{QLR}_{n(j)}(\phi_0)\right)\right]\right|$. Hence, to conclude the proof we need to show that

$$\lim_{j\to\infty}\sup_{f\in BL_{c^{-2}}}\left|E\left[f\left(\mathbb{Z}^2\right)\right]-E\left[f\left(||u^*_{n(j)}||^2\times\widehat{QLR}_{n(j)}(\phi_0)\right)\right]\right|=0. \tag{B.30}$$

Theorem 4.3 (i.e., $||u^*_n||^2\times\widehat{QLR}_n(\phi_0)=[\sqrt{n}\mathbb{Z}_n]^2+o_P(1)$) and Assumption 3.6(ii) directly imply that the above equation (B.30) actually holds for the whole sequence, which readily implies that for any sub-sequence $(n(j))_j$ there is a sub-sub-sequence (which we still denote as $(n(j))_j$) for which the previous display holds.

Finally for **Result (2)**, we want to show that

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

Let $f_t(\cdot)\equiv 1\{\cdot\le t\}$ for $t\in\mathbb{R}$. Under this notation, the previous display can be cast as

$$A_n\equiv\sup_{t\in\mathbb{R}}\left|E_{P_{V^\infty|Z^\infty}}\left[f_t\left(\frac{\widehat{QLR}_n^B(\hat{\phi}_n)}{\sigma_\omega^2}\right)\mid Z^n\right]-E_{P_{Z^\infty}}\left[f_t\left(\widehat{QLR}_n(\phi_0)\right)\right]\right|=o_{P_{Z^\infty}}(1).$$

Denote $\mathbb{Z}^2\sim\chi_1^2$ and

$$\begin{aligned}
A_{1,n}&\equiv\sup_{t'\in\mathbb{R}}\left|E_{P_{V^\infty|Z^\infty}}\left[f_{t'}\left(||u_n^*||^2\times\frac{\widehat{QLR}_n^B(\hat{\phi}_n)}{\sigma_\omega^2}\right)\mid Z^n\right]-E\left[f_{t'}\left(\mathbb{Z}^2\right)\right]\right|,\\
A_{2,n}&\equiv\sup_{t'\in\mathbb{R}}\left|E_{P_{Z^\infty}}\left[f_{t'}\left(||u_n^*||^2\times\widehat{QLR}_n(\phi_0)\right)\right]-E\left[f_{t'}\left(\mathbb{Z}^2\right)\right]\right|.
\end{aligned}$$

Notice that

$$\begin{aligned}
A_n&=\sup_{t\in\mathbb{R}}\left|E_{P_{V^\infty|Z^\infty}}\left[f_{t||u_n^*||^2}\left(||u_n^*||^2\frac{\widehat{QLR}_n^B(\hat{\phi}_n)}{\sigma_\omega^2}\right)\mid Z^n\right]-E_{P_{Z^\infty}}\left[f_{t||u_n^*||^2}\left(||u_n^*||^2\widehat{QLR}_n(\phi_0)\right)\right]\right|\\
&\le\sup_{t\in\mathbb{R}}\sup_{d\in[c,C]}\left|E_{P_{V^\infty|Z^\infty}}\left[f_{td^2}\left(||u_n^*||^2\frac{\widehat{QLR}_n^B(\hat{\phi}_n)}{\sigma_\omega^2}\right)\mid Z^n\right]-E_{P_{Z^\infty}}\left[f_{td^2}\left(||u_n^*||^2\widehat{QLR}_n(\phi_0)\right)\right]\right|\\
&\le\sup_{t'\in\mathbb{R}}\left|E_{P_{V^\infty|Z^\infty}}\left[f_{t'}\left(||u_n^*||^2\times\frac{\widehat{QLR}_n^B(\hat{\phi}_n)}{\sigma_\omega^2}\right)\mid Z^n\right]-E_{P_{Z^\infty}}\left[f_{t'}\left(||u_n^*||^2\times\widehat{QLR}_n(\phi_0)\right)\right]\right|\\
&\le A_{1,n}+A_{2,n}
\end{aligned}$$

where the first line follows from the property that $f_t(\cdot)=f_{t\lambda}(\lambda\times\cdot)$ for any $\lambda\in\mathbb{R}_+$; the second line follows because by assumption, $||u_n^*||^2\in[c^2,C^2]$; the third line follows simply because $\{1\{\cdot\le t\lambda\}: t\in\mathbb{R}\ \textit{and}\ \lambda\in\mathbb{R}_+\}\subseteq\{1\{\cdot\le t\}: t\in\mathbb{R}\}$. Finally, the last line is due to the triangle inequality and the definitions of $A_{1,n}$ and $A_{2,n}$.

By Theorem 4.3, under the null, $||u_n^*||^2\times\widehat{QLR}_n(\phi_0)$ converges weakly to $\mathbb{Z}^2\sim\chi_1^2$, whose distribution is continuous. Therefore, by Polya's theorem, $A_{2,n}=o(1)$. Similarly,

$$A_{1,n}=\sup_{t'\in\mathbb{R}}\left|P_{V^\infty|Z^\infty}\left(||u_n^*||^2\times\frac{\widehat{QLR}_n^B(\hat{\phi}_n)}{\sigma_\omega^2}\le t'\mid Z^n\right)-P\left(\mathbb{Z}^2\le t'\right)\right|=o_{P_{Z^\infty}}(1)$$

by equation (B.26) and by the same arguments in Lemma 10.11 in Kosorok (2008) (which are in turn analogous to those used in the proof of Polya's theorem). *Q.E.D.*

We first recall some notation introduced in the main text. Let $\mathcal{T}_n\equiv\{t\in\mathbb{R}\colon |t|\le 4M_n^2\delta_n\}$. For $t_n\in\mathcal{T}_n$, $\alpha(t_n)\equiv\alpha+t_nu_n^*$ where $u_n^*=v_n^*/\left\|v_n^*\right\|_{sd}$ and $v_n^*=(v_{\theta,n}^{*\prime},v_{h,n}^*(\cdot))'$. To simplify presentation we use $r_n = r_n(t_n) \equiv \left(\max\{t_n^2, t_n n^{-1/2}, o(n^{-1})\}\right)^{-1}$.

Proof of Lemma 5.1: For **Result (1)**, if $\omega \equiv 1$, then Assumption Boot.3(i) simplifies to

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

iff

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

where $\widehat{\Lambda}_n(\alpha(t_n),\alpha) \equiv 0.5(\widehat{Q}_n(\alpha(t_n)) - \widehat{Q}_n(\alpha))$ and B_n is a Z^n measurable random variable with $B_n = O_{P_{Z^\infty}}(1)$. Therefore, if we could verify Assumption Boot.3(i) in Result (2), we also verify Assumption 3.6(i).

For **Result (2)**, we divide its proof in several steps.

Step 1: We first introduce some notation. Let

$$P_n(Z^n) \equiv P_{V^\infty|Z^\infty}\left(\sup_{(\alpha,t_n)\in\mathcal{N}_{osn}\times\mathcal{T}_n} r_n\left|\widehat{\Lambda}_n^B(\alpha(t_n),\alpha) - t_n\left\{\mathbb{Z}_n^\omega + \langle u_n^*, \alpha-\alpha_0\rangle\right\} - \frac{B_n^\omega}{2}t_n^2\right| \geq \delta \mid Z^n\right).$$

Recall that $\ell_n^B(x,\alpha) \equiv \widetilde{m}(x,\alpha) + \widehat{m}^B(x,\alpha_0)$. Let

$$\widehat{L}_n^B(\alpha(t_n),\alpha) \equiv \frac{1}{2n}\sum_{i=1}^n\left\{\ell_n^B(X_i,\alpha(t_n))'\widehat{\Sigma}(X_i)^{-1}\ell_n^B(X_i,\alpha(t_n)) - \ell_n^B(X_i,\alpha)'\widehat{\Sigma}(X_i)^{-1}\ell_n^B(X_i,\alpha)\right\}.$$

We need to show that $P_{Z^\infty}(P_n(Z^n) < \delta) \geq 1-\delta$ eventually which is equivalent to show that $P_{Z^\infty}(P_n(Z^n) > \delta) \leq \delta$ eventually. Hence, it suffices to show that

$$P_{Z^\infty}(\{P_n'(Z^n) > \delta\} \cap S_n) + P_{Z^\infty}(S_n^C) \leq \delta, \ \textit{eventually},$$

for some event S_n that is measurable with respect to Z^n, and some $P_n'(Z^n) \geq P_n(Z^n)$ *a.s.*, here S_n^C denotes the complement of S_n. In the following we take

$$S_n \equiv \left\{Z^n : P_{V^\infty|Z^\infty}\left(\sup_{(\alpha,t_n)\in\mathcal{N}_{osn}\times\mathcal{T}_n} r_n\left|\widehat{\Lambda}_n^B(\alpha(t_n),\alpha) - \widehat{L}_n^B(\alpha(t_n),\alpha)\right| \geq 0.5\delta \mid Z^n\right) < 0.5\delta\right\},$$

and

$$\begin{aligned} P_n'(Z^n) \equiv & P_{V^\infty|Z^\infty}\left(\sup_{(\alpha,t_n)\in\mathcal{N}_{osn}\times\mathcal{T}_n} r_n\left|\widehat{L}_n^B(\alpha(t_n),\alpha) - t_n\left\{\mathbb{Z}_n^\omega + \langle u_n^*, \alpha-\alpha_0\rangle\right\} - \frac{B_n^\omega}{2}t_n^2\right| \geq 0.5\delta \mid Z^n\right) \\ & + P_{V^\infty|Z^\infty}\left(\sup_{(\alpha,t_n)\in\mathcal{N}_{osn}\times\mathcal{T}_n} r_n\left|\widehat{\Lambda}_n^B(\alpha(t_n),\alpha) - \widehat{L}_n^B(\alpha(t_n),\alpha)\right| \geq 0.5\delta \mid Z^n\right). \end{aligned}$$

It follows that we "only" need to show that

$$P_{Z^\infty}(S_n^C) \leq 0.5\delta \quad \text{and} \quad P_{Z^\infty}(\{P_n'(Z^n) > \delta\} \cap S_n) \leq 0.5\delta, \ \textit{eventually}.$$

Since $P_{Z^\infty}(S_n^C)$ can be expressed as

$$P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\sup_{(\alpha,t_n)\in\mathcal{A}_n\times\mathcal{T}_n} r_n\left|\widehat{\Lambda}_n^B(\alpha(t_n),\alpha)-\widehat{L}_n^B(\alpha(t_n),\alpha)\right|\geq 0.5\delta \mid Z^n\right)\geq 0.5\delta\right),$$

which, by Lemma A.2(3), is in fact less than 0.5δ. We only need to verify

$$P_{Z^\infty}(\{P_n'(Z^n)>\delta\}\cap S_n)\leq 0.5\delta,\ \textit{eventually}.$$

It is easy to see that

$$\begin{aligned}&P_{Z^\infty}(\{P_n'(Z^n)>\delta\}\cap S_n)\\ \leq&P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\sup_{(\alpha,t_n)\in\mathcal{N}_{osn}\times\mathcal{T}_n} r_n\left|\widehat{L}_n^B(\alpha(t_n),\alpha)-t_n\left\{\mathbb{Z}_n^\omega+\langle u_n^*,\alpha-\alpha_0\rangle\right\}-\frac{B_n^\omega}{2}t_n^2\right|\geq 0.5\delta\mid Z^n\right)>0.5\delta\right).\end{aligned}$$

Hence, in order to prove the desired result, it suffices to show that

$$P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\sup_{(\alpha,t_n)\in\mathcal{N}_{osn}\times\mathcal{T}_n} r_n\left|\widehat{L}_n^B(\alpha(t_n),\alpha)-t_n\left\{\mathbb{Z}_n^\omega+\langle u_n^*,\alpha-\alpha_0\rangle\right\}-\frac{B_n^\omega}{2}t_n^2\right|\geq \delta\mid Z^n\right)>\delta\right)<\delta \tag{B.31}$$

eventually.

Step 2: For any $\alpha\in\mathcal{N}_{osn}$ and $t_n\in\mathcal{T}_n$, $\alpha(t_n)=\alpha+t_nu_n^*$, under Assumption A.7(i), we can apply the mean value theorem (wrt t_n) and obtain

$$\begin{aligned}\widehat{L}_n^B(\alpha(t_n),\alpha)=&\frac{t_n}{n}\sum_{i=1}^n\left(\frac{d\widetilde{m}(X_i,\alpha)}{d\alpha}[u_n^*]\right)'\widehat{\Sigma}(X_i)^{-1}\ell_n^B(X_i,\alpha)\\ &+\frac{t_n^2}{2n}\int_0^1\sum_{i=1}^n\left(\frac{d\widetilde{m}(X_i,\alpha(s))}{d\alpha}[u_n^*]\right)'\widehat{\Sigma}(X_i)^{-1}\left(\frac{d\widetilde{m}(x,\alpha(s))}{d\alpha}[u_n^*]\right)ds\\ &+\frac{t_n^2}{2n}\int_0^1\sum_{i=1}^n\left(\frac{d^2\widetilde{m}(X_i,\alpha(s))}{d\alpha^2}[u_n^*,u_n^*]\right)'\widehat{\Sigma}(X_i)^{-1}\ell_n^B(Z_i,\alpha(s))ds\\ \equiv&t_nT_{1n}^B(\alpha)+\frac{t_n^2}{2}\{T_{2n}(\alpha)+T_{3n}^B(\alpha)\},\end{aligned}$$

where $\alpha(s)\equiv\alpha+st_nu_n^*\in\mathcal{N}_{osn}$.

From these calculations and the fact that $P_{V^\infty|Z^\infty}(a_n+b_n\geq d|Z^n)\leq P_{V^\infty|Z^\infty}(a_n\geq 0.5d|Z^n)+P_{V^\infty|Z^\infty}(b_n\geq 0.5d|Z^n)$ a.s. for any two measurable random variables a_n *and* b_n, it follows that

$$\begin{aligned}&P_{V^\infty|Z^\infty}\left(\sup_{(\alpha,t_n)\in\mathcal{N}_{osn}\times\mathcal{T}_n} r_n\left|\widehat{L}_n^B(\alpha(t_n),\alpha)-t_n\left\{\mathbb{Z}_n^\omega+\langle u_n^*,\alpha-\alpha_0\rangle\right\}-\frac{B_n^\omega}{2}t_n^2\right|\geq 0.5\delta\mid Z^n\right)\\ \leq&P_{V^\infty|Z^\infty}\left(\sup_{(\alpha,t_n)\in\mathcal{N}_{osn}\times\mathcal{T}_n} r_nt_n\left|T_{1n}^B(\alpha)-\{\mathbb{Z}_n^\omega+\langle u_n^*,\alpha-\alpha_0\rangle\}\right|\geq 0.25\delta\mid Z^n\right)\\ &+P_{V^\infty|Z^\infty}\left(\sup_{(\alpha,t_n)\in\mathcal{N}_{osn}\times\mathcal{T}_n} r_n\frac{t_n^2}{2}\left|\{T_{2n}(\alpha)+T_{3n}^B(\alpha)\}-B_n^\omega\right|\geq 0.25\delta\mid Z^n\right).\end{aligned}$$

Hence, in order to show equation (B.31), it suffices to show that

$$P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\sup_{(\alpha,t_n)\in\mathcal{N}_{osn}\times\mathcal{T}_n} r_n t_n \left|T_{1n}^B(\alpha)-\{\mathbb{Z}_n^\omega+\langle u_n^*,\alpha-\alpha_0\rangle\}\right|\geq\delta \mid Z^n\right)\geq\delta\right)<\delta$$

and

$$P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\sup_{(\alpha,t_n)\in\mathcal{N}_{osn}\times\mathcal{T}_n} \frac{r_n t_n^2}{2}\left|\{T_{2n}(\alpha)+T_{3n}^B(\alpha)\}-B_n^\omega\right|\geq\delta \mid Z^n\right)\geq\delta\right)<\delta$$

eventually.

Since $r_n t_n \leq n^{1/2}$, by Lemma A.3, the first equation holds. Since $r_n t_n^2 \leq 1$, then in order to verify the second equation it suffices to verify that, for any $\delta>0$,

$$P_{Z^\infty}\left(\sup_{\alpha\in\mathcal{N}_{osn}}|T_{2n}(\alpha)-B_n^\omega|\geq\delta\right)<\delta,\ \forall n\geq N(\delta),$$

and

$$P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\sup_{\alpha\in\mathcal{N}_{osn}}\left|T_{3n}^B(\alpha)\right|\geq\delta \mid Z^n\right)\geq\delta\right)<\delta,\ \forall n\geq N(\delta).$$

By Lemmas A.5(1) and A.4, these two equations hold.

By our choice of $\ell_n^B()$ (in particular the fact that $\widetilde{m}$ is measurable with respect to Z^n), it follows that $B_n^\omega=B_n=O_{P_{V^\infty|Z^\infty}}(1)$ $wpa1(P_{Z^\infty})$. Thus we verified Assumption Boot.3(i).

Finaally, Lemma A.5(2) implies $\left|B_n^\omega-||u_n^*||^2\right|=o_{P_{V^\infty|Z^\infty}}(1)$ $wpa1(P_{Z^\infty})$ and $\left|B_n-||u_n^*||^2\right|=o_{P_{Z^\infty}}(1)$. *Q.E.D.*

The following lemma is a LLN for triangular arrays.

Lemma B.4. *Let $((X_{i,n})_{i=1}^n)_{n=1}^\infty$ be a triangular array of real valued random variables such that (a) $X_{1,n},...,X_{n,n}$ are independent and $X_{i,n}\sim P_{i,n}$, for all n, (b) $E[X_{i,n}]=0$ for all i and n, and (c) there is a sequence of non-negative real numbers $(b_n)_n$ such that $b_n=o(\sqrt{n})$ and*

$$\limsup_{n\to\infty} n^{-1}\sum_{i=1}^n E[|X_{i,n}|1\{|X_{i,n}|\geq b_n\}]=0.$$

Then: for all $\epsilon>0$, there is a $N(\epsilon)$ such that

$$\Pr\left(\left|n^{-1}\sum_{i=1}^n X_{i,n}\right|\geq\epsilon\right)<\epsilon$$

for all $n\geq N(\epsilon)$.

It is easy to see that a sufficient condition for condition (c) in Lemma B.4 is that: $E[|X_{i,n}|\varrho(|X_{i,n}|)]\leq\eta_n<\infty$ for all n, where $\varrho:\mathbb{R}_+\to\mathbb{R}_+$ is a strictly increasing bounded function, and $(\eta_n)_n$ a sequence of non-negative real numbers, such that $\varrho(b_n)=(4\eta_n/\epsilon^2)$ and $\frac{(\varrho^{-1}(\eta_n c))^2}{n}=o(1)$, for some constant $c>0$.

Proof of Lemma B.4: We obtain the result by modifying the proofs of Billingsley (1995) theorem 22.1 and of Feller (1970) (p. 248). For any $\epsilon>0$, let

$$X_{i,n}=X_{i,n}1\{|X_{i,n}|\leq b_n\}+X_{i,n}1\{|X_{i,n}|>b_n\}\equiv X_{i,n}^B+X_{i,n}^U.$$

Thus,

$$\Pr\left(\left|n^{-1}\sum_{i=1}^{n}X_{i,n}\right|\geq\epsilon\right)\leq\Pr\left(\left|n^{-1}\sum_{i=1}^{n}X_{i,n}^{B}\right|\geq0.5\epsilon\right)+\Pr\left(\left|n^{-1}\sum_{i=1}^{n}X_{i,n}^{U}\right|\geq0.5\epsilon\right)$$
$$\equiv T_{1,\epsilon}+T_{2,\epsilon}.$$

By conditions (b) and (c), it is easy to see that, for large enough n,

$$\begin{aligned}T_{1,\epsilon}&\leq\Pr\left(\left|n^{-1}\sum_{i=1}^{n}\{X_{i,n}^{B}-E[X_{i,n}^{B}]\}\right|\geq0.25\epsilon\right)+1\{E[X_{i,n}^{B}]\geq0.25\epsilon\}\\&=\Pr\left(\left|n^{-1}\sum_{i=1}^{n}\{X_{i,n}^{B}-E[X_{i,n}^{B}]\}\right|\geq0.25\epsilon\right)\leq2\exp\left(-const.\frac{\epsilon^{2}n}{b_{n}^{2}}\right),\end{aligned}$$

for some finite constant $const>0$, where the last inequality is due to Hoeffding inequality (cf. Van der Vaart and Wellner (1996) Appendix A.6). Thus, there is a $N(\epsilon)$ such that for all $n\geq N(\epsilon)$, $T_{1,\epsilon}<0.5\epsilon$.

For $T_{2,\epsilon}$, by Markov inequality and then by condition (c), we have:

$$\begin{aligned}T_{2,\epsilon}&\leq(\epsilon/2)^{-1}n^{-1}\sum_{i=1}^{n}\int_{\{|x|\geq b_{n}\}}|x|P_{i,n}(dx)\\&=(\epsilon/2)^{-1}n^{-1}\sum_{i=1}^{n}\int|x|1\{|x|\geq b_{n}\}P_{i,n}(dx)<0.5\epsilon\end{aligned}$$

eventually. *Q.E.D.*

Proof of Lemma 5.2: We divide the proof into several steps.

Step 1. We first show that

$$S_{n}\equiv\left\{Z^{n}:\left|n^{-1}\sum_{i=1}^{n}(g(X_{i},u_{n}^{*})\rho(Z_{i},\alpha_{0}))^{2}-E[g(X,u_{n}^{*})\Sigma_{0}(X)g(X,u_{n}^{*})']\right|\leq\delta\right\}$$

occurs wpa1($P_{Z^{\infty}}$). For this we apply Lemma B.4. Using the notation in the lemma, we let $X_{i,n}\equiv(g(X_{i},u_{n}^{*})\rho(Z_{i},\alpha_{0}))^{2}-E[g(X,u_{n}^{*})\Sigma_{0}(X)g(X,u_{n}^{*})']$, and thus conditions (a) and (b) of Lemma B.4 immediately follow (note that $E[g(X,u_{n}^{*})\Sigma_{0}(X)g(X,u_{n}^{*})']=1$). In order to check condition (c), note first that for any generic random variable X with mean $\mu<\infty$, it follows

$$E[|x-\mu|1\{|x-\mu|\geq b_{n}\}]\leq E[|x|1\{|x|\geq b_{n}-|\mu|\}]+|\mu|\Pr\{|x|\geq b_{n}-|\mu|\}.$$

Since b_{n} is taken to diverge, we can "redefine" b_{n} as $b_{n}-|\mu|$. Moreover,

$$\Pr\{|x|\geq b_{n}-|\mu|\}\leq E[\max\{|x|,1\}1\{|x|\geq b_{n}-|\mu|\}].$$

Again, since b_{n} is taken to diverge the only relevant case is $|x|\geq1$. Therefore, it suffices to study $E[|x|1\{|x|\geq b_{n}\}]$ in order to bound $E[|x-\mu|1\{|x-\mu|\geq b_{n}\}]$. Thus, applied to our case, it is sufficient to verify that

$$\limsup_{n\to\infty} n^{-1}\sum_{i=1}^{n} E\left[(g(X_i,u_n^*)\rho(Z_i,\alpha_0))^2 1\left\{(g(X_i,u_n^*)\rho(Z_i,\alpha_0))^2 \geq b_n\right\}\right] = 0.$$

This holds under our equation (5.1).

Step 2. Step 3 below shows that under assumption Boot.1,

$$\sup_{f\in BL_1}\left|E\left[f\left(\sqrt{n}\frac{\mathbb{Z}_n^{\omega-1}}{\sigma_\omega}\right)\mid Z^n\right] - E\left[f\left(\mathbb{Z}\right)\right]\right| = o_{P_{Z^\infty}}(1),$$

where $\mathbb{Z}\sim N(0,1)$. Also, Step 4 below shows that the same result holds under assumption Boot.2.

Step 3. Let $\sqrt{n}\frac{\mathbb{Z}_n^{\omega-1}}{\sigma_\omega} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\zeta_i\mathbf{s}_{i,n}$ where $\zeta_i = (\omega_i - 1)\sigma_\omega^{-1}$ and $\mathbf{s}_{i,n}\equiv g(X_i,u_n^*)\rho(Z_i,\alpha_0)$. We first want to show that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\zeta_i\mathbf{s}_{i,n}\Rightarrow\mathbb{Z},\ wpa1(P_{Z^\infty}).$$

Thus, it suffices to show that any sub-sequence, contains a further sub-sequence, $(n_k)_k$, such that (see Billingsley (1995) Theorem 20.5, p. 268)

$$\frac{1}{\sqrt{n_k}}\sum_{i=1}^{n_k}\zeta_i\mathbf{s}_{i,n_k}\Rightarrow\mathbb{Z},\ a.s.-(P_{Z^\infty}).$$

Since S_n occurs $wpa1(P_{Z^\infty})$ (Step 1), it follows that any sub-sequence, contains a further sub-sequence such that $n_k^{-1}\sum_{i=1}^{n_k}(\mathbf{s}_{i,n_k})^2\to 1$, $a.s.-(P_{Z^\infty})$. Moreover, $\max_{i\leq n_k}|\mathbf{s}_{i,n_k}|/\sqrt{n_k} = o(1)$, $a.s.-(P_{Z^\infty})$. This follows since, for any $\epsilon>0$,

$$\begin{aligned} P_{Z^\infty}\left(\max_{i\leq n}|\mathbf{s}_{i,n}|\geq\epsilon\sqrt{n}\right) &\leq \sum_{i=1}^{n}\int_{|s|\geq\epsilon\sqrt{n}}P_{i,n}(ds)\leq\epsilon^{-2}n^{-1}\sum_{i=1}^{n}\int_{|s|\geq\epsilon\sqrt{n}}s^2P_{i,n}(ds)\\ &=\epsilon^{-2}n^{-1}\sum_{i=1}^{n}E[\mathbf{s}_{i,n}^2 1\{|\mathbf{s}_{i,n}|\geq\epsilon\sqrt{n}\}]. \end{aligned}$$

We note that $1\{|\mathbf{s}_{i,n}|\geq\epsilon\sqrt{n}\}\leq 1\{|\mathbf{s}_{i,n}|^2\geq b_n\}$ (provided that $|\mathbf{s}_{i,n}|\geq 1$, but if it is not, then the proof is trivial). Hence by equation (5.1) and the fact that $\mathbf{s}_{i,n}$ are row-wise iid, the RHS is of order $o(1)$. Going to a sub-sequence establishes the result. Hence, for any $\epsilon>0$.

$$n_k^{-1}\sum_{i=1}^{n_k}(\mathbf{s}_{i,n_k})^2E_{P_\Omega}[\zeta_i^2 1\{\zeta_i|\mathbf{s}_{i,n_k}|>\epsilon\sqrt{n_k}\}]\to 0.$$

By Lindeberg-Feller CLT, it follows that $\frac{1}{\sqrt{n_k}}\sum_{i=1}^{n_k}\zeta_i\mathbf{s}_{i,n_k}\Rightarrow\mathbb{Z}$, $a.s.-(P_{Z^\infty})$ where $\mathbb{Z}\sim N(0,1)$.

We have thus showed that any sub-sequence, contains a further sub-sequence such that the above equation holds; therefore

$$\sup_{f\in BL_1}\left|E\left[f\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\zeta_i\mathbf{s}_{i,n}\right)\mid Z^n\right]-E\left[f\left(\mathbb{Z}\right)\right]\right|=o_{P_{Z^\infty}}(1).$$

Step 4. We proceed as in Step 3. The difference is that now $(\zeta_i)_i$ are not iid, but exchangeable. To overcome this, we follow lemma 3.6.15 in VdV-W. As before, we want to show that any sub-sequence, contains a further sub-sequence, $(n_k)_k$, such that

$$\frac{1}{\sqrt{n_k}}\sum_{i=1}^{n_k}\zeta_i\mathbf{s}_{i,n_k}\Rightarrow\mathbb{Z},\ a.s.-(P_{Z^\infty}).$$

To do this we follow lemma 3.6.15 (or rather their lemma A.5.3) in VdV-W for a given sub-sequence $(n_k)_k$. Let (using their notation) $n=n_k$, $a_{ni}\equiv\mathbf{s}_{i,n_k}$ and $W_{ni}\equiv\zeta_i=(\omega_{i,n}-1)$. By assumption Boot.2, $n^{-1}\sum_{i=1}^{n}W_{ni}=0$, $n^{-1}\sum_{i=1}^{n}W_{ni}^2\to 1$ and $n^{-1}\max_{1\leq i\leq n}W_{ni}^2=o_{P_\Omega}(1)$. And $n^{-1}\sum_{i=1}^{n}a_{ni}\equiv\overline{a}_n<\infty$ (in their lemma, VdV-W require $\overline{a}_n=0$, but is easy to see that it is not necessary if $n^{-1}\sum_{i=1}^{n}W_{ni}=0$, see Preastgaard (1991) lemma 5, p. 35), $n^{-1}\max_{1\leq i\leq n}a_{ni}^2=o(1)$ (this can be establish in the exact same way as it was done in step 3), and finally we need:

$$\limsup_{n\to\infty}n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}(W_{nj}a_{ni})^2 1\{|a_{ni}W_{nj}|>\epsilon\sqrt{n}\}=0,\ a.s.-P_{Z^\infty}.$$

To show this, we note that

$$P_\Omega\left(\max_{1\leq i\leq n}W_{ni}\geq L_n n^{0.5-c}\right)=o(1)$$

for any $c<0.5$ and $L_n=\log(n)$ or $\log(\log(n))$. This follows from the same calculations used to bound $n^{-1}\max_{1\leq i\leq n}a_{ni}^2$ and the fact that $E[|W_{ni}|^{\frac{2}{1-2c}}]\leq const.<\infty$ for n large enough. We can obtain a a.s. version of this result by going to a sub-sequence.

Since

$$n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}(W_{nj}a_{ni})^2 1\{|a_{ni}W_{nj}|>\epsilon\sqrt{n}\}\leq n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}(a_{ni})^2 1\{|a_{ni}|>\epsilon\frac{\sqrt{n}L_n}{|W_{ni}|n^c}n^c/L_n\},$$

then by the previous result $\frac{\sqrt{n}L_n}{|W_{ni}|n^c}\geq 1$ a.s.-$-P_{Z^\infty}$ and the desired result follows from equation (5.1) and choosing c such that $b_n\leq n^c/L_n$ (since $b_n=o(n^{1/2})$ we can always find such $c<0.5$).

So, by lemma 3.6.15 (or lemma A.5.3) in Van der Vaart and Wellner (1996),

$$\frac{1}{\sqrt{n_k}}\sum_{i=1}^{n_k}\zeta_i\mathbf{s}_{i,n_k}\Rightarrow\mathbb{Z},\ a.s.-(P_{Z^\infty}).$$

The rest of the steps are analogous to those in Step 3 and will not be repeated here. *Q.E.D.*

B.3.1 Alternative bootstrap sieve t statistics

In this subsection we present additional 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. The first one is $\widehat{W}_{1,n}^B \equiv \sqrt{n}\frac{\phi(\widehat{\alpha}_n^B)-\phi(\widehat{\alpha}_n)}{\sigma_\omega||\widehat{v}_n^*||_{n,sd}}$. In the definition of $\widehat{W}_{2,n}^B$ 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 equation (B.5).

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:

$$||\widehat{v}_n^B||_{B,sd}^2 \equiv \frac{1}{n}\sum_{i=1}^n \left(\frac{d\widehat{m}^B(X_i,\widehat{\alpha}_n^B)}{d\alpha}[\widehat{v}_n^B]\right)' \hat{\Sigma}_i^{-1}\varrho(V_i,\widehat{\alpha}_n^B)\varrho(V_i,\widehat{\alpha}_n^B)'\hat{\Sigma}_i^{-1}\left(\frac{d\widehat{m}^B(X_i,\widehat{\alpha}_n^B)}{d\alpha}[\widehat{v}_n^B]\right) \quad \text{(B.32)}$$

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 equation (4.7). 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 equation (B.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.

The 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

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

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 (4.9) 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.

Recall that $\hat{M}_i^B = (\omega_{i,n}-1)^2\hat{M}_i$ and $\hat{M}_i = \hat{\Sigma}_i^{-1}\rho(Z_i,\widehat{\alpha}_n)\rho(Z_i,\widehat{\alpha}_n)'\hat{\Sigma}_i^{-1}$.

Assumption B.3. *(i)* $\sup_{v_1,v_2\in\overline{\mathbf{V}}^1_{k(n)}} |\langle v_1,v_2\rangle_{B,\Sigma^{-1}} - \langle v_1,v_2\rangle_{n,\Sigma^{-1}}| = o_{P_{V^\infty|Z^\infty}}(1)$ *wpa*$1(P_{Z^\infty})$;
(ii) $\sup_{v\in\overline{\mathbf{V}}^1_{k(n)}} |\langle v,v\rangle_{B,\hat{M}^B} - \sigma^2_\omega\langle v,v\rangle_{n,\hat{M}}| = o_{P_{V^\infty|Z^\infty}}(1)$ *wpa*$1(P_{Z^\infty})$;
(iii) $\sup_{v\in\overline{\mathbf{V}}^1_{k(n)}} n^{-1}\sum_{i=1}^n(\omega_{i,n}-1)^2\left\|\frac{d\widehat{m}^B(X_i,\hat{\alpha}^B_n)}{d\alpha}[v]\right\|_e^2 = O_{P_{V^\infty|Z^\infty}}(1)$ *wpa*$1(P_{Z^\infty})$.

Assumption B.3(i)(ii) is analogous to Assumption 4.1(ii)(v). Assumption B.3(iii) is a mild one, for example, it is implied by Assumptions for Lemma A.1 and uniformly bounded bootstrap weights (i.e., $|\omega_{i,n}|\leq C<\infty$ for all i).

The following result is a bootstrap version of Theorem 4.2.

Theorem B.3. *Let Conditions for Theorem 4.2(1) and Lemma A.1, Assumption B.3 hold. Then:*

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

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

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

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

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

Proof of Theorem B.3. For **Result (1)**, the proof is analogous to the one for Theorem 4.2(1). As in the proof of Theorem 4.2(1), it suffices to show that

$$\frac{||\hat{v}^B_n - v^*_n||}{||v^*_n||} = o_{P_{V^\infty|Z^\infty}}(1)\ wpa1(P_{Z^\infty}), \tag{B.33}$$

and

$$\left|\frac{||\hat{v}^B_n||_{B,sd} - ||\hat{v}^B_n||_{sd}}{||v^*_n||}\right| = o_{P_{V^\infty|Z^\infty}}(1)\ wpa1(P_{Z^\infty}). \tag{B.34}$$

Following the same derivations as in the proof of theorem 4.2(1) step 1, for equation (B.33), it suffices to show

$$|\langle\hat{\varpi}^B_n,\varpi\rangle_{B,\hat{\Sigma}^{-1}} - \langle\hat{\varpi}^B_n,\varpi\rangle_{B,\Sigma^{-1}}| = o_{P_{V^\infty|Z^\infty}}(1)\ and\ |\langle\hat{\varpi}^B_n,\varpi\rangle_{B,\Sigma^{-1}} - \langle\hat{\varpi}^B_n,\varpi\rangle_{\Sigma^{-1}}| = o_{P_{V^\infty|Z^\infty}}(1)$$

wpa$1(P_{Z^\infty})$, uniformly over $\varpi\in\overline{\mathbf{V}}^1_{k(n)}$; where $\hat{\varpi}^B_n = \frac{\hat{v}^B_n}{||\hat{v}^B_n||}$. The first term follows by Assumptions 4.1(iii) and 3.1(iv) and the fact that $\langle\varpi,\varpi\rangle_{B,\Sigma^{-1}} = O_{P_{V^\infty|Z^\infty}}(1)$ *wpa*$1(P_{Z^\infty})$ (by Assumptions B.3(i) and 4.1(ii)). The second term follows directly from these two assumptions.

Regarding equation (B.34), following the same derivations as in the proof of Theorem 4.2 step 2, it suffices to show that $\left| ||\hat{\varpi}_n^B||^2_{B,sd} - ||\hat{\varpi}_n^B||^2_{sd} \right| = o_{P_{V^\infty|Z^\infty}}(1)\ wpa1(P_{Z^\infty})$. By the triangle inequality,

$$\begin{aligned}
\sup_{v\in\overline{\mathbf{V}}^1_{k(n)}} |\langle v,v\rangle_{B,\hat{W}^B} - \sigma^2_\omega \langle v,v\rangle_{n,\hat{M}}| \leq & \sup_{v\in\overline{\mathbf{V}}^1_{k(n)}} \left|\langle v,v\rangle_{B,\hat{W}^B} - \langle v,v\rangle_{B,\hat{M}^B}\right| + \sup_{v\in\overline{\mathbf{V}}^1_{k(n)}} \left|\langle v,v\rangle_{B,\hat{M}^B} - \sigma^2_\omega\langle v,v\rangle_{n,\hat{M}}\right| \\
\equiv & A^B_{1n} + A^B_{2n}
\end{aligned}$$

with $\hat{W}^B_i \equiv \hat{\Sigma}^{-1}_i \varrho(V_i,\hat{\alpha}^B_n)\varrho(V_i,\hat{\alpha}^B_n)'\hat{\Sigma}^{-1}_i = (\omega_{i,n}-1)^2\hat{\Sigma}^{-1}_i\rho(Z_i,\hat{\alpha}^B_n)\rho(Z_i,\hat{\alpha}^B_n)'\hat{\Sigma}^{-1}_i$ and $\hat{M}^B_i = (\omega_{i,n}-1)^2\hat{M}_i$ and $\hat{M}_i = \hat{\Sigma}^{-1}_i\rho(Z_i,\widehat{\alpha}_n)\rho(Z_i,\widehat{\alpha}_n)'\hat{\Sigma}^{-1}_i$.

It is easy to see that A^B_{1n} is bounded above by

$$\begin{aligned}
& \sup_x ||\hat{\Sigma}^{-1}(x)\{\rho(z,\hat{\alpha}^B_n)\rho(z,\hat{\alpha}^B_n)' - \rho(z,\hat{\alpha}_n)\rho(z,\hat{\alpha}_n)'\}\hat{\Sigma}^{-1}(x)||_e n^{-1}\sum_{i=1}^n(\omega_{i,n}-1)^2\left\|\hat{T}^B_i[v]\right\|^2_e \\
\leq & 2\sup_x \sup_{\alpha\in\mathcal{N}_{osn}} ||\hat{\Sigma}^{-1}(x)\{\rho(z,\alpha)\rho(z,\alpha)' - \rho(z,\alpha_0)\rho(z,\alpha_0)'\}\hat{\Sigma}^{-1}(x)||_e n^{-1}\sum_{i=1}^n(\omega_{i,n}-1)^2\left\|\hat{T}^B_i[v]\right\|^2_e
\end{aligned}$$

where $\hat{T}^B_i[v] \equiv \frac{d\widehat{m}^B(X_i,\hat{\alpha}^B_n)}{d\alpha}[v]$. The second line follows because $\hat{\alpha}^B \in \mathcal{N}_{osn}$ wpa1. The first term in the RHS is of order $o_{P_{Z^\infty}}(1)$ by Assumption 4.1(iv). The second term is $O_{P_{V^\infty|Z^\infty}}(1)$ by Assumption B.3(iii).

A^B_{2n} is of order $o_{P_{V^\infty|Z^\infty}}(1)\ wpa1(P_{Z^\infty})$ by Assumption B.3(ii).

Result (1) now follows from the same derivations as in the proof of Theorem 4.2(1) step 2a.

Given Result (1), **Result (2)** follows from exactly the same proof as that of Theorem 5.2(1), and is omitted. *Q.E.D.*

B.4 Proofs for Section 6 on examples

Proof of Proposition 6.1. By our assumption over $clsp\{p_j : j=1,...,J\}$, $\frac{dm(x,\alpha_0)}{d\alpha}[u^*_n] \in clsp\{p_j : j=1,...,J_n\}$ provided $k(n) \leq J_n$, and thus Assumption A.6 (i) trivially holds. Since $\Sigma = 1$, Assumption A.6 (ii) is the same as Assumption A.6 (i).

We now show that Assumption A.6(iii)(iv) holds under condition 6.1. First, condition 6.1(i) implies that $\{(E[h(Y_2)-h_0(Y_2)|\cdot])^2 : h\in\mathcal{H}\}$ is a P-Donsker class and, moreover,

$$E[(E[h(Y_2)-h_0(Y_2)|X])^4] \leq 2c\times||h-h_0||^2 \to 0$$

as $||h-h_0||_{L^2(f_{Y_2})}\to 0$. So by Lemma 1 in Chen et al. (2003), Assumption A.6(iii) holds. Regarding Assumption A.6(iv). By Theorem 2.14.2 in VdV-W, (up to omitted constants)

$$E\left[\left|\sup_{f\in\mathcal{F}_n} n^{-1/2}\sum_{i=1}^n\{f(X_i)-E[f(X_i)]\}\right|\right] \leq \int_0^{||F_n||_{L^2(f_X)}}\sqrt{1+\log N_{[]}(u,\mathcal{F}_n,||\cdot||_{L^2(f_X)})}du$$

where $\mathcal{F}_n \equiv \{f : f = g(\cdot,u^*_n)(m(\cdot,\alpha)-m(\cdot,\alpha_0)),\ some\ \alpha\in\mathcal{N}_{osn}\}$ and

$$F_n(x) \equiv \sup_{\mathcal{F}_n}|f(x)| = \sup_{\alpha\in\mathcal{N}_{osn}}|g(x,u^*_n)\{m(x,\alpha)-m(x,\alpha_0)\}|.$$

We claim that, under our assumptions,

$$N_{[]}(u,\mathcal{F}_n,||\cdot||_{L^2(f_X)}) \leq N_{[]}(u,\Lambda^\gamma_c(\mathcal{X}),||\cdot||_{L^\infty}).$$

To show this claim, it suffices to show that given a radius $\delta > 0$, if we take $\{[l_j, u_j]\}_{j=1}^{N(\delta)}$ to be brackets of $\Lambda_c^\gamma(\mathcal{X})$ under $||\cdot||_{L^\infty}$, then we can construct $\{[l_{n,j}, u_{n,j}]\}_{j=1}^{N(\delta)}$ such that: they are valid brackets of $\mathcal{F}_n$, under $||\cdot||_{L^2(f_X)}$. To show this, observe that, for any $f_n \in \mathcal{F}_n$, there exists a $\alpha \in \mathcal{N}_{osn}$, such that $f_n = g(\cdot, u_n^*)\{m(\cdot,\alpha) - m(\cdot,\alpha_0)\}$, and under condition 6.1, it follows that there exists a $j \in \{1, ..., N(\delta)\}$ such that

$$l_j \leq m(\cdot,\alpha) - m(\cdot,\alpha_0) \leq u_j, \tag{B.35}$$

hence, there exists a $[l_{n,j}, u_{n,j}]$ such that, for all x,

$$l_{n,j}(x) = (1\{g(x,u_n^*) > 0\}l_j(x) + 1\{g(x,u_n^*) < 0\}u_j(x))g(x,u_n^*),$$

and

$$u_{n,j}(x) = (1\{g(x,u_n^*) > 0\}u_j(x) + 1\{g(x,u_n^*) < 0\}l_j(x))g(x,u_n^*).$$

such that $l_{n,j} \leq f_n \leq u_{n,j}$. Also, observe that

$$||l_{n,j} - u_{n,j}||_{L^2(f_X)} = \sqrt{E[(g(X,u_n^*))^2(u_j(X) - l_j(X))^2]} \leq ||u_j - l_j||_{L^\infty} \leq \delta$$

because $E[(g(X,u_n^*))^2] = ||u_n^*||^2 = 1$ and $||u_j - l_j||_{L^\infty} \leq \delta$ by construction.

Therefore,

$$E\left[\left|\sup_{f\in\mathcal{F}_n} n^{-1/2}\sum_{i=1}^n \{f(X_i) - E[f(X_i)]\}\right|\right] \leq \int_0^{||F_n||_{L^2(f_X)}} \sqrt{1 + \log N_{[]}(u, \Lambda_c^\gamma(\mathcal{X}), ||\cdot||_{L^\infty})}du.$$

Since by assumption $\gamma > 0.5$, it is well-known that $\sqrt{1 + \log N_{[]}(u, \Lambda_c^\gamma(\mathcal{X}), ||\cdot||_{L^\infty})}$ is integrable, so in order to show that $E\left[\left|\sup_{f\in\mathcal{F}_n} n^{-1/2}\sum_{i=1}^n\{f(X_i) - E[f(X_i)]\}\right|\right] = o(1)$, it suffices to show that $||F_n||_{L^2(f_X)} = o(1)$. In order to show this,

$$\begin{aligned}
||F_n||_{L^2(f_X)} \leq & \sqrt{E[(g(X,u_n^*))^2(\sup_{\mathcal{N}_{osn}} |m(X,\alpha) - m(X,\alpha_0)|)^2]} \\
= & \sqrt{E[(g(X,u_n^*))^2(\sup_{\mathcal{N}_{osn}} |E[h(Y_2) - h_0(Y_2)|X]|)^2]} \\
= & \sqrt{E[(g(X,u_n^*))^2 \sup_{\mathcal{N}_{osn}} \int (h(y_2) - h_0(y_2))^2 f_{Y_2|X}(y_2, X)dy_2]} \\
= & \sqrt{E[(g(X,u_n^*))^2 \sup_{\mathcal{N}_{osn}} \int (h(y_2) - h_0(y_2))^2 \frac{f_{Y_2X}(y_2,X)}{f_{Y_2}(y_2)f_X(X)} f_{Y_2}(y_2)dy_2]} \\
\leq & \sup_{x,y_2} \frac{f_{Y_2X}(y_2,x)}{f_{Y_2}(y_2)f_X(x)} \sup_{\mathcal{N}_{osn}} ||h - h_0||_{L^2(f_{Y_2})} \sqrt{E[(g(X,u_n^*))^2]} \\
\leq & Const. \times M_n\delta_{s,n} \to 0
\end{aligned}$$

where the last expression follows from the fact that $E[(g(X,u_n^*))^2] = ||u_n^*||^2 = 1$ and condition 6.1(ii), that states that

$$\sup_{x,y_2} \frac{f_{Y_2X}(y_2,x)}{f_{Y_2}(y_2)f_X(x)} \leq Const. < \infty.$$

Hence, $E\left[\left|\sup_{f\in\mathcal{F}_n} n^{-1/2}\sum_{i=1}^n\{f(X_i)-E[f(X_i)]\}\right|\right]=o(1)$ which implies assumption A.6(iv). Finally, Assumption A.7 is automatically satisfied with the NPIV model. *Q.E.D.*

Proof of Proposition 6.2. Assumptions A.6(i) and (ii) hold by the same calculations as those in the proof of Proposition 6.1 (for the NPIV model). Also, under Condition 6.2(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 $\gamma>0.5$, Assumptions A.6(iii) and (iv) hold by similar calculations to those in the proof of Proposition 6.1.

Assumption A.7(i) is standard in the literature. Regarding Assumption A.7(ii), observe that for any $h\in\mathcal{N}_{osn}$,

$$\begin{aligned}
&\left|\frac{dm(x,h)}{dh}[u_n^*]-\frac{dm(x,h_0)}{dh}[u_n^*]\right|\\
=&\left|E\left[\{f_{Y_1|Y_2X}(h(Y_2),Y_2,x)-f_{Y_1|Y_2X}(h_0(Y_2),Y_2,x)\}u_n^*(Y_2)\mid X=x\right]\right|\\
=&\left|\int\left\{\int_0^1\frac{df_{Y_1|Y_2X}(h_0(t)(y_2),y_2,x)}{dy_1}(h(y_2)-h_0(y_2))u_n^*(y_2)dt\right\}f_{Y_2|X}(y_2,x)dy_2\right|\\
=&\left|\int\left(\int_0^1\frac{df_{Y_1|Y_2X}(h_0(t)(y_2),y_2,x)}{dy_1}dt\right)(h(y_2)-h_0(y_2))u_n^*(y_2)f_{Y_2}(y_2)\left(\frac{f_{Y_2X}(y_2,x)}{f_{Y_2}(y_2)f_X(x)}\right)dy_2\right|\\
=&\left|\int\Gamma_1(y_2,x)\Gamma_2(y_2,x)(h(y_2)-h_0(y_2))u_n^*(y_2)f_{Y_2}(y_2)dy_2\right|\\
\leq&||\Gamma_1(\cdot,x)\Gamma_2(\cdot,x)||_{L^\infty}\times||h-h_0||_{L^2(f_{Y_2})}||u_n^*||_{L^2(f_{Y_2})}
\end{aligned}$$

where $h_0(t)\equiv h_0+t\{h-h_0\}$ and $\Gamma_1(y_2,x)\equiv\left(\int_0^1\frac{df_{Y_1|Y_2X}(h_0(t)(y_2),y_2,x)}{dy_1}dt\right)$ and $\Gamma_2(y_2,x)\equiv\frac{f_{Y_2X}(y_2,x)}{f_{Y_2}(y_2)f_X(x)}$; the last line follows from Cauchy-Swarchz inequality.

Under Condition 6.2(ii), it follows that

$$\sup_{y_1,y_2,x}\left|\frac{df_{Y_1|Y_2X}(y_1,y_2,x)}{dy_1}\right|\leq C<\infty$$

and, under Condition 6.1(ii), it follows that

$$\sup_{x,y_2}\left|\frac{f_{Y_2X}(y_2,x)}{f_{Y_2}(y_2)f_X(x)}\right|\leq C<\infty.$$

Then it is easy to see that $||\Gamma_j(\cdot,x)||_{L^\infty(f_{Y_2})}\leq C<\infty$ for both $j=1,2$. Thus

$$\left|\frac{dm(x,h)}{dh}[u_n^*]-\frac{dm(x,h_0)}{dh}[u_n^*]\right|\leq C^2\times||h-h_0||_{L^2(f_{Y_2})}||u_n^*||_{L^2(f_{Y_2})}$$

and thus, Assumption A.7(ii) is satisfied provided that $n\times M_n^2\delta_n^2\sup_{h\in\mathcal{N}_{osn}}||h-h_0||^2_{L^2(f_{Y_2})}||u_n^*||^2_{L^2(f_{Y_2})}=o(1)$. Since $||u_n^*||_{L^2(f_{Y_2})}\leq c\mu_{k(n)}^{-1}$ it suffices to show that

$$nM_n^4\delta_n^2(||\Pi_nh_0-h_0||_{L^2(f_{Y_2})}+\mu_{k(n)}^{-1}\delta_n)^2\mu_{k(n)}^{-2}=o(1).$$

By assumption, $||\Pi_nh_0-h_0||_{L^2(f_{Y_2})}\leq Const.\times\mu_{k(n)}^{-1}\delta_n=O(\delta_{s,n})$ and $\delta_n^2\asymp Const.k(n)/n$, then it suffices to show that

$$nM_n^4\delta_{s,n}^4=o(1),$$

which holds by Condition 6.3.

Regarding Assumption A.7(iii), observe that for any $h \in \mathcal{N}_{osn}$,

$$\frac{d^2 m(x,h)}{dh^2}[u_n^*, u_n^*] = \int \frac{df_{Y_1|Y_2X}(h(y_2), y_2, x)}{dy_1}(u_n^*(y_2))^2 f_{Y_2|X}(y_2, x)dy_2.$$

Again by Conditions 6.2(ii) and 6.1(ii), it follows that $\left|\frac{d^2m(x,h)}{dh^2}[u_n^*, u_n^*]\right| \leq C^2 \times ||u_n^*||^2_{L^2(f_{Y_2})}$. Since $||u_n^*||_{L^2(f_{Y_2})} \leq const \times \mu_{k(n)}^{-1}$, Assumption A.7(iii) holds because

$$\mu_{k(n)}^{-2} \times (M_n \delta_n)^2 = o(1), \ or \ M_n^2 \delta_{s,n}^2 = o(1).$$

Finally, we verify Assumption A.7(iv). By our previous calculations

$$\begin{aligned}
&\left|\frac{dm(x,h_1)}{dh}[h_2 - h_0] - \frac{dm(x,h_0)}{dh}[h_2 - h_0]\right| \\
=&\left|\int \left(\int \frac{df_{Y_1|Y_2X}(h_0(y_2) + t[h_1(y_2) - h_0(y_2)], y_2, x)}{dy_1}dt\right)(h_1(y_2) - h_0(y_2))(h_2(y_2) - h_0(y_2))f_{Y_2|X}(y_2, x)dy_2\right| \\
\leq& C^2 \times \int |(h_1(y_2) - h_0(y_2))(h_2(y_2) - h_0(y_2))| f_{Y_2}(y_2)dy_2 \\
\leq& C^2 \times ||h_1 - h_0||_{L^2(f_{Y_2})}||h_2 - h_0||_{L^2(f_{Y_2})},
\end{aligned}$$

where the first inequality follows from Conditions 6.2(ii) and 6.1(ii), and the last one from Cauchy-Swarchz inequality. This result and Cauchy-Swarchz inequality together imply that

$$\begin{aligned}
&\left|E\left[g(X, u_n^*)\left(\frac{dm(X,h_1)}{dh}[h_2 - h_0] - \frac{dm(X,h_0)}{dh}[h_2 - h_0]\right)\right]\right| \\
\leq& C^2\sqrt{E[(g(X, u_n^*))^2]}||h_1 - h_0||_{L^2(f_{Y_2})}||h_2 - h_0||_{L^2(f_{Y_2})} \\
\leq& const \times ||h_1 - h_0||_{L^2(f_{Y_2})}||h_2 - h_0||_{L^2(f_{Y_2})},
\end{aligned}$$

where the last line follows from $E[(g(X, u_n^*))^2] = ||u^*||^2 \asymp 1$. Thus, Assumption A.7(iv) follows if

$$\delta_{s,n}^2 = (||\Pi_n h_0 - h_0||_{L^2(f_{Y_2})} + \mu_{k(n)}^{-1}\delta_n)^2 = o(n^{-1/2})$$

which holds by Condition 6.3. *Q.E.D.*

C Supplement: Proofs of the Results in Appendix A

In Appendix C, we provide the proofs of all the lemmas, theorems and propositions stated in Appendix A.

C.1 Proofs for Section A.2 on convergence rates of bootstrap PSMD estimators

Proof of Lemma A.1: For **Result (1)**, we prove this result in two steps. First, we show that $\widehat{\alpha}_n^B \in \mathcal{A}_{k(n)}^{M_0}$ wpa1-$P_{V^\infty|Z^\infty}$ for any Z^∞ in a set that occurs probability approaching P_{Z^∞} one, where $\mathcal{A}_{k(n)}^{M_0}$ is defined in the text. Second, we establish consistency, using the fact that we are in the $\mathcal{A}_{k(n)}^{M_0}$ set.

Step 1. We show that for any $\delta > 0$, there exists a $N(\delta)$ such that

$$P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\hat{\alpha}_n^B \notin \mathcal{A}_{k(n)}^{M_0}|Z^n\right) < \delta\right) \geq 1-\delta,\ \forall n \geq N(\delta).$$

To show this, note that, by definition of $\widehat{\alpha}_n^B$,

$$\lambda_n Pen(\widehat{h}_n^B) \leq \widehat{Q}_n^B(\widehat{\alpha}_n) + \lambda_n Pen(\widehat{h}_n) + o_{P_{V^\infty|Z^\infty}}(\frac{1}{n}),\ wpa1(P_{Z^\infty}).$$

By Assumption A.1(i) and the definition of $\widehat{\alpha}_n \in \mathcal{A}_{k(n)}$,

$$\begin{aligned}
\lambda_n Pen(\widehat{h}_n^B) &\leq c_0^*\left(\widehat{Q}_n(\widehat{\alpha}_n) + \lambda_n Pen(\widehat{h}_n)\right) + o_{P_{V^\infty|Z^\infty}}(\frac{1}{n}),\ wpa1(P_{Z^\infty}) \\
&\leq c_0^*\left(\widehat{Q}_n(\Pi_n\alpha_0) + \lambda_n Pen(\Pi_n h_0)\right) + o_{P_{V^\infty|Z^\infty}}(\frac{1}{n}),\ wpa1(P_{Z^\infty}).
\end{aligned}$$

By Assumptions 3.2(i)(ii) and 3.3(i),

$$\lambda_n Pen(\widehat{h}_n^B) \leq c_0^* c_0 Q(\Pi_n\alpha_0) + \lambda_n Pen(h_0) + O_{P_{V^\infty|Z^\infty}}(\lambda_n + o(\frac{1}{n})),\ wpa1(P_{Z^\infty}).$$

By the fact that $Q(\Pi_n\alpha_0) + o(\frac{1}{n}) = O(\lambda_n)$, the desired result follows.

Step 2. We want to show that for any $\delta > 0$, there exists a $N(\delta)$ such that

$$P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(||\hat{\alpha}_n^B - \alpha_0||_s \geq \delta|Z^n\right) < \delta\right) \geq 1-\delta,\ \forall n \geq N(\delta),$$

which is equivalent to show that $P_{Z^\infty}(P_{V^\infty|Z^\infty}\left(||\hat{\alpha}_n^B - \alpha_0||_s \geq \delta|Z^n\right) > \delta) \leq \delta$ eventually. Note that

$$\begin{aligned}
&P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(||\hat{\alpha}_n^B - \alpha_0||_s \geq \delta|Z^n\right) > \delta\right) \\
\leq &P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\{||\hat{\alpha}_n^B - \alpha_0||_s \geq \delta\} \cap \{\hat{\alpha}_n^B \in \mathcal{A}_{k(n)}^{M_0}\}|Z^n\right) > 0.5\delta\right) \\
&+ P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\hat{\alpha}_n^B \notin \mathcal{A}_{k(n)}^{M_0}|Z^n\right) > 0.5\delta\right).
\end{aligned}$$

By step 1, the second summand in the RHS is negligible. Thus, it suffices to show that

$$P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\hat{\alpha}_n^B \in \mathcal{A}_{k(n)}^{M_0} \colon ||\hat{\alpha}_n^B - \alpha_0||_s \geq \delta|Z^n\right) < \delta\right) \geq 1-\delta,\ \forall n \geq N(\delta).$$

(henceforth, we omit $\hat{\alpha}_n^B \in \mathcal{A}_{k(n)}^{M_0}$). Note that, conditioning on Z^n, by Assumption A.1(i), the definition of $\widehat{\alpha}_n \in \mathcal{A}_{k(n)}^{M_0}$, Assumption 3.2(i)(ii) and $\max\{\lambda_n, o(\frac{1}{n})\} = O(\lambda_n)$, we have:

$$
\begin{aligned}
& P_{V^\infty|Z^\infty}\left(||\hat{\alpha}_n^B - \alpha_0||_s \geq \delta | Z^n\right) \\
& \leq P_{V^\infty|Z^\infty}\left(\inf_{\{\mathcal{A}_{k(n)}^{M_0}:\,||\alpha-\alpha_0||_s\geq\delta\}} \left\{\widehat{Q}_n^B(\alpha) + \lambda_n Pen(h)\right\} \leq \widehat{Q}_n^B(\widehat{\alpha}) + \lambda_n Pen(\widehat{h}) + o(\frac{1}{n}) | Z^n\right) \\
& \leq P_{V^\infty|Z^\infty}\left(\inf_{\{\mathcal{A}_{k(n)}^{M_0}:\,||\alpha-\alpha_0||_s\geq\delta\}} \left\{c^*\widehat{Q}_n(\alpha) + \lambda_n Pen(h)\right\} \leq c_0^*\left[\widehat{Q}_n(\widehat{\alpha}) + \lambda_n Pen(\widehat{h})\right] + O(\lambda_n) + (\overline{\delta}_{m,n}^*)^2 | Z^n\right) \\
& \leq P_{V^\infty|Z^\infty}\left(\inf_{\{\mathcal{A}_{k(n)}^{M_0}:\,||\alpha-\alpha_0||_s\geq\delta\}} \left\{c^*\widehat{Q}_n(\alpha)\right\} \leq c_0^*\left[\widehat{Q}_n(\Pi_n\alpha_0) + \lambda_n Pen(\Pi_n h_0)\right] + O(\lambda_n) + (\overline{\delta}_{m,n}^*)^2 | Z^n\right).
\end{aligned}
$$

Thus

$$
\begin{aligned}
& P_{V^\infty|Z^\infty}\left(||\hat{\alpha}_n^B - \alpha_0||_s \geq \delta | Z^n\right) \\
\leq \; & P_{V^\infty|Z^\infty}\left(\inf_{\{\mathcal{A}_{k(n)}^{M_0}:\,||\alpha-\alpha_0||_s\geq\delta\}} c^*\widehat{Q}_n(\alpha) \leq c_0^*\widehat{Q}_n(\Pi_n\alpha_0) + M(\lambda_n + (\overline{\delta}_{m,n}^*)^2) | Z^n\right) \\
& + P_{V^\infty|Z^\infty}\left(\sup_{\{\mathcal{A}_{k(n)}^{M_0}:\,||\alpha-\alpha_0||_s\geq\delta\}} \widehat{Q}_n^B(\alpha) - c^*\widehat{Q}_n(\alpha) < -M(\overline{\delta}_{m,n}^*)^2 | Z^n\right) \\
& + P_{V^\infty|Z^\infty}\left(\widehat{Q}_n^B(\widehat{\alpha}) - c_0^*\widehat{Q}_n(\widehat{\alpha}) > -o(\frac{1}{n}) | Z^n\right),
\end{aligned}
$$

where the second and third terms in the RHS are negligible (wpa1(P_{Z^∞})) by Assumption A.1(i)(ii). Regarding the first term, by similar algebra, it can be bounded above by

$$
\begin{aligned}
& P_{V^\infty|Z^\infty}\left(\inf_{\{\mathcal{A}_{k(n)}^{M_0}:\,||\alpha-\alpha_0||_s\geq\delta\}} c^* c Q(\alpha) \leq c_0^* c_0 Q(\Pi_n\alpha_0) + M(\lambda_n + (\overline{\delta}_{m,n} + \overline{\delta}_{m,n}^*)^2) | Z^n\right) \\
& + P_{V^\infty|Z^\infty}\left(\sup_{\{\mathcal{A}_{k(n)}^{M_0}:\,||\alpha-\alpha_0||_s\geq\delta\}} \widehat{Q}_n(\alpha) - cQ(\alpha) < -M(\overline{\delta}_{m,n})^2 | Z^n\right) \\
& + P_{V^\infty|Z^\infty}\left(\widehat{Q}_n(\Pi_n\alpha_0) - c_0 Q(\Pi_n\alpha_0) > -o(\frac{1}{n}) | Z^n\right).
\end{aligned}
$$

Therefore, for sufficiently large n,

$$
\begin{aligned}
&P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(||\hat{\alpha}_n^B-\alpha_0||_s\geq\delta|Z^n\right)<\delta\right)\leq 0.25\delta\\
&+P_{Z^\infty}\left(\inf_{\{\mathcal{A}_{k(n)}^{M_0}:\,||\alpha-\alpha_0||_s\geq\delta\}} c^*cQ(\alpha)\leq c_0^*c_0Q(\Pi_n\alpha_0)+M(\lambda_n+(\overline{\delta}_{m,n}+\overline{\delta}^*_{m,n})^2)\right)\\
&+P_{Z^\infty}\left(\sup_{\{\mathcal{A}_{k(n)}^{M_0}:\,||\alpha-\alpha_0||_s\geq\delta\}}\widehat{Q}_n(\alpha)-cQ(\alpha)<-M(\overline{\delta}_{m,n})^2\right)+P_{Z^\infty}\left(\widehat{Q}_n(\Pi_n\alpha_0)-c_0Q(\Pi_n\alpha_0)>-o(\frac{1}{n})\right).
\end{aligned}
$$

By Assumption 3.3, the third and fourth terms in the RHS are less than 0.5δ. The second term in the RHS is not random. By Assumptions 3.1(ii) and 3.2(iii), $\mathcal{A}_{k(n)}^{M_0}$ is compact, and so is $\mathcal{A}^{M_0}\equiv\{\alpha=(\theta',h)\in\mathcal{A}:\lambda_nPen(h)\leq\lambda_nM_0\}$. This fact, and Assumption 3.1(iii) imply that $\inf_{\{\mathcal{A}_{k(n)}^{M_0}:\,||\alpha-\alpha_0||_s\geq\delta\}}c^*cQ(\alpha)\geq Q(\alpha(\delta))$ some $\alpha(\delta)\in\mathcal{A}^{M_0}\cap\{||\alpha-\alpha_0||_s\geq\delta\}$. By Assumption 3.1(i), $Q(\alpha(\delta))>0$, so eventually, since $c_0^*c_0Q(\Pi_n\alpha_0)+M(\lambda_n+(\overline{\delta}_{m,n}+\overline{\delta}^*_{m,n})^2)=o(1)$, $P_{Z^\infty}\left(\inf_{\{\mathcal{A}_{k(n)}^{M_0}:\,||\alpha-\alpha_0||_s\geq\delta\}}c^*cQ(\alpha)\leq c_0^*c_0Q(\Pi_n\alpha_0)+M(\lambda_n+(\overline{\delta}_{m,n}+\overline{\delta}^*_{m,n})^2)\right)=0$.

For **Result (2)**, we want to show that for any $\delta>0$, there exists a $M(\delta)$ such that

$$
P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\delta_n^{-1}||\hat{\alpha}_n^B-\alpha_0||\geq M'\ |Z^n\right)<\delta\right)\geq 1-\delta,\ \forall M'\geq M(\delta)
$$

eventually. By Assumptions 3.4(iii) and A.1(iii), following the similar algebra as before, we have: for M' large enough,

$$
\begin{aligned}
&P_{V^\infty|Z^\infty}\left(\delta_n^{-1}||\hat{\alpha}_n^B-\alpha_0||\geq M'\ |Z^n\right)\\
\leq&P_{V^\infty|Z^\infty}\left(\inf_{\{\mathcal{A}_{osn}:\,\delta_n^{-1}||\alpha-\alpha_0||\geq M'\}}c^*cQ(\alpha)\leq M(\lambda_n+\delta_n^2)|Z^n\right)+\delta.
\end{aligned}
$$

By Assumption 3.4(i)(ii) and $\delta_n=\sqrt{\max\{\lambda_n,\delta_n^2\}}$, we have:

$$
\begin{aligned}
&P_{V^\infty|Z^\infty}\left(\inf_{\{\mathcal{A}_{osn}:\,\delta_n^{-1}||\alpha-\alpha_0||\geq M'\}}c^*cQ(\alpha)\leq M(\lambda_n+\delta_n^2)|Z^n\right)\\
\leq&1\{c^*cc_1\left(M'\delta_n\right)^2\leq M(\lambda_n+\delta_n^2)\},
\end{aligned}
$$

which is eventually naught, because M' can be chosen to be large. The rate under $||\cdot||_s$ immediately follows from this result and the definition of the sieve measure of local ill-posedness τ_n.

For **Result (3)**, we note that both $\hat{\alpha}_n^{R,B},\widehat{\alpha}_n\in\{\alpha\in\mathcal{A}_{k(n)}\colon\phi(\alpha)=\phi(\widehat{\alpha}_n)\}$, and hence all the above proofs go through with $\hat{\alpha}_n^{R,B}$ replacing $\hat{\alpha}_n^B$. In particular, let $\mathcal{A}_{k(n)}^{M_0}(\widehat{\phi})\equiv\{\alpha\in\mathcal{A}_{k(n)}^{M_0}\colon\phi(\alpha)=$

$\phi(\widehat{\alpha}_n)\} \subseteq \mathcal{A}^{M_0}_{k(n)}$. Then: for any $\delta > 0$,

$$
\begin{aligned}
& P_{V^\infty|Z^\infty}\left(\hat{\alpha}_n^{R,B} \in \mathcal{A}^{M_0}_{k(n)}(\widehat{\phi}) : ||\hat{\alpha}_n^{R,B} - \alpha_0||_s \geq \delta | Z^n\right) \\
\leq\ & P_{V^\infty|Z^\infty}\left(\inf_{\{\mathcal{A}^{M_0}_{k(n)}(\widehat{\phi}):\ ||\alpha-\alpha_0||_s \geq \delta\}} \left\{\widehat{Q}_n^B(\alpha) + \lambda_n Pen(h)\right\} \leq \widehat{Q}_n^B(\widehat{\alpha}) + \lambda_n Pen(\widehat{h}) + o(\frac{1}{n}) | Z^n\right) \\
\leq\ & P_{V^\infty|Z^\infty}\left(\inf_{\{\mathcal{A}^{M_0}_{k(n)}(\widehat{\phi}):\ ||\alpha-\alpha_0||_s \geq \delta\}} \left\{c^* \widehat{Q}_n(\alpha) + \lambda_n Pen(h)\right\} \leq c_0^* \left[\widehat{Q}_n(\widehat{\alpha}) + \lambda_n Pen(\widehat{h})\right] + O(\lambda_n) + (\overline{\delta}^*_{m,n})^2 | Z^n\right) \\
\leq\ & P_{V^\infty|Z^\infty}\left(\inf_{\{\mathcal{A}^{M_0}_{k(n)}:\ ||\alpha-\alpha_0||_s \geq \delta\}} \left\{c^* \widehat{Q}_n(\alpha)\right\} \leq c_0^* \left[\widehat{Q}_n(\Pi_n \alpha_0) + \lambda_n Pen(\Pi_n h_0)\right] + O(\lambda_n) + (\overline{\delta}^*_{m,n})^2 | Z^n\right).
\end{aligned}
$$

The rest follows from the proof of Results (1) and (2). *Q.E.D.*

C.2 Proofs for Section A.3 on behaviors under local alternatives

Proof of Theorem A.1: The proof is analogous to that of Theorem 4.3, hence we only present the main steps. Let $\boldsymbol{\alpha}_n = \alpha_0 + d_n \Delta_n$ with $\frac{d\phi(\alpha_0)}{d\alpha}[\Delta_n] = \langle v_n^*, \Delta_n \rangle = \kappa_n = \kappa \times (1 + o(1)) \neq 0$.

Step 1. By assumption 3.6(i) under the local alternatives, for any $t_n \in \mathcal{T}_n$,

$$0 \leq 0.5\left(\widehat{Q}_n(\widehat{\alpha}_n(t_n)) - \widehat{Q}_n(\widehat{\alpha}_n)\right) = t_n \left\{\mathbb{Z}_n(\boldsymbol{\alpha}_n) + \langle u_n^*, \widehat{\alpha}_n - \boldsymbol{\alpha}_n\rangle\right\} + \frac{B_n}{2} t_n^2 + o_{P_{n,Z^\infty}}([r_n(t_n)]^{-1}) \quad \text{(C.1)}$$

where $[r_n(t_n)]^{-1} = \max\{t_n^2, t_n n^{-1/2}, s_n^{-1}\}$ and $s_n^{-1} = o(n^{-1})$. The LHS is always positive (up to possibly a negligible terms given by the penalty function, see the proof of Theorem 4.1(1) for details) by definition of $\widehat{\alpha}_n$. Hence, by choosing $t_n = \pm\{s_n^{-1/2} + o(n^{-1/2})\}$, it follows that $\{\mathbb{Z}_n(\boldsymbol{\alpha}_n) + \langle u_n^*, \widehat{\alpha}_n - \boldsymbol{\alpha}_n\rangle\} = o_{P_{n,Z^\infty}}(n^{-1/2})$. Since $\langle u_n^*, \boldsymbol{\alpha}_n - \alpha_0\rangle = \frac{d_n \kappa_n}{||v_n^*||_{sd}}$ by the definition of local alternatives $\boldsymbol{\alpha}_n$, we obtain equation (C.2):

$$\left\{\mathbb{Z}_n(\boldsymbol{\alpha}_n) + \langle u_n^*, \widehat{\alpha}_n - \alpha_0\rangle - \frac{d_n \kappa_n}{||v_n^*||_{sd}}\right\} = \mathbb{Z}_n(\boldsymbol{\alpha}_n) + \langle u_n^*, \widehat{\alpha}_n - \boldsymbol{\alpha}_n\rangle = o_{P_{n,Z^\infty}}(n^{-1/2}), \qquad \text{(C.2)}$$

where $\mathbb{Z}_n(\boldsymbol{\alpha}_n)$ is defined as that of $\mathbb{Z}_n$ but using $\rho(z, \boldsymbol{\alpha}_n)$ instead of $\rho(z, \alpha_0)$ (since $m(X, \boldsymbol{\alpha}_n) = 0$ a.s.-X under the local alternative).

Next, by Assumption 3.6(i) under the local alternative, we have: for any $t_n \in \mathcal{T}_n$,

$$0.5\left(\widehat{Q}_n(\widehat{\alpha}_n^R(t_n)) - \widehat{Q}_n(\widehat{\alpha}_n^R)\right) = t_n \left\{\mathbb{Z}_n(\boldsymbol{\alpha}_n) + \langle u_n^*, \widehat{\alpha}_n^R - \boldsymbol{\alpha}_n\rangle\right\} + \frac{B_n}{2} t_n^2 + o_{P_{n,Z^\infty}}([r_n(t_n)]^{-1}). \quad \text{(C.3)}$$

By Assumption 3.5(ii)

$$\sup_{\alpha \in \mathcal{N}_{0n}} \left|\phi(\alpha) - \phi(\alpha_0) - \frac{d\phi(\alpha_0)}{d\alpha}[\alpha - \alpha_0]\right| = o(n^{-1/2}||v_n^*||),$$

and assumption $\widehat{\alpha}_n^R \in \mathcal{N}_{osn}$ wpa1-P_{n,Z^∞}, and the fact that $\phi(\widehat{\alpha}_n^R) - \phi(\alpha_0) = 0$, following the same calculations as those in Step 1 of the proof of Theorem 4.3, we have:

$$\langle u_n^*, \widehat{\alpha}_n^R - \alpha_0\rangle = o_{P_{n,Z^\infty}}(n^{-1/2}).$$

Since $\boldsymbol{\alpha}_n = \alpha_0 + d_n\Delta_n \in \mathcal{N}_{osn}$ with $\frac{d\phi(\alpha_0)}{d\alpha}[\Delta_n] = \langle v_n^*, \Delta_n\rangle = \kappa_n$, we have:

$$\langle u_n^*, \widehat{\alpha}_n^R - \boldsymbol{\alpha}_n\rangle = \langle u_n^*, \widehat{\alpha}_n^R - \alpha_0\rangle - \frac{d_n\kappa_n}{||v_n^*||_{sd}} + o_{P_{n,Z^\infty}}(n^{-1/2}) = -\frac{d_n\kappa_n}{||v_n^*||_{sd}} + o_{P_{n,Z^\infty}}(n^{-1/2}).$$

Therefore, by choosing $t_n \equiv -(\mathbb{Z}_n(\boldsymbol{\alpha}_n) - \frac{d_n\kappa_n}{||v_n^*||_{sd}})B_n^{-1}$ in (C.3) with $[r_n(t_n)]^{-1} = \max\{t_n^2, t_n n^{-1/2}, o(n^{-1})\}$ (which is a valid choice), we obtain:

$$\begin{aligned} 0.5\left(\widehat{Q}_n(\widehat{\alpha}_n) - \widehat{Q}_n(\widehat{\alpha}_n^R)\right) &\leq 0.5\left(\widehat{Q}_n(\widehat{\alpha}_n^R(t_n)) - \widehat{Q}_n(\widehat{\alpha}_n^R)\right) + o_{P_{n,Z^\infty}}(n^{-1}) \\ &= -\frac{1}{2}\left(\frac{(\mathbb{Z}_n(\boldsymbol{\alpha}_n) - \frac{d_n\kappa_n}{||v_n^*||_{sd}})}{\sqrt{B_n}}\right)^2 + o_{P_{n,Z^\infty}}([r_n(t_n)]^{-1}). \end{aligned}$$

By our assumption and the fact that $||u_n^*|| \geq c > 0$ for all n, it follows that $B_n \geq c > 0$ eventually, so

$$0.5\left(\widehat{Q}_n(\widehat{\alpha}_n) - \widehat{Q}_n(\widehat{\alpha}_n^R)\right) \leq -\frac{1}{2}\left(\frac{(\mathbb{Z}_n(\boldsymbol{\alpha}_n) - \frac{d_n\kappa_n}{||v_n^*||_{sd}})}{||u_n^*||}\right)^2 \times \left(1 + o_{P_{n,Z^\infty}}(1)\right).$$

Step 2. On the other hand, suppose there exists a t_n^*, such that (a) $\phi(\widehat{\alpha}_n(t_n^*)) = \phi(\alpha_0)$, $\widehat{\alpha}_n(t_n^*) \in \mathcal{A}_{k(n)}$, and (b) $t_n^* = (\mathbb{Z}_n(\boldsymbol{\alpha}_n) - \frac{d_n\kappa_n}{||v_n^*||_{sd}})\,(||u_n^*||)^{-2} + o_{P_{n,Z^\infty}}(n^{-1/2})$. Substituting this into (C.1) with $[r_n(t_n^*)]^{-1} = \max\{(t_n^*)^2, t_n^* n^{-1/2}, o(n^{-1})\}$, we obtain:

$$\begin{aligned} 0.5\left(\widehat{Q}_n(\widehat{\alpha}_n) - \widehat{Q}_n(\widehat{\alpha}_n^R)\right) &\geq 0.5\left(\widehat{Q}_n(\widehat{\alpha}_n) - \widehat{Q}_n(\widehat{\alpha}_n(t_n^*))\right) - o_{P_{n,Z^\infty}}(n^{-1}) \\ &= -\frac{B_n}{2}(t_n^*)^2 + o_{P_{n,Z^\infty}}([r_n(t_n^*)]^{-1}) \\ &= -\frac{B_n}{2}(\mathbb{Z}_n(\boldsymbol{\alpha}_n) - \frac{d_n\kappa_n}{||v_n^*||_{sd}})^2\,(||u_n^*||)^{-4} + o_{P_{n,Z^\infty}}([r_n(t_n^*)]^{-1}) \\ &= -\frac{1}{2}\left(\frac{\mathbb{Z}_n(\boldsymbol{\alpha}_n) - \frac{d_n\kappa_n}{||v_n^*||_{sd}}}{||u_n^*||}\right)^2 \times \left(1 + o_{P_{n,Z^\infty}}(1)\right) \end{aligned}$$

where the second line follows from equation (C.2). Finally, we observe that point (a) follows from Lemma B.2, with $r = 0$, which is of order $n^{-1/2}||v_n^*||$ and thus a valid choice. Point (b) follows by analogous calculations to those in Step 3 of the proof of Theorem 4.3, except that now with $\widehat{\alpha}(t_n^*) = \widehat{\alpha}_n + t_n^* u_n^*$,

$$\begin{aligned} \phi(\widehat{\alpha}(t_n^*)) - \phi(\alpha_0) &= \frac{d\phi(\alpha_0)}{d\alpha}[\widehat{\alpha}_n - \alpha_0] + t_n^*\frac{||v_n^*||^2}{||v_n^*||_{sd}} + o_{P_{n,Z^\infty}}(n^{-1/2}||v_n^*||) \\ &= -\mathbb{Z}_n(\boldsymbol{\alpha}_n)||v_n^*||_{sd} + \frac{d_n\kappa_n}{||v_n^*||_{sd}}||v_n^*||_{sd} + \left((\mathbb{Z}_n(\boldsymbol{\alpha}_n) - \frac{d_n\kappa_n}{||v_n^*||_{sd}})\frac{||v_n^*||_{sd}^2}{||v_n^*||^2}\right)\frac{||v_n^*||^2}{||v_n^*||_{sd}} \\ &\quad + o_{P_{n,Z^\infty}}(n^{-1/2}||v_n^*||) \\ &= o_{P_{n,Z^\infty}}(n^{-1/2}||v_n^*||) \end{aligned}$$

where the second line follows from equation (C.2) and some straightforward algebra.

STEP 3. Finally, the above calculations and $\kappa_n = \kappa(1+o(1))$ imply that

$$||u_n^*||^2 \times \left(\widehat{Q}_n(\widehat{\alpha}_n^R) - \widehat{Q}_n(\widehat{\alpha}_n)\right) = \left(\mathbb{Z}_n(\boldsymbol{\alpha}_n) - \frac{d_n\kappa(1+o(1))}{||v_n^*||_{sd}}\right)^2 \times \left(1 + o_{P_{n,Z^\infty}}(1)\right). \tag{C.4}$$

For **Result (1),** equation (C.4) with $d_n = n^{-1/2}||v_n^*||_{sd}$ implies that

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

which is due to $\sqrt{n}\mathbb{Z}_n(\boldsymbol{\alpha}_n) \Rightarrow N(0,1)$ under the local alternatives.

For **Result (2),** equation (C.4) with $\sqrt{n}\frac{d_n}{||v_n^*||_{sd}} \to \infty$ implies that

$$\begin{aligned} ||u_n^*||^2 \times \widehat{QLR}_n(\phi_0) &= \left(\sqrt{n}\mathbb{Z}_n(\boldsymbol{\alpha}_n) - \sqrt{n}\frac{d_n\kappa(1+o(1))}{||v_n^*||_{sd}}\right)^2 \times \left(1 + o_{P_{n,Z^\infty}}(1)\right) \\ &= \left(O_{P_{n,Z^\infty}}(1) - \sqrt{n}\frac{d_n\kappa(1+o(1))}{||v_n^*||_{sd}}\right)^2 \times \left(1 + o_{P_{n,Z^\infty}}(1)\right), \end{aligned}$$

where the second line is due to $\sqrt{n}\mathbb{Z}_n(\boldsymbol{\alpha}_n) \Rightarrow N(0,1)$ under the local alternatives. Since $\sqrt{n}\frac{d_n\kappa(1+o(1))}{||v_n^*||_{sd}} \to \infty$ (or $-\infty$) if $\kappa > 0$ (or $\kappa < 0$), we have that $\lim_{n\to\infty}\left(||u_n^*||^2 \times \widehat{QLR}_n(\phi_0)\right) = \infty$ in probability (under the alternative). *Q.E.D.*

Proof of Proposition A.1. Recall that $\widehat{QLR}_n^0(\phi_0)$ denotes the optimally-weighted SQLR statistic. By inspection of the proof of Theorem A.1, it is easy to see that

$$||u_n^*||^2 \times \widehat{QLR}_n(\phi_0) = \left(\sqrt{n}\mathbb{Z}_n(\boldsymbol{\alpha}_n) - \kappa\right)^2 + o_{P_{n,Z^\infty}}(1)$$

and

$$\widehat{QLR}_n^0(\phi_0) = \left(\sqrt{n}\mathbb{Z}_n(\boldsymbol{\alpha}_n) - \kappa\frac{||v_n^*||_{sd}}{||v_n^0||_0}\right)^2 + o_{P_{n,Z^\infty}}(1)$$

for local alternatives of the form described in equation (A.2) with $d_n = n^{-1/2}||v_n^*||_{sd}$. Hence, the distribution of $||u_n^*||^2 \times \widehat{QLR}_n(\phi_0)$ is, asympotically close to $\chi_1^2(\kappa^2)$ and the distribution of $\widehat{QLR}_n^0(\phi_0)$ is, asympotically close to $\chi_1^2\left(\frac{||v_n^*||_{sd}^2}{||v_n^0||_0^2}\kappa^2\right)$.

Observe that $\frac{||v_n^*||_{sd}}{||v_n^0||_0} \geq 1$ for all n, and that for a noncentral chi-square, $\chi_p^2(r)$, $\Pr(\chi_p^2(r) \leq t)$ is decreasing in the noncentrality parameter r for each t; thus $\Pr(\chi_p^2(r_1) > t) > \Pr(\chi_p^2(r_2) > t)$ for $r_1 > r_2$. Therefore, the previous results imply that, for any t,

$$\begin{aligned} \lim_{n\to\infty} P_{n,Z^\infty}\left(||u_n^*||^2 \times \widehat{QLR}_n(\phi_0) \geq t\right) &= \Pr(\chi_1^2(\kappa^2) \geq t) \\ &\leq \liminf_{n\to\infty} \Pr\left(\chi_1^2\left(\frac{||v_n^*||_{sd}^2}{||v_n^0||_0^2}\kappa^2\right) \geq t\right) \\ &= \liminf_{n\to\infty} P_{n,Z^\infty}(\widehat{QLR}_n^0(\phi_0) \geq t). \end{aligned}$$

Q.E.D.

Proof of Theorem A.2: The proof of **Result (1)** is similar to that of Theorem 5.3, so we only present a sketch here. By assumptions 3.6(i) and Boot.3(i) under local alternative, it follows that

$$
\begin{aligned}
&0.5\left(\widehat{Q}_n^B(\widehat{\alpha}_n^{R,B}(-\frac{\mathbb{Z}_n^{\omega-1}(\boldsymbol{\alpha}_n)}{B_n^\omega}))-\widehat{Q}_n^B(\widehat{\alpha}_n^{R,B})\right)\\
&=-\frac{\mathbb{Z}_n^{\omega-1}(\boldsymbol{\alpha}_n)}{B_n^\omega}\{\mathbb{Z}_n^\omega(\boldsymbol{\alpha}_n)+\langle u_n^*,\widehat{\alpha}_n^{R,B}-\boldsymbol{\alpha}_n\rangle\}+\frac{(\mathbb{Z}_n^{\omega-1}(\boldsymbol{\alpha}_n))^2}{2B_n^\omega}+o_{P_{V^\infty|Z^\infty}}(r_n^{-1}),\ wpa1(P_{n,Z^\infty}),
\end{aligned}
$$

where $r_n^{-1}=\max\left\{\left(-\frac{\mathbb{Z}_n^{\omega-1}(\boldsymbol{\alpha}_n)}{B_n^\omega}\right)^2,\left|-\frac{\mathbb{Z}_n^{\omega-1}(\boldsymbol{\alpha}_n)}{B_n^\omega}\right|n^{-1/2},o(n^{-1})\right\}=O_{P_{V^\infty|Z^\infty}}(n^{-1}),\ wpa1(P_{n,Z^\infty})$ under assumption Boot.3(i)(ii) with $\boldsymbol{\alpha}_n$ (instead of α_0).

By similar calculations to those in the proof of Result (1) of Theorem 5.3 (equation (B.23)),

$$
\sqrt{n}\langle u_n^*,\widehat{\alpha}_n^{R,B}-\widehat{\alpha}_n\rangle=o_{P_{V^\infty|Z^\infty}}(1),\ wpa1(P_{n,Z^\infty}),
$$

i.e., the restricted bootstrap estimator $\widehat{\alpha}_n^{R,B}$ centers at $\widehat{\alpha}_n$, regardless of the local alternative. Thus

$$
\langle u_n^*,\widehat{\alpha}_n^{R,B}-\boldsymbol{\alpha}_n\rangle=\langle u_n^*,\widehat{\alpha}_n^{R,B}-\widehat{\alpha}_n\rangle+\langle u_n^*,\widehat{\alpha}_n-\boldsymbol{\alpha}_n\rangle=\langle u_n^*,\widehat{\alpha}_n-\boldsymbol{\alpha}_n\rangle+o_{P_{V^\infty|Z^\infty}}(n^{-1/2}),\ wpa1(P_{n,Z^\infty}).
$$

This result and equation (C.2) (i.e., $\mathbb{Z}_n(\boldsymbol{\alpha}_n)+\langle u_n^*,\widehat{\alpha}_n-\boldsymbol{\alpha}_n\rangle=o_{P_{n,Z^\infty}}(n^{-1/2})$) imply that

$$
\begin{aligned}
&0.5\left(\widehat{Q}_n^B(\widehat{\alpha}_n^{R,B}(-\frac{\mathbb{Z}_n^{\omega-1}(\boldsymbol{\alpha}_n)}{B_n^\omega}))-\widehat{Q}_n^B(\widehat{\alpha}_n^{R,B})\right)\\
&=-\frac{\mathbb{Z}_n^{\omega-1}(\boldsymbol{\alpha}_n)}{B_n^\omega}\{\mathbb{Z}_n^\omega(\boldsymbol{\alpha}_n)+\langle u_n^*,\widehat{\alpha}_n-\boldsymbol{\alpha}_n\rangle\}+\frac{(\mathbb{Z}_n^{\omega-1}(\boldsymbol{\alpha}_n))^2}{2B_n^\omega}+o_{P_{V^\infty|Z^\infty}}(r_n^{-1}),\ wpa1(P_{n,Z^\infty})\\
&=-\frac{\mathbb{Z}_n^{\omega-1}(\boldsymbol{\alpha}_n)}{B_n^\omega}\{\mathbb{Z}_n^{\omega-1}(\boldsymbol{\alpha}_n)+o_{P_{n,Z^\infty}}(n^{-1/2})\}+\frac{(\mathbb{Z}_n^{\omega-1}(\boldsymbol{\alpha}_n))^2}{2B_n^\omega}+o_{P_{V^\infty|Z^\infty}}(r_n^{-1}),\ wpa1(P_{n,Z^\infty})\\
&=-\frac{(\mathbb{Z}_n^{\omega-1}(\boldsymbol{\alpha}_n))^2}{2B_n^\omega}\times\left(1+o_{P_{V^\infty|Z^\infty}}(1)\right)\ wpa1(P_{n,Z^\infty}).
\end{aligned}
$$

Following the proof of Result (1) of Theorem 5.3 step 3 with $\mathbb{Z}_n^{\omega-1}(\boldsymbol{\alpha}_n)$ replacing $\mathbb{Z}_n^{\omega-1}$, we obtain:

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

This shows that, since for the bootstrap SQLR the "null hypothesis is $\phi(\alpha)=\widehat{\phi}_n\equiv\phi(\widehat{\alpha}_n)$", it always centers correctly.

By similar calculations to those in the proof of Result (2) of Theorem 5.3, the law of $\left(\sqrt{n}\frac{\mathbb{Z}_n^{\omega-1}(\boldsymbol{\alpha}_n)}{\sigma_\omega\sqrt{B_n^\omega}}\right)^2$ is asymptotically (and $wpa1(P_{n,Z^\infty})$) equal to the law of $\left(\frac{\mathbb{Z}}{||u_n^*||}\right)^2$ where $\mathbb{Z}\sim N(0,1)$. This implies that the a-th quantile of the distribution of $\frac{\widehat{QLR}_n^B(\widehat{\phi}_n)}{\sigma_\omega^2}$, $\widehat{c}_n(a)$, is uniformly bounded $wpa1(P_{n,Z^\infty})$. Also, following the proof of Result (2) of Theorem 5.3 we obtain:

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

This and Theorem A.1 (and the fact that $||u_n^*|| \leq c < \infty$) immediately imply **Results (2)**. *Q.E.D.*

Proof of Theorem A.3: The proof is analogous to that of Theorems 4.2 and A.1 so we only present a sketch here.

Under our assumptions, Theorem 4.2 still holds under the local alternatives $\boldsymbol{\alpha}_n$. Observe that, with $\boldsymbol{\alpha}_n = \alpha_0 + d_n\Delta_n \in \mathcal{N}_{osn}$ and $d_n = o(1)$,

$$\begin{aligned}
\mathcal{T}_n \equiv \sqrt{n}\frac{\phi(\widehat{\alpha}_n) - \phi_0}{||\widehat{v}_n^*||_{n,sd}} &= \sqrt{n}\frac{\phi(\widehat{\alpha}_n) - \phi_0}{||v_n^*||_{sd}} \times (1 + o_{P_{n,Z^\infty}}(1)) \\
&= \sqrt{n}\langle u_n^*, \widehat{\alpha}_n - \alpha_0\rangle \times (1 + o_{P_{n,Z^\infty}}(1)) + o_{P_{n,Z^\infty}}(1) \\
&= \left(-\sqrt{n}\mathbb{Z}_n(\boldsymbol{\alpha}_n) + \sqrt{n}\frac{d_n\kappa\,(1+o(1))}{||v_n^*||_{sd}}\right) \times (1 + o_{P_{n,Z^\infty}}(1)) + o_{P_{n,Z^\infty}}(1),
\end{aligned}$$

where the second line follows from assumption 3.5; the third line follows from equation (C.2), and $\sqrt{n}\mathbb{Z}_n(\boldsymbol{\alpha}_n) \Rightarrow N(0,1)$ under the local alternatives (i.e., assumption 3.6(ii) under the alternatives).

For **Result (1)**, under local alternatives with $d_n = n^{-1/2}||v_n^*||_{sd}$ we have:

$$\mathcal{T}_n = -\left(\sqrt{n}\mathbb{Z}_n(\boldsymbol{\alpha}_n) - \kappa\,(1+o(1))\right) \times \left(1 + o_{P_{n,Z^\infty}}(1)\right) + o_{P_{n,Z^\infty}}(1), \text{ and } \mathcal{W}_n \equiv (\mathcal{T}_n)^2 \Rightarrow \chi_1^2(\kappa^2).$$

For **Result (2),** under local alternatives with $\sqrt{n}\frac{d_n}{||v_n^*||_{sd}} \to \infty$ we have:

$$\mathcal{W}_n \equiv (\mathcal{T}_n)^2 = \left(O_{P_{n,Z^\infty}}(1) - \sqrt{n}\frac{d_n\kappa\,(1+o(1))}{||v_n^*||_{sd}}\right)^2 \times \left(1 + o_{P_{n,Z^\infty}}(1)\right) + o_{P_{n,Z^\infty}}(1) \to \infty \text{ in probability.}$$

Q.E.D.

Proof of Theorem A.4 For **Result (1)**, following the proofs of Theorems 5.2(1) and A.2, we have: under local alternatives $\boldsymbol{\alpha}_n$ defined in (A.2), for $j = 1, 2$,

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

By similar calculations to those in the proof of Theorem 5.2(1), the law of $\sqrt{n}\frac{\mathbb{Z}_n^{\omega-1}(\boldsymbol{\alpha}_n)}{\sigma_\omega\sqrt{B_n^\omega}}$ is asymptotically (and $wpa1(P_{n,Z^\infty})$) equal to the law of $\mathbb{Z} \sim N(0,1)$. Then under the local alternatives $\boldsymbol{\alpha}_n$,

$$\sup_{t\in\mathbb{R}}\left|P_{V^\infty|Z^\infty}\left(\widehat{W}_{j,n}^B \leq t \mid Z^n\right) - P_{Z^\infty}\left(\widehat{W}_n \leq t\right)\right| = o_{P_{V^\infty|Z^\infty}}(1) \; wpa1(P_{n,Z^\infty}), \tag{C.5}$$

where $\lim_{n\to\infty} P_{Z^\infty}\left(\widehat{W}_n \leq t\right) = \Phi(t)$ (i.e., the standard normal cdf). Thus the a-th quantile of the distribution of $\left(\widehat{W}_{j,n}^B\right)^2$, $\widehat{c}_{j,n}(a)$, is uniformly bounded $wpa1(P_{n,Z^\infty})$.

For **Result (2a),** by Theorem A.3(2), Result (1) (i.e., equation (C.5)) and the continuous mapping theorem, we have:

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

Thus by the definition of $\widehat{c}_{j,n}(1-\tau)$ we obtain:

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

Result (2b) directly follows from Theorem A.3(2), equation (C.5) and the continuous mapping theorem. *Q.E.D.*

C.3 Proofs for Section A.4 on asymptotic theory under increasing dimension of ϕ

Lemma C.1. *Let Assumption 3.1(iv) hold. Then: there exist positive finite constants c, C such that*

$$c^2 I_{d(n)} \leq \mathbb{D}_n^2 \leq C^2 I_{d(n)},$$

where $I_{d(n)}$ is the $d(n) \times d(n)$ identity and for matrices $A \leq B$ means that $B - A$ is positive semi-definite.

Proof of Lemma C.1. By Assumption 3.1(iv), the eigenvalues of $\Sigma_0(x)$ and $\Sigma(x)$ are bounded away from zero and infinity uniformly in x. Therefore, for any matrix A,

$$A'\Sigma^{-1}(x)\Sigma_0(x)\Sigma^{-1}(x)A \geq dA'\Sigma^{-1}(x)A$$

and

$$A'\Sigma^{-1}(x)\Sigma_0(x)\Sigma^{-1}(x)A \leq DA'\Sigma^{-1}(x)A$$

for some finite constant $0 < d \leq D < \infty$, and for all x (for matrices $A \leq B$ means that $B - A$ is positive semi-definite). Taking expectations at both sides and choosing $A' \equiv \frac{dm(x,\alpha_0)}{d\alpha}[\mathbf{v}_n^*]'$, these displays imply that

$$\Omega_{sd,n} \geq d\Omega_n \text{ and } \Omega_{sd,n} \leq D\Omega_n$$

Thus

$$\mathbb{D}_n^2 = \Omega_{sd,n}^{1/2}\Omega_n^{-1}\Omega_{sd,n}\Omega_n^{-1}\Omega_{sd,n}^{1/2} \geq d\{\Omega_{sd,n}^{1/2}\Omega_n^{-1}\Omega_{sd,n}^{1/2}\} \geq d^2\Omega_{sd,n}^{1/2}\Omega_{sd,n}^{-1}\Omega_{sd,n}^{1/2} = d^2 I_{d(n)}.$$

Similarly, $\mathbb{D}_n^2 \leq D^2 I_{d(n)}$. *Q.E.D.*

Lemma C.2. *Let $\mathcal{T}_n^M \equiv \{t \in \mathbb{R}^{d(n)} : ||t||_e \leq M_n n^{-1/2}\sqrt{d(n)}\}$. Then:*

$$||\Omega_{sd,n}^{-1/2}\mathbf{Z}_n||_e = O_P\left(n^{-1/2}\sqrt{d(n)}\right) \quad \text{and} \quad \Omega_{sd,n}^{-1/2}\mathbf{Z}_n \in \mathcal{T}_n^M \ wpa1.$$

Proof of Lemma C.2. Let $\Omega_{sd,n}^{-1/2}\mathbf{Z}_n \equiv n^{-1}\sum_{i=1}^n \zeta_{in}$ where $\zeta_{in} \in \mathbb{R}^{d(n)}$. Observe that $E[\zeta_{in}\zeta_{in}'] = I_{d(n)}$. It follows that

$$E_P[(\Omega_{sd,n}^{-1/2}\mathbf{Z}_n)'(\Omega_{sd,n}^{-1/2}\mathbf{Z}_n)] = tr\left\{E_P[\Omega_{sd,n}^{-1/2}\mathbf{Z}_n\mathbf{Z}_n'\Omega_{sd,n}^{-1/2}]\right\} = n^{-2}\sum_{i=1}^n tr\left\{E_P[\zeta_{in}\zeta_{in}']\right\} = n^{-1}d(n),$$

and thus the desired result follows by the Markov inequality. *Q.E.D.*

Lemma C.3. *Let Conditions for Lemma 3.2 and Assumption A.3 hold. Denote $\tilde{\gamma}_n \equiv \sqrt{s_n}(1 + b_n) + a_n$. Then:*

(1) $\left\|\Omega_{sd,n}^{-1/2}\{\mathbf{Z}_n + \langle \mathbf{v}_n^{*\prime}, \widehat{\alpha}_n - \alpha_0\rangle\}\right\|_e = O_P(\sqrt{d(n)}\tilde{\gamma}_n) = o_P(n^{-1/2})$*;*
(2) further let Assumption A.2 hold. Then

$$\left\|\Omega_{sd,n}^{-1/2}\{\mathbf{Z}_n + \phi(\widehat{\alpha}_n) - \phi(\alpha_0)\}\right\|_e = o_P(n^{-1/2}).$$

Proof of Lemma C.3: For **Result (1),** note that $||t||_e^2 = \sum_{l=1}^{d(n)} |t_l|^2$ and if we obtain $|t_l| = O_P(\tilde{\gamma}_n)$ for $\tilde{\gamma}_n$ uniformly over l, then $||t||_e^2 = O_P(d(n)\tilde{\gamma}_n^2)$.

The rest of the proof follows closely the proof of Theorem 4.1 so we only present the main steps. By definition of the approximate PSMD estimator $\widehat{\alpha}_n$, and Assumption A.3(i),

$$0 \leq t'\Omega_{sd,n}^{-1/2}\left(\mathbf{Z}_n + \langle \mathbf{v}_n^{*\prime}, \widehat{\alpha}_n - \alpha_0\rangle\right) + \frac{1}{2}t'\mathbb{B}_n t + O_P(r^{-1}(t)).$$

We now choose $t = \sqrt{s_n}e$ where $e \in \{(1,0,...,0),(0,1,0,...,0),...,(0,...,1)\}$, it is easy to see that this $t \in \mathcal{T}_n^M$, and thus the display above implies

$$0 \leq e'\Omega_{sd,n}^{-1/2}\left(\mathbf{Z}_n + \langle \mathbf{v}_n^{*\prime}, \widehat{\alpha}_n - \alpha_0\rangle\right) + O_P(\tilde{\gamma}_n).$$

By changing the sign of t, it follows that

$$\left|e'\Omega_{sd,n}^{-1/2}\left(\mathbf{Z}_n + \langle \mathbf{v}_n^{*\prime}, \widehat{\alpha}_n - \alpha_0\rangle\right)\right| = O_P(\tilde{\gamma}_n).$$

Observe that the RHS holds uniformly over e, thus, since $e \in \{(1,0,...,0),(0,1,0,...,0),...,(0,...,1)\}$, it follows that

$$\left\|\Omega_{sd,n}^{-1/2}\left(\mathbf{Z}_n + \langle \mathbf{v}_n^{*\prime}, \widehat{\alpha}_n - \alpha_0\rangle\right)\right\|_e = O_P(\sqrt{d(n)}\tilde{\gamma}_n) = o_P(n^{-1/2}),$$

where the second equal sign is due to Assumption A.3(ii).

For **Result (2).** In view of Result (1), it suffices to show that

$$\left\|\Omega_{sd,n}^{-1/2}\{\phi(\widehat{\alpha}_n) - \phi(\alpha_0) - \langle \mathbf{v}_n^{*\prime}, \widehat{\alpha}_n - \alpha_0\rangle\}\right\|_e = o_P(n^{-1/2}).$$

Following the proof of Theorem 4.1 we have:

$$\langle \mathbf{v}_n^{*\prime}, \widehat{\alpha}_n - \alpha_0\rangle = \frac{d\phi(\alpha_0)}{d\alpha}[\widehat{\alpha}_n - \alpha_{0,n}] = \frac{d\phi(\alpha_0)}{d\alpha}[\widehat{\alpha}_n - \alpha_0] - \frac{d\phi(\alpha_0)}{d\alpha}[\alpha_{0,n} - \alpha_0].$$

Since Assumption A.2(ii)(iii) (with $t = 0$) implies that

$$\left\|\Omega_{sd,n}^{-1/2}\{\phi(\widehat{\alpha}_n) - \phi(\alpha_0) - \frac{d\phi(\alpha_0)}{d\alpha}[\widehat{\alpha}_n - \alpha_0] + \frac{d\phi(\alpha_0)}{d\alpha}[\alpha_{0,n} - \alpha_0]\}\right\|_e = O_P(c_n),$$

the desired result now follows from Assumption A.2(iv) of $c_n = o(n^{-1/2})$. *Q.E.D.*

Proof of Theorem A.5. Throughout the proof let $\hat{W}_n \equiv n(\phi(\widehat{\alpha}_n) - \phi(\alpha_0))'\Omega_{sd,n}^{-1}(\phi(\widehat{\alpha}_n) - \phi(\alpha_0))$. By Lemma C.3(2),

$$T_n \equiv (\phi(\widehat{\alpha}_n) - \phi(\alpha_0) + \mathbf{Z}_n)'\Omega_{sd,n}^{-1}(\phi(\widehat{\alpha}_n) - \phi(\alpha_0) + \mathbf{Z}_n) = o_P(n^{-1}).$$

Observe that

$$
\begin{aligned}
&|(\phi(\widehat{\alpha}_n)-\phi(\alpha_0))'\Omega_{sd,n}^{-1}(\phi(\widehat{\alpha}_n)-\phi(\alpha_0))-(\mathbf{Z}_n)'\Omega_{sd,n}^{-1}(\mathbf{Z}_n)| \\
&\leq T_n+2||(\phi(\widehat{\alpha}_n)-\phi(\alpha_0)+\mathbf{Z}_n)'\Omega_{sd,n}^{-1/2}||_e \times ||\Omega_{sd,n}^{-1/2}\mathbf{Z}_n||_e \\
&= o_P(n^{-1})+2||(\phi(\widehat{\alpha}_n)-\phi(\alpha_0)+\mathbf{Z}_n)'\Omega_{sd,n}^{-1/2}||_e \times ||\Omega_{sd,n}^{-1/2}\mathbf{Z}_n||_e = o_P(n^{-1})+o_P(n^{-1}\sqrt{d(n)})
\end{aligned}
$$

where the last equality is due to Lemmas C.2 and C.3(2). Therefore we obtain **Result (1)**:

$$
\hat{W}_n=(\sqrt{n}\mathbf{Z}_n)'\Omega_{sd,n}^{-1}(\sqrt{n}\mathbf{Z}_n)+o_P(\sqrt{d(n)}) \equiv \mathbf{W}_n+o_P(\sqrt{d(n)}).
$$

Result (2) follows directly from Result (1) when $d(n)=d$ is fixed and finite.

Result (3) follows from Result (1) and the following property:

$$
\Xi_n \equiv (2d(n))^{-1/2}(\mathbf{W}_n-d(n)) \Rightarrow N(0,1)
$$

where $\mathbf{W}_n \equiv (\sqrt{n}\mathbf{Z}_n)'\Omega_{sd,n}^{-1}(\sqrt{n}\mathbf{Z}_n)$, or formally,

$$
\sup_{f\in BL_1(\mathbb{R})}|E[f(\Xi_n)]-E[f(\mathbb{Z})]|=o(1)
$$

where $\mathbb{Z}\sim N(0,1)$ and $BL_1(\mathbb{R})$ is the space of bounded (by 1) Lipschitz functions from $\mathbb{R}$ to $\mathbb{R}$.

By triangle inequality it suffices to show that

$$
\sup_{f\in BL_1(\mathbb{R})}|E[f(\Xi_n)]-E[f(\xi_n)]|=o(1) \tag{C.6}
$$

and

$$
\sup_{f\in BL_1(\mathbb{R})}|E[f(\xi_n)]-E[f(\mathbb{Z})]|=o(1) \tag{C.7}
$$

where $\xi_n \equiv (2d(n))^{-1/2}(\sum_{j=1}^{d(n)}\mathbb{Z}_j^2-d(n))$ with $\mathbb{Z}_j\sim N(0,1)$ and independent across $j=1,...,d(n)$. We now show that both equations hold.

Equation C.6. Let $t\mapsto \nu_M(t)\equiv \min\{t't,M\}$ for some $M>0$. Observe

$$
\begin{aligned}
&\left|E[f(\Xi_n)]-E\left[f\left((2d(n))^{-1/2}(\nu_M(\Omega_{sd,n}^{-1/2}\sqrt{n}\mathbf{Z}_n)-d(n))\right)\right]\right| \\
=&\left|E\left[f\left((2d(n))^{-1/2}(\nu_\infty(\Omega_{sd,n}^{-1/2}\sqrt{n}\mathbf{Z}_n)-d(n))\right)-f\left((2d(n))^{-1/2}(\nu_M(\Omega_{sd,n}^{-1/2}\sqrt{n}\mathbf{Z}_n)-d(n))\right)\right]\right| \\
=&\left|\int_{\{z:nz'\Omega_{sd,n}^{-1}z>M\}}\left[f\left(\frac{\nu_\infty(\Omega_{sd,n}^{-1/2}\sqrt{n}z_n)-d(n)}{\sqrt{2d(n)}}\right)-f\left(\frac{M-d(n)}{\sqrt{2d(n)}}\right)\right]P_{Z^\infty}(dz)\right| \\
\leq& 2P_{Z^\infty}\left((\sqrt{n}\mathbf{Z}_n)'\Omega_{sd,n}^{-1}(\sqrt{n}\mathbf{Z}_n)>M\right)
\end{aligned}
$$

where the last line follows from the fact that f is bounded by 1. Therefore, by the Markov inequality, for any ϵ, there exists a M such that the $\left|E[f(\Xi_n)]-E\left[f\left((2d(n))^{-1/2}(\nu_M(\Omega_{sd,n}^{-1/2}\sqrt{n}\mathbf{Z}_n)-d(n))\right)\right]\right|<\epsilon$ for sufficiently large n. A similar result holds if we replace $\Omega_{sd,n}^{-1/2}\sqrt{n}\mathbf{Z}_n$ by $\mathcal{Z}_n=(\mathbb{Z}_1,...,\mathbb{Z}_{d(n)})'$ with $\mathbb{Z}_j\sim N(0,1)$ and independent across $j=1,...,d(n)$. Therefore, in order to show equation

C.6, it suffices to show

$$\sup_{f \in BL_1(\mathbb{R})} |E\left[f\left(\Xi_{M,n}\right)\right] - E[f(\xi_{M,n})]| = o(1)$$

where $\Xi_{M,n} \equiv (2d(n))^{-1/2}(\nu_M(\Omega_{sd,n}^{-1/2}\sqrt{n}\mathbf{Z}_n) - d(n))$ and $\xi_{M,n} \equiv (2d(n))^{-1/2}(\nu_M(\mathcal{Z}_n) - d(n))$.

Since f is uniformly bounded and continuous, it is clear that in order to show the previous display, it suffices to show that

$$(2d(n))^{-1/2}|\nu_M(\Omega_{sd,n}^{-1/2}\sqrt{n}\mathbf{Z}_n) - \nu_M(\mathcal{Z}_n)| = o_P(1). \tag{C.8}$$

It turns out that $|\nu_M(t) - \nu_M(r)| \leq 2\sqrt{M}||t-r||_e$, so $t \mapsto \nu_M(t)$ is Lipschitz (and uniformly bounded). So in order to show equation C.8 it is sufficient to show that for any $\delta > 0$, there exists a $N(\delta)$ such that

$$\Pr\left((2d(n))^{-1/2}||\Omega_{sd,n}^{-1/2}\sqrt{n}\mathbf{Z}_n - \mathcal{Z}_n||_e > \delta\right) < \delta$$

for all $n \geq N(\delta)$. Note that $\Omega_{sd,n}^{-1/2}\sqrt{n}\mathbf{Z}_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n \Psi_n(Z_i)$, with $\Psi_n(z) \equiv \left(\frac{dm(x,\alpha_0)}{d\alpha}[\mathbf{v}_n^*]\Omega_{sd,n}^{-1/2}\right)' \rho(z,\alpha_0)$, and that $\mathcal{Z}_n$ can be cast as $\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathcal{Z}_{n,i}$ with $\mathcal{Z}_{n,i} \sim N(0, I_{d(n)})$, iid across $i = 1, ..., n$. Following the arguments in Section 10.4 of Pollard (2001), we obtain: for any $\delta > 0$,

$$\Pr\left(\left\|\sqrt{n}\mathbf{Z}_n'\Omega_{sd,n}^{-1/2} - \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathcal{Z}_{n,i}\right\|_e > 3\delta\right) \leq \mathcal{Y}_{d(n)}\left(\frac{\mu_{3,n}nd(n)^{5/2}}{(\delta\sqrt{n})^3}\right),$$

for any n, where $x \mapsto \mathcal{Y}_{d(n)}(x) \equiv Cx \times (1 + |\log(1/x)|/d(n))$ and $\mu_{3,n} \equiv E\left[\left\|\left(\frac{dm(X,\alpha_0)}{d\alpha}[\mathbf{v}_n^*]\Omega_{sd,n}^{-1/2}\right)' \rho(Z,\alpha_0)\right\|_e^3\right]$. Therefore,

$$\Pr\left((2d(n))^{-1/2}||\Omega_{sd,n}^{-1/2}\sqrt{n}\mathbf{Z}_n - \mathcal{Z}_n||_e > \delta\right) \leq \mathcal{Y}_{d(n)}\left(\frac{\mu_{3,n}nd(n)^{5/2}}{(\delta/3)^3 d(n)^{3/2} n^{3/2}}\right) = \mathcal{Y}_{d(n)}\left(n^{-1/2}d(n)\frac{\mu_{3,n}}{(\delta/3)^3 8}\right) \to 0$$

provided that $d(n) = o(\sqrt{n}\mu_{3,n}^{-1})$ which is assumed in the Theorem Result (3).

Equation C.7. Observe that $\xi_n \equiv (2d(n))^{-1/2}(\sum_{j=1}^{d(n)} \mathbb{Z}_j^2 - d(n))$ with $\mathbb{Z}_j \sim N(0,1)$ i.i.d. across $j = 1, ..., d(n)$, $E[(\mathbb{Z}_l^2 - 1)] = 0$ and $E[(\mathbb{Z}_l^2 - 1)^2] = 2$. Thus, $\xi_n \Rightarrow N(0,1)$ by a standard CLT. *Q.E.D.*

Lemma C.4. *Let all conditions for Theorem A.6(1) hold. Then there exists a t_n (possibly random) such that: (1) $t_n \in \mathcal{T}_n^M$ wpa1, (2) $\widehat{\alpha}_n(t_n) \in \mathcal{A}_{k(n)}^R = \{\alpha \in \mathcal{A}_{k(n)} : \phi(\alpha) = \phi_0\}$ wpa1, and (3)*

$$n\left|\widehat{Q}_n(\widehat{\alpha}_n(t_n)) - \widehat{Q}_n(\widehat{\alpha}_n(t_n^*))\right| = o_P(\sqrt{d(n)}) \quad \textit{where} \quad t_n^* = \mathbb{D}_n\Omega_{sd,n}^{-1/2}\mathbf{Z}_n.$$

Proof of Lemma C.4: The proof is very similar to Step 3 in the proof of Theorem 4.3. We choose as a candidate

$$t_n = -\mathbb{D}_n\Omega_{sd,n}^{-1/2}\langle \mathbf{v}_n^{*\prime}, \widehat{\alpha}_n - \alpha_0\rangle + \mathbb{D}_n\Omega_{sd,n}^{-1/2}\mathbf{c}_n$$

where $\{\mathbf{c}_n \in \mathbb{R}^{d(n)} : n = 1, 2, 3, ...\}$ is a sequence to be determined later, but has the property that $||\Omega_{sd,n}^{-1/2}\mathbf{c}_n||_e = O_P(c_n)$.

Part (1). By Lemmas C.1, C.2 and C.3, and the choice of t_n, we have:

$$||t_n||_e \leq O_P(\sqrt{d(n)}\{\tilde{\gamma}_n + n^{-1/2}\}) + O_P(c_n)$$

where $\tilde{\gamma}_n \equiv \sqrt{|s_n|}(1+b_n) + a_n = o(n^{-1/2}d(n)^{-1/2})$ (by assumption A.3(ii)) and $c_n = o(n^{-1/2})$ (assumption A.2(iv)). Thus $||t_n||_e = O_P(\sqrt{d(n)}n^{-1/2})$ so $t_n \in \mathcal{T}_n^M$ wpa1.

Part (2). We want to show that $\phi(\widehat{\alpha}_n(t_n)) - \phi(\alpha_0) = 0$ wpa1. then under null $H_0 : \phi(\alpha_0) = \phi_0$ we obtain $\widehat{\alpha}_n(t_n) \in \mathcal{A}^R_{k(n)}$ wpa1.

Under Assumption A.2(i)(ii) and $t_n \in \mathcal{T}_n^M$ wpa1, we have:

$$\left\|\Omega_{sd,n}^{-1/2}\left\{\phi\left(\widehat{\alpha}_n(t_n)\right) - \phi(\alpha_0) - \frac{d\phi(\alpha_0)}{d\alpha}[\widehat{\alpha}_n(t_n) - \alpha_0]\right\}\right\|_e = O_P\left(c_n\right).$$

Since $\widehat{\alpha}_n(t_n) = \widehat{\alpha}_n - \mathbf{v}_n^*\Omega_{sd,n}^{-1/2}\mathbb{D}_n\Omega_{sd,n}^{-1/2}\langle \mathbf{v}_n^{*\prime}, \widehat{\alpha}_n - \alpha_0\rangle + \mathbf{v}_n^*\Omega_{sd,n}^{-1/2}\mathbb{D}_n\Omega_{sd,n}^{-1/2}\mathbf{c}_n$, and $\Omega_{sd,n}^{-1/2}\mathbb{D}_n\Omega_{sd,n}^{-1/2} = \Omega_n^{-1}$ the previous display implies that

$$\begin{aligned}
O_P\left(c_n\right) &= \left\|\Omega_{sd,n}^{-1/2}\left\{\phi\left(\widehat{\alpha}_n(t_n)\right) - \phi(\alpha_0) - \frac{d\phi(\alpha_0)}{d\alpha}[\widehat{\alpha}_n - \alpha_0] + \langle \mathbf{v}_n^{*\prime}, \mathbf{v}_n^*\rangle\Omega_n^{-1}\langle \mathbf{v}_n^{*\prime}, \widehat{\alpha}_n - \alpha_0\rangle - \langle \mathbf{v}_n^{*\prime}, \mathbf{v}_n^*\rangle\Omega_n^{-1}\mathbf{c}_n\right\}\right\|_e \\
&= \left\|\Omega_{sd,n}^{-1/2}\left\{\phi\left(\widehat{\alpha}_n(t_n)\right) - \phi(\alpha_0) - \frac{d\phi(\alpha_0)}{d\alpha}[\widehat{\alpha}_n - \alpha_0] + \langle \mathbf{v}_n^{*\prime}, \widehat{\alpha}_n - \alpha_0\rangle - \mathbf{c}_n\right\}\right\|_e \\
&= \left\|\Omega_{sd,n}^{-1/2}\left\{\phi\left(\widehat{\alpha}_n(t_n)\right) - \phi(\alpha_0) - \frac{d\phi(\alpha_0)}{d\alpha}[\alpha_{0,n} - \alpha_0] - \mathbf{c}_n\right\}\right\|_e.
\end{aligned}$$

Therefore, there exists a $(F_n)_n$ such that $F_n \in \mathbb{R}^{d(n)}$, $||F_n||_e = O_P(c_n)$ and

$$F_n = (\Omega_{sd,n})^{-1/2}\left\{\phi\left(\widehat{\alpha}_n(t_n)\right) - \phi(\alpha_0) - \frac{d\phi(\alpha_0)}{d\alpha}[\alpha_{0,n} - \alpha_0] - \mathbf{c}_n\right\}.$$

If we set $\mathbf{c}_n = (\Omega_{sd,n})^{1/2}F_n - \frac{d\phi(\alpha_0)}{d\alpha}[\alpha_{0,n} - \alpha_0]$, then $||\Omega_{sd,n}^{-1/2}\mathbf{c}_n||_e \leq ||F_n||_e + ||\Omega_{sd,n}^{-1/2}\frac{d\phi(\alpha_0)}{d\alpha}[\alpha_{0n} - \alpha_0]||_e = O_P(c_n)$ by Assumption A.2(iii), so it is indeed a valid choice. From this choice it is easy to see that $\phi\left(\widehat{\alpha}_n(t_n)\right) - \phi(\alpha_0) = 0$ wpa1, as desired.

Part (3). Recall that $\widehat{\alpha}_n \in \mathcal{N}_{osn}$ wpa1 and $\widehat{\alpha}_n(t_n), \widehat{\alpha}_n(t_n^*) \in \mathcal{A}_{k(n)}$. Note that $||t_n||_e = O_P(\sqrt{d(n)}n^{-1/2})$ (by part (1)) and that $||t_n^*||_e = O_P(\sqrt{d(n)}n^{-1/2})$ (by Lemmas C.1 and C.2. So by Assumption A.3(i),

$$\begin{aligned}
&n\left[\widehat{Q}_n(\widehat{\alpha}_n(t_n)) - \widehat{Q}_n(\widehat{\alpha}_n(t_n^*))\right] \\
&= n(t_n - t_n^*)'\Omega_{sd,n}^{-1/2}\{\langle \mathbf{v}_n^{*\prime}, \widehat{\alpha}_n - \alpha_0\rangle + \mathbf{Z}_n\} + 0.5n\{t_n'\mathbb{B}_nt_n - (t_n^*)'\mathbb{B}_nt_n^*\} \\
&+ n \times O_P(s_n + (||t_n||_e + ||t_n^*||_e)a_n + (||t_n||_e^2 + ||t_n^*||_e^2)b_n) \\
&\equiv nT_1 + nT_2 + n \times O_P(s_n + (||t_n||_e + ||t_n^*||_e)a_n + (||t_n||_e^2 + ||t_n^*||_e^2)b_n).
\end{aligned}$$

Observe that $t_n - t_n^* = -\mathbb{D}_n\Omega_{sd,n}^{-1/2}\{\langle \mathbf{v}_n^{*\prime}, \widehat{\alpha}_n - \alpha_0\rangle + \mathbf{Z}_n\} + \mathbb{D}_n\Omega_{sd,n}^{-1/2}\mathbf{c}_n$, so, by Lemmas C.1 and

C.3(1) and Assumption A.2(iv),

$$|T_1| \le ||\Omega_{sd,n}^{-1/2}\{\langle \mathbf{v}_n^{*\prime}, \widehat{\alpha}_n - \alpha_0\rangle + \mathbf{Z}_n\}||_e^2 + ||\Omega_{sd,n}^{-1/2}\{\langle \mathbf{v}_n^{*\prime}, \widehat{\alpha}_n - \alpha_0\rangle + \mathbf{Z}_n\}||_e \times ||\Omega_{sd,n}^{-1/2}\mathbf{c}_n||_e$$
$$= O_P\left(d(n)\tilde{\gamma}_n^2 + \sqrt{d(n)}\tilde{\gamma}_n c_n\right) = o_P(n^{-1}).$$

Regarding T_2, by Lemmas C.1 and C.3 and the definitions of t_n and t_n^*, it follows that

$$\begin{aligned}
|T_2| &= |(t_n - t_n^*)'\mathbb{B}_n(t_n - t_n^*) + 2(t_n - t_n^*)'\mathbb{B}_n t_n^*| \\
&\le O_P((b_n+1)||t_n - t_n^*||_e^2) + O_P((b_n+1)||t_n - t_n^*||_e \times ||t_n^*||_e) \\
&= o_P(n^{-1}) + o_P(n^{-1/2}) \times O_P\left(\sqrt{d(n)}n^{-1/2}\right) = o_P(n^{-1}\sqrt{d(n)}).
\end{aligned}$$

Finally, by the definitions of t_n and t_n^* and Assumption A.3(ii) it follows that

$$ns_n + n(||t_n||_e + ||t_n^*||_e)a_n + n(||t_n||_e^2 + ||t_n^*||_e^2)b_n = ns_n + O_P\left(\sqrt{d(n)}n^{1/2}a_n + d(n)b_n\right) = o_P(\sqrt{d(n)}).$$

Therefore

$$n\left|\widehat{Q}_n(\widehat{\alpha}_n(t_n)) - \widehat{Q}_n(\widehat{\alpha}_n(t_n^*))\right| = o_P(1) + o_P(\sqrt{d(n)}) + o_P(\sqrt{d(n)}) = o_P(\sqrt{d(n)})$$

and the desired result follows. *Q.E.D.*

Proof of Theorem A.6: The proof is very similar to that of Theorem 4.3 and we only provide main steps here.

Step 1. Similar to Steps 1 and 2 in the proof of Theorem 4.3, by the definitions of $\widehat{\alpha}_n^R$ and $\widehat{\alpha}_n$ and Assumption A.3(i), it follows that for any (possibly random) $t \in \mathcal{T}_n^M$,

$$\begin{aligned}
0.5\widehat{QLR}_n(\phi_0) &\ge 0.5n\left(\widehat{Q}_n(\widehat{\alpha}_n^R) - \widehat{Q}_n(\widehat{\alpha}_n^R(t))\right) - o_P(1) \\
&= -n\left(t'\Omega_{sd,n}^{-1/2}\{\mathbf{Z}_n + \langle \mathbf{v}_n^{*\prime}, \widehat{\alpha}_n^R - \alpha_0\rangle\} + 0.5t'\mathbb{B}_n t\right) + O_P(s_n n + n||t||_e a_n + n||t||_e^2 b_n).
\end{aligned}$$

By Assumption A.2(i)(ii),

$$\left\|\Omega_{sd,n}^{-1/2}\left(\underbrace{\phi(\widehat{\alpha}_n^R) - \phi(\alpha_0)}_{=0} - \frac{d\phi(\alpha_0)}{d\alpha}[\widehat{\alpha}_n^R - \alpha_0]\right)\right\|_e = O_P(c_n).$$

Hence, by Assumption A.2(iii),

$$\left\|\Omega_{sd,n}^{-1/2}\langle \mathbf{v}_n^{*\prime}, \widehat{\alpha}_n^R - \alpha_0\rangle\right\|_e = O_P(c_n). \tag{C.9}$$

Since $\sup_{t:||t||_e=1}|t'\{\mathbb{B}_n - \mathbb{D}_n^{-1}\}t| = O_P(b_n)$ by assumption, we have: $t'\mathbb{B}_n t \le |t'\{\mathbb{B}_n - \mathbb{D}_n^{-1}\}t| + t'\mathbb{D}_n^{-1}t \le ||t||_e^2 O_P(b_n) + t'\mathbb{D}_n^{-1}t$ uniformly over $t \in \mathbb{R}^{d(n)}$ with $||t||_e = 1$. This, Assumption A.3(i) and equation (C.9) together imply that

$$0.5\widehat{QLR}_n(\phi_0) \ge -n\left(t'\Omega_{sd,n}^{-1/2}\mathbf{Z}_n + 0.5t'\mathbb{D}_n^{-1}t\right) + O_P(s_n n + n||t||_e(a_n + c_n) + n||t||_e^2 b_n).$$

In the above display we let $t' = -\mathbf{Z}_n'\Omega_{sd,n}^{-1/2}\mathbb{D}_n$, which, by Lemmas C.1 and C.2, is an admissible choice and $||t||_e = O_P\left(n^{-1/2}\sqrt{d(n)}\right)$. Observe that $t_n'\Omega_{sd,n}^{-1/2}\mathbf{Z}_n = -\mathbf{Z}_n'\Omega_{sd,n}^{-1/2}\mathbb{D}_n\Omega_{sd,n}^{-1/2}\mathbf{Z}_n$ and $t_n'\mathbb{D}_n^{-1}t_n = \mathbf{Z}_n'\Omega_{sd,n}^{-1/2}\mathbb{D}_n\Omega_{sd,n}^{-1/2}\mathbf{Z}_n$, we obtain:

$$\begin{aligned}0.5\widehat{QLR}_n(\phi_0) &\geq 0.5(\sqrt{n}\Omega_{sd,n}^{-1/2}\mathbf{Z}_n)'\mathbb{D}_n(\sqrt{n}\Omega_{sd,n}^{-1/2}\mathbf{Z}_n) + O_P\left(s_n n + n^{1/2}\sqrt{d(n)}(a_n+c_n) + d(n)b_n\right)\\ &= 0.5(\sqrt{n}\Omega_{sd,n}^{-1/2}\mathbf{Z}_n)'\mathbb{D}_n(\sqrt{n}\Omega_{sd,n}^{-1/2}\mathbf{Z}_n) + o_P(\sqrt{d(n)}),\end{aligned}$$

where the last equal sign is due to Assumptions A.2(iv) and A.3(ii).

Step 2. Similar to Step 3 in the proof of Theorem 4.3, by the definitions of $\widehat{\alpha}_n^R$ and $\widehat{\alpha}_n$ and the result that $\widehat{\alpha}_n(t_n) \in \mathcal{A}_{k(n)}^R$ (Lemma C.4), with t_n and t_n^* given in Lemma C.4, we obtain:

$$\begin{aligned}0.5\widehat{QLR}_n(\phi_0) &\leq 0.5n\left(\widehat{Q}_n(\widehat{\alpha}_n(t_n)) - \widehat{Q}_n(\widehat{\alpha}_n)\right) + o_P(1)\\ &= 0.5n\left(\widehat{Q}_n(\widehat{\alpha}_n(t_n^*)) - \widehat{Q}_n(\widehat{\alpha}_n)\right) + o_P(\sqrt{d(n)}).\end{aligned}$$

This, Assumption A.3(i)(ii) and the fact that $||t_n^*||_e = O_P\left(n^{-1/2}\sqrt{d(n)}\right)$ together imply:

$$\begin{aligned}0.5\widehat{QLR}_n(\phi_0) &\leq n(t_n^*)'\Omega_{sd,n}^{-1/2}\{\mathbf{Z}_n + \langle \mathbf{v}_n^{*\prime}, \widehat{\alpha}_n - \alpha_0\rangle\} + 0.5n(t_n^*)'\mathbb{B}_n t_n^*\\ &\quad + n \times O_P(s_n + ||t_n^*||_e a_n + ||t_n^*||_e^2 b_n) + o_P(\sqrt{d(n)})\\ &= n(t_n^*)'\Omega_{sd,n}^{-1/2}\{\mathbf{Z}_n + \langle \mathbf{v}_n^{*\prime}, \widehat{\alpha}_n - \alpha_0\rangle\} + 0.5n(t_n^*)'\mathbb{B}_n t_n^* + o_P(\sqrt{d(n)}).\end{aligned}$$

By Lemma C.3 (given that $t_n^* = \mathbb{D}_n\Omega_{sd,n}^{-1/2}\mathbf{Z}_n$, $||t_n^*||_e = O_P\left(n^{-1/2}\sqrt{d(n)}\right)$) and the assumption that $\sup_{t:||t||_e=1}|t'\{\mathbb{B}_n - \mathbb{D}_n^{-1}\}t| = O_P(b_n)$, it follows

$$\begin{aligned}0.5\widehat{QLR}_n(\phi_0) &\leq 0.5n(t_n^*)'\mathbb{D}_n^{-1}t_n^* + o_P\left(\sqrt{d(n)}\right)\\ &= 0.5(\sqrt{n}\Omega_{sd,n}^{-1/2}\mathbf{Z}_n)'\mathbb{D}_n(\sqrt{n}\Omega_{sd,n}^{-1/2}\mathbf{Z}_n) + o_P\left(\sqrt{d(n)}\right).\end{aligned}$$

Step 3. The results in steps 1 and 2 together imply that

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

which establishes **Result (1)**.

Result (2) directly follows from Result (1) and the fact that $\mathbb{D}_n = I_{d(n)}$, $\Omega_{sd,n} = \Omega_{0,n}$ when $\Sigma = \Sigma_0$.

Result (3) follows from Result (2), $\Omega_{sd,n} = \Omega_{0,n}$ when $\Sigma = \Sigma_0$, and the following property of $\mathbf{W}_n \equiv n\mathbf{Z}_n'\Omega_{sd,n}^{-1}\mathbf{Z}_n$:

$$(2d(n))^{-1/2}(\mathbf{W}_n - d(n)) \Rightarrow N(0,1),$$

which has been established in the proof of Theorem A.5 Result (3). *Q.E.D.*

C.4 Proofs for Section A.6 on series LS estimator $\widehat{m}$ and its bootstrap version

Proof of Lemma A.2: For **Result (1)**, since

$$
\begin{aligned}
M_n(Z^n) \equiv & P_{V^\infty|Z^\infty}\left(\sup_{\mathcal{N}_{osn}} \frac{\tau_n}{n}\sum_{i=1}^{n}\left\|\widehat{m}^B(X_i,\alpha)-\widetilde{m}(X_i,\alpha)-\widehat{m}^B(X_i,\alpha_0)\right\|_e^2 \geq M \mid Z^n\right) \\
\leq & P_{V^\infty|Z^\infty}\left(\sup_{\mathcal{N}_{osn}} \frac{\tau_n}{n}\sum_{i=1}^{n}\left\|\widehat{m}^B(X_i,\alpha)-\widehat{m}(X_i,\alpha)-\{\widehat{m}^B(X_i,\alpha_0)-\widehat{m}(X_i,\alpha_0)\}\right\|_e^2 \geq \frac{M}{2} \mid Z^n\right) \\
& + P_{V^\infty|Z^\infty}\left(\sup_{\mathcal{N}_{osn}} \frac{\tau_n}{n}\sum_{i=1}^{n}\left\|\widehat{m}(X_i,\alpha)-\widetilde{m}(X_i,\alpha)-\widehat{m}(X_i,\alpha_0)\right\|_e^2 \geq \frac{M}{2} \mid Z^n\right) \\
\equiv & M_{1,n}(Z^n)+M_{2,n}(Z^n),
\end{aligned}
$$

we have: for all $\delta>0$, there is a $M(\delta)>0$ such that for all $M \geq M(\delta)$,

$$
P_{Z^\infty}\left(M_n(Z^n)\geq 2\delta\right) \leq P_{Z^\infty}\left(M_{1,n}(Z^n)\geq \delta\right)+P_{Z^\infty}\left(M_{2,n}(Z^n)\geq \delta\right).
$$

By following the proof of Lemma C.3(ii) of Chen and Pouzo (2012a), we have that $P_{Z^\infty}\left(M_{2,n}(Z^n)\geq\delta\right)<\delta/2$ eventually. Thus, to establish Result (1), it suffices to bound

$$
P_{Z^\infty}\left(\{M_{1,n}(Z^n)\geq\delta\}\cap\{\lambda_{\min}((P'P)/n)>c\}\right)+P_{Z^\infty}(\lambda_{\min}((P'P)/n)\leq c).
$$

By Assumption A.4(ii)(iii) and theorem 1 in Newey (1997) $\lambda_{\min}((P'P)/n)\geq c>0$ with probability P_{Z^∞} approaching one, hence $P_{Z^\infty}(\lambda_{\min}((P'P)/n)\leq c)<\delta/4$ eventually. To bound the term corresponding to $M_{1,n}$, we note that[21]

$$
\begin{aligned}
& \sum_{i=1}^{n}\left\|\widehat{m}^B\left(X_i,\alpha\right)-\widehat{m}\left(X_i,\alpha\right)-\{\widehat{m}^B(X_i,\alpha_0)-\widehat{m}\left(X_i,\alpha_0\right)\}\right\|_e^2 \\
& =\sum_{i=1}^{n}\Delta\zeta^B(\alpha)'P(P'P)^{-}p^{J_n}(X_i)p^{J_n}(X_i)'(P'P)^{-}P'\Delta\zeta^B(\alpha) \\
& =\Delta\zeta^B(\alpha)'P(P'P)^{-}P'\Delta\zeta^B(\alpha) \\
& \leq\frac{1}{\lambda_{min}((P'P)/n)}\{n^{-1}\Delta\zeta^B(\alpha)'PP'\Delta\zeta^B(\alpha)\};
\end{aligned}
$$

where $\Delta\zeta^B(\alpha)=((\omega_1-1)\Delta\rho(Z_1,\alpha),...,(\omega_n-1)\Delta\rho(Z_n,\alpha))'$ with $\Delta\rho(Z,\alpha)\equiv\rho(Z,\alpha)-\rho(Z,\alpha_0)$. It is thus sufficient to show that, for large enough n,

$$
P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\sup_{\mathcal{N}_{osn}}\frac{\tau_n}{n^2}\Delta\zeta^B(\alpha)'PP'\Delta\zeta^B(\alpha)\geq M \mid Z^n\right)\geq\delta\right)<\delta, \tag{C.10}
$$

which is established in Lemma C.5.

For **Result (2)**, recall that $\ell_n^B(x,\alpha)\equiv\widetilde{m}(x,\alpha)+\widehat{m}^B(x,\alpha_0)$. By similar calculations to those

[21] To ease the notational burden in the proof, we assume $d_\rho=1$; when $d_\rho>1$ the same proof steps hold, component by component.

in Ai and Chen (2003) (p. 1824) it follows

$$
\begin{aligned}
&E_{P_{V^\infty}}\left[n^{-1}\sum_{i=1}^{n}\left\|\widehat{m}^B(X_i,\alpha_0)\right\|_e^2\right]\\
=&E_{P_{V^\infty}}\left[p^{J_n}(X_i)'(P'P)^{-}P'E_{P_{V^\infty|X^\infty}}\left[\rho^B(\alpha_0)\rho^B(\alpha_0)'|X^n\right]P(P'P)^{-}p^{J_n}(X_i)\right]
\end{aligned}
$$

where $\rho^B(\alpha)\equiv(\rho^B(V_1,\alpha),...,\rho^B(V_n,\alpha))'$ with $\rho^B(V_{i.},\alpha)\equiv\omega_i\rho(Z_i,\alpha)$. Note that

$$
\begin{aligned}
E_{P_{V|X^\infty}}[\rho^B(V_i,\alpha_0)\rho^B(V_j,\alpha_0)'|X^n] &= E_{P_\Omega}[\omega_i\omega_j E_{P_{V|X}}[\rho(Z_i,\alpha_0)\rho(Z_j,\alpha_0)'|X_i,X_j]]\\
&=0 \quad \text{for all } i\neq j,
\end{aligned}
$$

and

$$
E_{P_{V|X^\infty}}[\rho^B(V_i,\alpha_0)\rho^B(V_i,\alpha_0)'|X^n]=\sigma_\omega^2\Sigma_0(X_i).
$$

So under Assumption Boot.1 or Boot.2, and by Assumptions 3.1(iv), and A.4(ii), we have: for all $\delta>0$, there is a $M(\delta)>0$ such that for all $M\geq M(\delta)$,

$$
P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\frac{J_n}{n}n^{-1}\sum_{i=1}^{n}\left\|\widehat{m}^B(X_i,\alpha_0)\right\|_e^2\geq M\mid Z^n\right)\geq\delta\right)<\delta.
$$

To establish Result (2), with $(\tau_n')^{-1}=\max\{\frac{J_n}{n},b^2_{m,J_n},(M_n\delta_n)^2\}$, it remains to show that

$$
P_{Z^\infty}\left(\sup_{\mathcal{N}_{osn}}\tau_n'n^{-1}\sum_{i=1}^{n}\|\widetilde{m}(X_i,\alpha)\|_e^2\geq M\right)<\delta. \tag{C.11}
$$

By Lemma SM.1 of Chen and Pouzo (2012b), under Assumptions A.4 and A.5(i), we have: there are finite constants $c,c'>0$ such that, for all $\delta>0$, there is a $N(\delta)$, such that for all $n\geq N(\delta)$,

$$
P_{Z^\infty}\left(\forall\alpha\in\mathcal{N}_{osn}:\ cE_{P_X}\left[||\widetilde{m}\left(X,\alpha\right)||_e^2\right]\leq\frac{1}{n}\sum_{i=1}^{n}||\widetilde{m}\left(X_i,\alpha\right)||_e^2\leq c'E_{P_X}\left[||\widetilde{m}\left(X,\alpha\right)||_e^2\right]\right)>1-\delta.
$$

Thus to show (C.11), it suffices to show that

$$
\sup_{\mathcal{N}_{osn}}\tau_n'E_{P_X}\left[\|\widetilde{m}(X,\alpha)\|_e^2\right]=O(1).
$$

By Assumption A.4(ii) it follows

$$
\begin{aligned}
\sup_{\alpha\in\mathcal{N}_{osn}}E_{P_X}\left[\|\widetilde{m}(X,\alpha)\|_e^2\right]&\leq const.\sup_{\mathcal{N}_{osn}}m(\alpha)'P(P'P)^{-2}P'm(\alpha)\\
&\leq const.\sup_{\mathcal{N}_{osn}}\{||(P'P)^{-}P'(m(\alpha)-P\pi(\alpha))||_e^2+||\pi(\alpha)||_e^2\}\\
&\leq const.\sup_{\alpha\in\mathcal{N}_{osn}}\max\left\{b^2_{m,J_n},||\alpha-\alpha_0||^2\right\}=O((\tau_n')^{-1}),
\end{aligned}
$$

where π is chosen as in Assumption A.4(iv). The last inequality follows from Assumptions A.4(ii)(iii)(iv) and 3.4. We thus obtain Result (2).

For **Result (3)**, we note that

$$\frac{1}{n}\sum_{i=1}^{n}\left\|\widehat{m}^{B}\left(X_i,\alpha\right)\right\|^2_{\widehat{\Sigma}^{-1}}-\frac{1}{n}\sum_{i=1}^{n}\left\|\ell_n^{B}\left(X_i,\alpha\right)\right\|^2_{\widehat{\Sigma}^{-1}}=R^B_{1n}(\alpha)+2R^B_{2n}(\alpha),$$

where

$$R^B_{1n}(\alpha)\equiv\frac{1}{n}\sum_{i=1}^{n}\left\|\widehat{m}^{B}\left(X_i,\alpha\right)-\ell_n^{B}\left(X_i,\alpha\right)\right\|^2_{\widehat{\Sigma}^{-1}},\ R^B_{2n}(\alpha)\le\sqrt{R^B_{1n}}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\|\ell_n^{B}\left(X_i,\alpha\right)\|^2_{\widehat{\Sigma}^{-1}}}.$$

By Result (1) and Assumption 4.1(iii), we have:

$$P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\sup_{\mathcal{N}_{osn}}\tau_n R^B_{1n}(\alpha)\ge M\mid Z^n\right)\ge\delta\right)<\delta$$

with $\tau_n^{-1}=\delta_n^2(M_n\delta_{s,n})^{2\kappa}C_n$. By Results (1) and (2), and Assumption 4.1(iii), we have:

$$P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\sup_{\mathcal{N}_{osn}}\tilde{\tau}_n R^B_{2n}(\alpha)\ge M\mid Z^n\right)\ge\delta\right)<\delta$$

with $\tilde{\tau}_n^{-1}\equiv M_n\delta_n^2(M_n\delta_{s,n})^{\kappa}\sqrt{C_n}$. By Assumption A.5(iii) and the fact that L_n diverges, we obtain the desired result. *Q.E.D.*

In the following we state Lemma C.5 and its proof.

Lemma C.5. *Let Assumptions 3.4(i)(ii), A.4(iii), A.5(i)(ii) and either Boot.1 or Boot.2 hold. Then: for all $\delta>0$, there is a $M(\delta)>0$ such that for all $M\ge M(\delta)$,*

$$P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\sup_{\mathcal{N}_{osn}}\frac{\tau_n}{n^2}\Delta\zeta^B(\alpha)'PP'\Delta\zeta^B(\alpha)\ge M\mid Z^n\right)\ge\delta\right)<0.5\delta$$

eventually, with $\tau_n^{-1}\equiv(\delta_n)^2\left(M_n\delta_{s,n}\right)^{2\kappa}C_n$, where $\Delta\zeta^B(\alpha)=((\omega_1-1)\Delta\rho(Z_1,\alpha),...,(\omega_n-1)\Delta\rho(Z_n,\alpha))'$ and $\Delta\rho(Z,\alpha)\equiv\rho(Z,\alpha)-\rho(Z,\alpha_0)$.

Proof of Lemma C.5: Denote

$$M'_{1n}(Z^n)\equiv P_{V^\infty|Z^\infty}\left(\sup_{\mathcal{N}_{osn}}\frac{\tau_n}{n^2}\Delta\zeta^B(\alpha)'PP'\Delta\zeta^B(\alpha)\ge M\mid Z^n\right).$$

By the Markov inequality

$$M'_{1n}(Z^n)\le M^{-1}E_{P_{V^\infty|Z^\infty}}\left[\sup_{\mathcal{N}_{osn}}\frac{\tau_n}{n^2}\Delta\zeta^B(\alpha)'PP'\Delta\zeta^B(\alpha)\right].$$

Hence it is sufficient to bound

$$\begin{aligned}P_{Z^\infty}\left(M'_{1n}(Z^n)\ge\delta\right)&\le\frac{1}{M\delta}E_{P_{V^\infty}}\left[\sup_{\mathcal{N}_{osn}}\frac{\tau_n}{n^2}\Delta\zeta^B(\alpha)'PP'\Delta\zeta^B(\alpha)\right]\\&=\frac{\tau_n}{nM\delta}\sum_{j=1}^{J_n}E_{P_{V^\infty}}\left[\sup_{\mathcal{N}_{osn}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\omega_i-1)f_j(Z_i,\alpha)\right)^2\right],\end{aligned}$$

where the first inequality follows from the law of iterated expectations and the Markov inequality, and the second equality is due to the notation $f_j(z,\alpha) \equiv p_j(x)\{\rho(z,\alpha) - \rho(z,\alpha_0)\}$.

Under assumption Boot.1, $\{(\omega_i - 1)f_j(Z_i,\alpha)\}_{i=1}^n$ are independent, and thus, by proposition A.1.6 in Van der Vaart and Wellner (1996) (VdV-W),

$$\frac{\tau_n}{nM\delta}\sum_{j=1}^{J_n} E_{P_V^\infty}\left[\sup_{\mathcal{N}_{osn}}\left(n^{-1/2}\sum_{i=1}^n(\omega_i-1)f_j(Z_i,\alpha)\right)^2\right]$$
$$\leq \frac{\tau_n}{nM\delta}\sum_{j=1}^{J_n}\left(E_{P_V^\infty}\left[\sup_{\mathcal{N}_{osn}}\left|n^{-1/2}\sum_{i=1}^n(\omega_i-1)f_j(Z_i,\alpha)\right|\right] + \sqrt{E[\max_{i\leq n}\sup_{\mathcal{N}_{osn}}\left|n^{-1/2}(\omega_i-1)f_j(Z_i,\alpha)\right|^2]}\right)^2.$$

The second term in the RHS is bounded above by

$$\sqrt{nn^{-1}E_{P_V^\infty}[(\omega_i-1)^2\sup_{\mathcal{N}_{osn}}|f_j(Z_i,\alpha)|^2]} \leq \sqrt{E_{P_\omega}[(\omega_i-1)^2]E_{P_Z^\infty}[\sup_{\mathcal{N}_{osn}}|f_j(Z_i,\alpha)|^2]} = O((M_n\delta_{s,n})^\kappa)$$

by Assumptions A.4(iii), A.5(ii) and Boot.1. Hence, under assumption Boot.1 we need to control

$$\frac{\tau_n}{nM\delta}\sum_{j=1}^{J_n}\left(E_{P_V^\infty}\left[\sup_{\mathcal{N}_{osn}}\left|n^{-1/2}\sum_{i=1}^n(\omega_i-1)f_j(Z_i,\alpha)\right|\right]\right)^2 + O\left(\frac{\tau_n J_n}{nM\delta}(M_n\delta_{s,n})^{2\kappa}\right). \tag{C.12}$$

Under Assumption Boot.2, $((\omega_i-1)f_j(Z_i,\alpha))_i$ are *not* independent. So we need to take some additional steps to arrive to an equation of the form of (C.12). Under Assumption Boot.2, it follows

$$\begin{aligned}
&\frac{\tau_n}{M\delta}\sum_{j=1}^{J_n} E_{P_V^\infty}\left[\sup_{\mathcal{N}_{osn}}\left(n^{-1}\sum_{i=1}^n(\omega_i-1)f_j(Z_i,\alpha)\right)^2\right]\\
=&\frac{\tau_n}{M\delta}\sum_{j=1}^{J_n} E_{P_V^\infty}\left[\sup_{\mathcal{N}_{osn}}\left(n^{-1}\sum_{i=1}^n\omega_i f_j(Z_i,\alpha) - n^{-1}\sum_{i=1}^n f_j(Z_i,\alpha)\right)^2\right]\\
=&\frac{\tau_n}{M\delta}\sum_{j=1}^{J_n} E_{P_Z^\infty\times P_{\hat{Z}}^\infty}\left[\sup_{\mathcal{N}_{osn}}\left(n^{-1}\sum_{i=1}^n(\delta_{\hat{Z}_i}-\mathbb{P}_n)[f_j(\cdot,\alpha)]\right)^2\right],
\end{aligned}$$

where the last line follows from the fact that ω_i are the number of times the variable Z_i appear on the bootstrap sample. Thus, the distribution of $\omega_i\delta_{Z_i}$ is the same as that of $\delta_{\hat{Z}_i}$ where $(\hat{Z}_i)_i$ is the bootstrap sample, i.e., an i.i.d. sample from $\mathbb{P}_n \equiv n^{-1}\sum_{i=1}^n\delta_{Z_i}$. By a slight adaptation of lemma 3.6.6 in VdV-W (allowing for square of the norm), it follows

$$E_{P_Z^\infty\times P_{\hat{Z}}^\infty}\left[\sup_{\mathcal{N}_{osn}}\left(n^{-1}\sum_{i=1}^n(\delta_{\hat{Z}_i}-\mathbb{P}_n)[f_j(\cdot,\alpha)]\right)^2\right] \leq E_{P_Z^\infty}\left[E_{P_{\tilde{N}}^\infty}\left[\sup_{\mathcal{N}_{osn}}\left(n^{-1}\sum_{i=1}^n\tilde{N}_i\delta_{Z_i}[f_j(\cdot,\alpha)]\right)^2\right]\right],$$

where $\tilde{N}_i = N_i - N_i'$ with N_i and N_i' being iid Poisson variables with parameter 0.5 ($P_{\tilde{N}}^\infty$ is the corresponding probability). Note that now, $\{\tilde{N}_i f_j(Z_i,\alpha)\}_{i=1}^n$ are independent. So by proposition

A.1.6 in VdV-W,

$$\frac{\tau_n}{nM\delta}\sum_{j=1}^{J_n}E_Q\left[\sup_{\mathcal{N}_{osn}}\left(n^{-1/2}\sum_{i=1}^{n}\tilde{N}_i f_j(Z_i,\alpha)\right)^2\right]$$
$$\leq\frac{\tau_n}{nM\delta}\sum_{j=1}^{J_n}\left(E_Q\left[\sup_{\mathcal{N}_{osn}}\left|n^{-1/2}\sum_{i=1}^{n}\tilde{N}_i f_j(Z_i,\alpha)\right|\right]+\sqrt{E[\max_{i\leq n}\sup_{\mathcal{N}_{osn}}\left|n^{-1/2}\tilde{N}_i f_j(Z_i,\alpha)\right|^2]}\right)^2,$$

where $Q\equiv P_{Z^\infty}\times P_{\tilde{N}^\infty}$. By Cauchy-Schwarz inequality, the second term in the RHS is bounded above by

$$\sqrt{nn^{-1}E_Q[|\tilde{N}|^2\sup_{\mathcal{N}_{osn}}|f_j(Z,\alpha)|^2]}\leq\sqrt{E_{P_{\tilde{N}}}[|\tilde{N}|^2]E_{P_Z}[\sup_{\mathcal{N}_{osn}}|f_j(Z,\alpha)|^2]}=O((M_n\delta_{s,n})^\kappa)$$

by Assumptions A.4(iii) and A.5(ii) and $E[|\tilde{N}|^2]<\infty$. Therefore, under Assumption Boot.2 we need to control

$$\frac{\tau_n}{nM\delta}\sum_{j=1}^{J_n}\left(E_Q\left[\sup_{\mathcal{N}_{osn}}\left|n^{-1/2}\sum_{i=1}^{n}\tilde{N}_i f_j(Z_i,\alpha)\right|\right]\right)^2+O\left(\frac{\tau_n J_n}{nM\delta}((M_n\delta_{s,n})^\kappa)^2\right).\tag{C.13}$$

For both equations, (C.12) and (C.13), we can invoke lemma 2.9.1 in VdV-W and bound the leading term in the equations as follows,

$$\frac{\tau_n}{nM\delta}\sum_{j=1}^{J_n}E_{P_{V^\infty}}\left[\sup_{\mathcal{N}_{osn}}\left|n^{-1/2}\sum_{i=1}^{n}(\omega_i-1)\delta_{Z_i}[f_j(\cdot,\alpha)]\right|\right]$$
$$\leq\frac{\tau_n}{nM\delta}\sum_{j=1}^{J_n}\left\{\int_0^\infty\sqrt{P(|\omega-1|\geq t)}dt\right\}\max_{1\leq l\leq n}E_{P_{Z^\infty}\times P_{\epsilon^\infty}}\left[\sup_{\mathcal{N}_{osn}}\left|l^{-1/2}\sum_{i=1}^{l}\epsilon_i\delta_{Z_i}[f_j(\cdot,\alpha)]\right|\right],\tag{C.14}$$

and

$$\frac{\tau_n}{nM\delta}\sum_{j=1}^{J_n}E_{P_{Z^\infty}}\left[E_{P_{\tilde{N}}}\left[\sup_{\mathcal{N}_{osn}}\left|n^{-1/2}\sum_{i=1}^{n}\tilde{N}_i\delta_{Z_i}[f_j(\cdot,\alpha)]\right|\right]\right]$$
$$\leq\frac{\tau_n}{nM\delta}\sum_{j=1}^{J_n}\left\{\int_0^\infty\sqrt{P(|\tilde{N}|\geq t)}dt\right\}\max_{1\leq l\leq n}E_{P_{Z^\infty}\times P_{\epsilon^\infty}}\left[\sup_{\mathcal{N}_{osn}}\left|l^{-1/2}\sum_{i=1}^{l}\epsilon_i\delta_{Z_i}[f_j(\cdot,\alpha)]\right|\right],\tag{C.15}$$

where $(\epsilon_i)_{i=1}^n$ is a sequence of Rademacher random variables.

Note that $\left\{\int_0^\infty\sqrt{P(|\omega-1|\geq t)}dt\right\}<\infty$ (under Assumption Boot.1), and also $\left\{\int_0^\infty\sqrt{P(|\tilde{N}|\geq t)}dt\right\}\leq$

$2\sqrt{2}$ (see VdV-W p. 351). Hence in both cases we need to bound

$$
\begin{aligned}
&\frac{\tau_n}{nM\delta}\sum_{j=1}^{J_n}\left(\max_{1\le l\le n} E_{P_{Z^\infty}\times P_{\epsilon^\infty}}\left[\sup_{\mathcal{N}_{osn}}\left|l^{-1/2}\sum_{i=1}^{l}\epsilon_i\delta_{Z_i}[f_j(\cdot,\alpha)]\right|\right]\right)^2\\
\le&\frac{\tau_n}{nM\delta}\sum_{j=1}^{J_n}\left(\max_{1\le l\le n} E_{P_{Z^\infty}\times P_{\epsilon^\infty}}\left[\sup_{\mathcal{N}_{osn}}\left|l^{-1/2}\sum_{i=1}^{l}\epsilon_i\delta_{Z_i}[\bar{f}_j(\cdot,\alpha)]\right|\right]\right.\\
&\left.+\max_{1\le l\le n} E_{P_{Z^\infty}\times P_{\epsilon^\infty}}\left[\sup_{\mathcal{N}_{osn}}\left|l^{-1/2}\sum_{i=1}^{l}\epsilon_i E_{P_Z}[f_j(Z,\alpha)]\right|\right]\right)^2\\
\le&2\frac{\tau_n}{nM\delta}\sum_{j=1}^{J_n}\left(\max_{1\le l\le n} E_{P_{Z^\infty}\times P_{\epsilon^\infty}}\left[\sup_{\mathcal{N}_{osn}}\left|l^{-1/2}\sum_{i=1}^{l}\epsilon_i\delta_{Z_i}[\bar{f}_j(\cdot,\alpha)]\right|\right]\right)^2\\
&+2\frac{\tau_n}{nM\delta}\sum_{j=1}^{J_n}\left(\max_{1\le l\le n} E_{P_{Z^\infty}\times P_{\epsilon^\infty}}\left[\sup_{\mathcal{N}_{osn}}\left|l^{-1/2}\sum_{i=1}^{l}\epsilon_i E_{P_Z}[f_j(Z,\alpha)]\right|\right]\right)^2.
\end{aligned}
\tag{C.16}
$$

Note that

$$
\begin{aligned}
&E_{P_{Z^\infty}\times P_{\epsilon^\infty}}\left[\sup_{\mathcal{N}_{osn}}\left|l^{-1/2}\sum_{i=1}^{l}\epsilon_i E_{P_Z}[f_j(Z,\alpha)]\right|\right]\\
&=\sup_{\mathcal{N}_{osn}}|E_{P_Z}[f_j(Z,\alpha)]|E_{P_{\epsilon^\infty}}\left[\left|l^{-1/2}\sum_{i=1}^{l}\epsilon_i\right|\right]\\
&\le\sup_{\mathcal{N}_{osn}}|E_{P_X}[p_j(X)\Delta m(X,\alpha)]|\sqrt{E_{P_{\epsilon^\infty}}\left[\left(l^{-1/2}\sum_{i=1}^{l}\epsilon_i\right)^2\right]}\\
&\le\sqrt{E_{P_Z}[|p_j(X)|^2]}\sup_{\mathcal{N}_{osn}}\sqrt{E_{P_X}[|\Delta m(X,\alpha)|^2]}\sqrt{E_{P_{\epsilon^\infty}}\left[l^{-1}\sum_{i=1}^{l}(\epsilon_i)^2\right]}\\
&=O(\sqrt{E_{P_Z}[|p_j(X)|^2]}M_n\delta_n)=O(M_n^2\delta_n^2),
\end{aligned}
$$

where $\Delta m(X,\alpha)\equiv m(X,\alpha)-m(X,\alpha_0)$ and the inequality follows from Cauchy-Schwarz and the fact that ϵ_i are independent, and the last two equal signs are due to Assumptions 3.4(i)(ii) and A.4(iii).

Hence, by the "desymmetrization lemma" 2.3.6 in VdV-W (note that $\bar{f}_j(Z_i,\alpha)$ are centered), equation (C.16) is bounded above (up to a constant) by

$$
\frac{\tau_n}{nM\delta}\sum_{j=1}^{J_n}\max_{1\le l\le n}\left(E_{P_{Z^\infty}}\left[\sup_{\mathcal{N}_{osn}}\left|l^{-1/2}\sum_{i=1}^{l}\bar{f}_j(Z_i,\alpha)\right|\right]\right)^2+(M_n\delta_n)^2\frac{\tau_nJ_n}{nM\delta}.
$$

Note that $\max\left\{(M_n\delta_n)^2,(M_n\delta_{s,n})^{2\kappa}\right\}=(M_n\delta_{s,n})^{2\kappa}$ (by assumption). and that $\tau_n^{-1}\equiv\frac{J_n}{n}(M_n\delta_{s,n})^{2\kappa}C_n$, it suffices to show that

$$\max_{1\le j\le J_n}\max_{1\le l\le n} E_{P_{Z^\infty}}\left[\sup_{\alpha\in\mathcal{N}_{osn}}\left(l^{-1/2}\sum_{i=1}^{l}\bar{f}_j(Z_i,\alpha)\right)^2\right]\le const.\times(M_n\delta_{s,n})^{2\kappa}C_n. \tag{C.17}$$

By Van der Vaart and Wellner (1996) theorem 2.14.2, we have (up to some omitted constant), for all j,

$$\begin{aligned}&E_{P_{Z^\infty}}\left[\sup_{\alpha\in\mathcal{N}_{osn}}\left|l^{-1/2}\sum_{i=1}^{l}\bar{f}_j(Z_i,\alpha)\right|\right]\\ \le&\left\{(M_n\delta_{s,n})^{\kappa}\int_0^1\sqrt{1+\log N_{[]}(w(M_n\delta_{s,n})^{\kappa},\mathcal{E}_{ojn},||\cdot||_{L^2(f_Z)})}dw\right\}\end{aligned}$$

where $\mathcal{E}_{ojn}=\{p_j(\cdot)(\rho(\cdot,\alpha)-\rho(\cdot,\alpha_0))-E[p_j(\cdot)(\rho(\cdot,\alpha)-\rho(\cdot,\alpha_0)])：\alpha\in\mathcal{N}_{osn}\}$.

Given any $w>0$, let $(\{g_l^m,g_u^m\})_{m=1,...,N(w)}$ be the $||.||_{L^2(f_Z)}$-norm brackets of $\mathcal{O}_{on}$. If $\{\rho(\cdot,\alpha)-\rho(\cdot,\alpha_0)\}\in\mathcal{O}_{on}$ belongs to a bracket $\{g_l^m,g_u^m\}$, then, since $|p_j(x)|<const<\infty$ by Assumption A.4(iii),

$$g_l^m(Z)\le p_j(X)\{\Delta\rho(Z,\alpha)\}\le g_u^m(Z)$$

(where $\{g_l^m,g_u^m\}$ are the original ones, buy scaled by a constant; we keep the same notation to ease the notational burden) and from the previous calculations it is easy to see that

$$\{g_l^m(Z)-E[g_u^m(Z)]\}\le p_j(X)\Delta\rho(Z,\alpha)-E[p_j(X)\Delta\rho(Z,\alpha)]\le\{g_u^m(Z)-E[g_l^m(Z)]\}.$$

So functions of the form $(\{(g_l^m(Z)-E[g_u^m(Z)]),(g_u^m(Z)-E[g_l^m(Z)])\})_{m=1,...,N(w)}$ form $||.||_{L^2(f_V)}$-norm brackets on $\mathcal{E}_{ojn}$. By construction, $N_{[]}(w,\mathcal{E}_{ojn},||.||_{L^2(f_Z)})\le N(w)$. Hence (up to some omitted constants)

$$\begin{aligned}&E_{P_{Z^\infty}}\left[\sup_{\alpha\in\mathcal{N}_{osn}}\left|l^{-1/2}\sum_{i=1}^{l}\bar{f}_j(Z_i,\alpha)\right|\right]\\ \le&(M_n\delta_{s,n})^{\kappa}\max_{j=1,...,J_n}\left\{\int_0^1\sqrt{1+\log N_{[]}(w(M_n\delta_{s,n})^{\kappa},\mathcal{O}_{on},||\cdot||_{L^2(f_Z)})}dw\right\}\\ \le&(M_n\delta_{s,n})^{\kappa}\sqrt{C_n},\end{aligned}$$

where the last inequality follows from assumption A.5(ii). Notice that the above RHS does not depend on l nor on j, so we obtain (C.17). The desired result follows. *Q.E.D.*

Proof of Lemma A.3: Denote

$$T_{nI}^B\equiv\sup_{\mathcal{N}_{osn}}\left|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{d\widetilde{m}(X_i,\alpha)}{d\alpha}[u_n^*]\right)'\widehat{\Sigma}(X_i)^{-1}\ell_n^B(X_i,\alpha)-\frac{1}{n}\sum_{i=1}^{n}\left(\frac{dm(X_i,\alpha_0)}{d\alpha}[u_n^*]\right)'\Sigma(X_i)^{-1}\ell_n^B(X_i,\alpha)\right|,$$

and

$$T_{nII}^B\equiv\sup_{\mathcal{N}_{osn}}\left|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{dm(X_i,\alpha_0)}{d\alpha}[u_n^*]\right)'\Sigma(X_i)^{-1}\ell_n^B(X_i,\alpha)-\{\mathbb{Z}_n^\omega+\langle u_n^*,\alpha-\alpha_0\rangle\}\right|.$$

It suffices to show that for all $\delta>0$, there is $N(\delta)$ such that for all $n\ge N(\delta)$,

$$P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\sqrt{n}T_{nI}^B\ge\delta\mid Z^n\right)\ge\delta\right)<\delta \tag{C.18}$$

and

$$P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\sqrt{n}T^B_{nII} \geq \delta \mid Z^n\right) \geq \delta\right) < \delta. \tag{C.19}$$

We first verify equation (C.18). Note that

$$\begin{aligned}
T^B_{nI} \leq & \sup_{\mathcal{N}_{osn}}\left|\frac{1}{n}\sum_{i=1}^n\left(\frac{d\widetilde{m}(X_i,\alpha)}{d\alpha}[u^*_n] - \frac{dm(X_i,\alpha_0)}{d\alpha}[u^*_n]\right)'\widehat{\Sigma}(X_i)^{-1}\ell^B_n(X_i,\alpha)\right| \\
& + \sup_{\mathcal{N}_{osn}}\left|\frac{1}{n}\sum_{i=1}^n\left(\frac{dm(X_i,\alpha_0)}{d\alpha}[u^*_n]\right)'\{\widehat{\Sigma}(X_i)^{-1} - \Sigma(X_i)^{-1}\}\ell^B_n(X_i,\alpha)\right| \\
\equiv & T^B_{nIa} + T^B_{nIb}.
\end{aligned}$$

By Assumption 4.1(iii) and the Cauchy-Schwarz inequality, it follows that, for some $C \in (0,\infty)$,

$$\begin{aligned}
& P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\sqrt{n}T^B_{nIa} \geq \delta \mid Z^n\right) \geq \delta\right) \leq \\
& P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\sup_{\mathcal{N}_{osn}}\sqrt{\frac{\sum_{i=1}^n\left\|\frac{dm(X_i,\alpha_0)}{d\alpha}[u^*_n] - \frac{d\widetilde{m}(X_i,\alpha)}{d\alpha}[u^*_n]\right\|^2_e}{n}}\sqrt{\frac{\sum_{i=1}^n\|\ell^B_n(X_i,\alpha)\|^2_e}{n}} \geq \frac{C\delta}{\sqrt{n}} \mid Z^n\right) \geq \delta\right) \\
& + P_{Z^\infty}\left(\lambda_{min}(\widehat{\Sigma}(X)) < c\right).
\end{aligned}$$

The second term in the RHS vanishes eventually, so we focus on the first term. It follows

$$\begin{aligned}
& P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\sup_{\mathcal{N}_{osn}}\sqrt{\frac{\sum_{i=1}^n\left\|\frac{dm(X_i,\alpha_0)}{d\alpha}[u^*_n] - \frac{d\widetilde{m}(X_i,\alpha)}{d\alpha}[u^*_n]\right\|^2_e}{n}}\sqrt{\frac{1}{n}\sum_{i=1}^n\|\ell^B_n(X_i,\alpha)\|^2_e} \geq \frac{C\delta}{\sqrt{n}} \mid Z^n\right) \geq \delta\right) \\
\leq & P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\sup_{\mathcal{N}_{osn}}\sqrt{\frac{1}{n}\sum_{i=1}^n\left\|\frac{dm(X_i,\alpha_0)}{d\alpha}[u^*_n] - \frac{d\widetilde{m}(X_i,\alpha)}{d\alpha}[u^*_n]\right\|^2_e}\sqrt{\frac{Mn}{\tau'_n}} \geq C\delta \mid Z^n\right) \geq 0.5\delta\right) \\
& + P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\sup_{\mathcal{N}_{osn}}\sqrt{\frac{\tau'_n}{n}\sum_{i=1}^n\|\ell^B_n(X_i,\alpha)\|^2_e} \geq \sqrt{M} \mid Z^n\right) \geq 0.5\delta\right).
\end{aligned}$$

By Lemma A.2(2) the second term on the RHS is less than 0.5δ eventually (with $(\tau'_n)^{-1} =$

$const.(M_n\delta_n)^2)$. Regarding the first term, note that

$$\sup_{\mathcal{N}_{osn}}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\|\frac{dm(X_i,\alpha_0)}{d\alpha}[u_n^*]-\frac{d\widetilde{m}(X_i,\alpha)}{d\alpha}[u_n^*]\right\|_e^2}\sqrt{\frac{n}{\tau_n'}}$$
$$\leq \sup_{\mathcal{N}_{osn}}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\|\frac{d\widetilde{m}(X_i,\alpha_0)}{d\alpha}[u_n^*]-\frac{d\widetilde{m}(X_i,\alpha)}{d\alpha}[u_n^*]\right\|_e^2\times\frac{n}{\tau_n'}}$$
$$+\sup_{\mathcal{N}_{osn}}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\|\frac{dm(X_i,\alpha_0)}{d\alpha}[u_n^*]-\frac{d\widetilde{m}(X_i,\alpha_0)}{d\alpha}[u_n^*]\right\|_e^2\times\frac{n}{\tau_n'}}$$
$$\leq \sup_{\mathcal{N}_{osn}}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\|\frac{dm(X_i,\alpha_0)}{d\alpha}[u_n^*]-\frac{dm(X_i,\alpha)}{d\alpha}[u_n^*]\right\|_e^2\times\frac{n}{\tau_n'}}+o_{P_{Z^\infty}}(1),$$

by the LS projection property and the definition of $\widetilde{m}$, as well as by the Markov inequality and Assumption A.6(i). Next, by the Markov inequality and Assumption A.7(ii), we have:

$$P_{Z^\infty}\left(\sup_{\mathcal{N}_{osn}}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\|\frac{dm(X_i,\alpha_0)}{d\alpha}[u_n^*]-\frac{dm(X_i,\alpha)}{d\alpha}[u_n^*]\right\|_e^2}\sqrt{\frac{n}{\tau_n'}}\geq 0.5\delta\right)$$
$$\leq \quad \frac{2}{\delta}\sqrt{E_{P_{Z^\infty}}\left[\sup_{\mathcal{N}_{osn}}\left\|\frac{dm(X,\alpha_0)}{d\alpha}[u_n^*]-\frac{dm(X,\alpha)}{d\alpha}[u_n^*]\right\|_e^2\right]\times\frac{n}{\tau_n'}}\to 0.$$

Thus, we established that

$$P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\sqrt{n}T_{nIa}^B\geq\delta\mid Z^n\right)\geq\delta\right)<\delta\quad\text{eventually.}$$

By similar arguments, Assumptions 4.1(iii) and A.5(iv), Lemma A.2(2), and that $\frac{1}{n}\sum_{i=1}^n\left\|\frac{dm(X_i,\alpha_0)}{d\alpha}[u_n^*]\right\|_e^2$ is bounded in probability, it can be shown that

$$P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\sqrt{n}T_{nIb}^B\geq\delta\mid Z^n\right)\geq\delta\right)<\delta,\ \textit{eventually.}$$

Therefore, we establish equation (C.18).

For equation (C.19), let $g(X,u_n^*)\equiv\left(\frac{dm(X,\alpha_0)}{d\alpha}[u_n^*]\right)'\Sigma^{-1}(X)$. Then

$$\begin{aligned}T_{nII}^B&\leq\sup_{\mathcal{N}_{osn}}\left|\frac{1}{n}\sum_{i=1}^{n}g(X_i,u_n^*)\widetilde{m}(X_i,\alpha)-\langle u_n^*,\alpha-\alpha_0\rangle\right|+\left|\frac{1}{n}\sum_{i=1}^{n}g(X_i,u_n^*)\widehat{m}^B(X_i,\alpha_0)-\mathbb{Z}_n^\omega\right|\\&\equiv T_{nIIa}+T_{nIIb}^B.\end{aligned}$$

Thus to show equation (C.19) it suffices to show that $\sqrt{n}T_{nIIa}=o_{P_{Z^\infty}}(1)$ and that

$$P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\sqrt{n}T_{nIIb}^B\geq\delta\mid Z^n\right)\geq\delta\right)<\delta\quad\text{eventually.}\qquad\text{(C.20)}$$

First we consider the term T_{nIIa}. This part of proof is similar to those in Ai and Chen (2003), Ai and Chen (2007) and Chen and Pouzo (2009) for their regular functional $\lambda'\theta$ case, and hence we shall be brief. By the orthogonality properties of the LS projection and the definition of $\widetilde{m}(X_i,\alpha)$ and $\widetilde{g}(X_i,u_n^*)$, we have:

$$n^{-1}\sum_{i=1}^{n} g(X_i,u_n^*)\widetilde{m}(X_i,\alpha) = n^{-1}\sum_{i=1}^{n}\widetilde{g}(X_i,u_n^*)m(X_i,\alpha).$$

By Cauchy-Schwarz inequality,

$$\begin{aligned}
&\sup_{\mathcal{N}_{osn}}\left|\frac{1}{n}\sum_{i=1}^{n}\{\widetilde{g}(X_i,u_n^*)-g(X_i,u_n^*)\}\{m(X_i,\alpha)-m(X_i,\alpha_0)\}\right| \\
\leq\ & \sqrt{\frac{1}{n}\sum_{i=1}^{n}||\widetilde{g}(X_i,u_n^*)-g(X_i,u_n^*)||_e^2}\,\sup_{\mathcal{N}_{osn}}\sqrt{\frac{1}{n}\sum_{i=1}^{n}||m(X_i,\alpha)-m(X_i,\alpha_0)||_e^2}.
\end{aligned}$$

By assumption A.6(iii), $\sqrt{n}\sup_{\mathcal{N}_{osn}}\frac{1}{n}\sum_{i=1}^{n}\left\{||m(X_i,\alpha)-m(X_i,\alpha_0)||_e^2 - E_{P_X}[||m(X_1,\alpha)-m(X_1,\alpha_0)||_e^2]\right\} = o_P(1)$. Thus, since $\sup_{\mathcal{N}_{osn}}E_{P_X}[||m(X_1,\alpha)-m(X_1,\alpha_0)||_e^2] = O(M_n^2\delta_n^2)$, it follows

$$\sup_{\mathcal{N}_{osn}}\frac{1}{n}\sum_{i=1}^{n}||m(X_i,\alpha)-m(X_i,\alpha_0)||_e^2 = O_{P_{Z^\infty}}\left((M_n\delta_n)^2 + o_{P_{Z^\infty}}(n^{-1/2})\right).$$

This, Assumption A.6(ii) and $\delta_n = o(n^{-1/4})$ (by assumption A.5(iv)) imply that

$$\begin{aligned}
&\sup_{\mathcal{N}_{osn}}\left|\frac{1}{n}\sum_{i=1}^{n}\{\widetilde{g}(X_i,u_n^*)-g(X_i,u_n^*)\}\{m(X_i,\alpha)-m(X_i,\alpha_0)\}\right| \\
\leq\ & o_{P_{Z^\infty}}\left(\frac{1}{\sqrt{n}M_n\delta_n}\right)\times O_{P_{Z^\infty}}\left(\sqrt{(M_n\delta_n)^2+o(n^{-1/2})}\right) = o_{P_{Z^\infty}}(n^{-1/2})
\end{aligned}$$

Therefore,

$$\sqrt{n}T_{nIIa} = \sqrt{n}\sup_{\mathcal{N}_{osn}}\left|\frac{1}{n}\sum_{i=1}^{n} g(X_i,u_n^*)m(X_i,\alpha) - \langle u_n^*,\alpha-\alpha_0\rangle\right| + o_{P_{Z^\infty}}(n^{-1/2}).$$

By assumption A.6(iv), $\sqrt{n}\sup_{\mathcal{N}_{osn}}\left|\frac{1}{n}\sum_{i=1}^{n} g(X_i,u_n^*)m(X_i,\alpha) - E_{P_X}\left[g(X_1,u_n^*)\{m(X_1,\alpha)-m(X_1,\alpha_0)\right]\right| = o_{P_{Z^\infty}}(1)$. Thus, by Assumption A.7(iv), we conclude that $\sqrt{n}T_{nIIa} = o_{P_{Z^\infty}}(1)$.

Next we consider the term T_{nIIb}^B. By the orthogonality properties of the LS projection,

$$n^{-1}\sum_{i=1}^{n} g(X_i,u_n^*)\widehat{m}^B(X_i,\alpha_0) = n^{-1}\sum_{i=1}^{n}\widetilde{g}(X_i,u_n^*)\rho^B(V_i,\alpha_0),$$

where $\rho^B(V_i,\alpha_0)\equiv\omega_{i,n}\rho(Z_i,\alpha_0)$ and $\{\omega_{i,n}\}_{i=1}^n$ is independent of $\{Z_i\}_{i=1}^n$.

Hence, by applying the Markov inequality twice, it follows that

$$\begin{aligned}
&P_{Z^\infty}\left(P_{V^\infty|Z^\infty}\left(\sqrt{n}T_{nIIb}^B \geq \delta \mid Z^n\right)\geq\delta\right) \\
\leq\ & \delta^{-4}E_{P_{V^\infty}}\left[n^{-1}\left(\sum_{i=1}^{n}\{g(X_i,u_n^*)-\widetilde{g}(X_i,u_n^*)\}\rho^B(V_i,\alpha_0)\right)^2\right].
\end{aligned}$$

Regarding the cross-products terms where $i \neq j$, note that

$$
\begin{aligned}
&E_{P_{V^\infty}}\left[\{g(X_j,u_n^*)-\widetilde{g}(X_j,u_n^*)\}\{g(X_i,u_n^*)-\widetilde{g}(X_i,u_n^*)\}\rho^B(V_i,\alpha_0)\rho^B(V_j,\alpha_0)\right]\\
=&E_{P_{V^\infty}}\left[\{g(X_j,u_n^*)-\widetilde{g}(X_j,u_n^*)\}\{g(X_i,u_n^*)-\widetilde{g}(X_i,u_n^*)\}E_{P_{V^\infty|X^\infty}}\left[\rho^B(V_i,\alpha_0)\rho^B(V_j,\alpha_0)\mid X^n\right]\right]\\
=&E_{P_{V^\infty}}\left[\{g(X_j,u_n^*)-\widetilde{g}(X_j,u_n^*)\}\{g(X_i,u_n^*)-\widetilde{g}(X_i,u_n^*)\}E_{P_{V^\infty|X^\infty}}\left[\omega_i\omega_j|X^n\right]E_{P_{Z^\infty|X^\infty}}\left[\rho(Z_i,\alpha_0)\rho(Z_j,\alpha_0)\mid X^n\right]\right]\\
=&0,
\end{aligned}
$$

since $E_{P_{Z^\infty|X^\infty}}\left[\rho(Z_i,\alpha_0)\rho(Z_j,\alpha_0)\mid X^n\right]=E_{P_{Z|X}}\left[\rho(Z_i,\alpha_0)X_i\right]E_{P_{Z|X}}\left[\rho(Z_j,\alpha_0)\mid X_j\right]=0$ for $i\neq j$. Thus, it suffices to study

$$
\begin{aligned}
&\delta^{-4}E_{P_{V^\infty}}\left[n^{-1}\sum_{i=1}^{n}\left(g(X_i,u_n^*)-\widetilde{g}(X_i,u_n^*)\right)^2\left(\rho^B(V_i,\alpha_0)\right)^2\right]\\
&=\delta^{-4}n^{-1}\sum_{i=1}^{n}E_{P_{V^\infty}}\left[\left(g(X_i,u_n^*)-\widetilde{g}(X_i,u_n^*)\right)^2E_{P_{V^\infty|X^\infty}}\left[\left(\omega_i\rho(Z_i,\alpha_0)\right)^2\mid X^n\right]\right].
\end{aligned}
$$

By the original-sample $\{Z_i\}_{i=1}^n$ being i.i.d., $\{\omega_{i,n}\}_{i=1}^n$ being independent of $\{Z_i\}_{i=1}^n$, Assumption 3.1(iv) and the fact that $\sigma_\omega^2<\infty$, we can majorize the previous expression (up to an omitted constant) by

$$\delta^{-4}E_{P_{V^\infty}}\left[\left(g(X_i,u_n^*)-\widetilde{g}(X_i,u_n^*)\right)^2\right]=o(1),$$

where the last equality is due to Assumption A.6(ii). Hence we established equation (C.20). The desired result now follows. *Q.E.D.*

Proof of Lemma A.4: By the Cauchy-Schwarz inequality and Assumption 4.1(iii), it suffices to show that

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

By Lemma A.2(2), it suffices to show that

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

Using Markov inequality twice and the LS projection properties, the LHS of the previous equation can be bounded above by

$$\frac{M_n^2\delta_n^2}{\delta^2}E_{P_X}\left[\sup_{\mathcal{N}_{osn}}\left\|\frac{d^2\widetilde{m}(X,\alpha)}{d\alpha^2}[u_n^*,u_n^*]\right\|_e^2\right]\leq\frac{M_n^2\delta_n^2}{\delta^2}E_{P_X}\left[\sup_{\mathcal{N}_{osn}}\left\|\frac{d^2m(X,\alpha)}{d\alpha^2}[u_n^*,u_n^*]\right\|_e^2\right]<\delta$$

eventually, which is satisfied given Assumption A.7(iii). The desired result follows. *Q.E.D.*

Proof of Lemma A.5: For **Result (1)**, we first want to show that

$$\sup_{\mathcal{N}_{osn}}\left|\frac{1}{n}\sum_{i=1}^{n}\left\{\left\|\frac{d\widetilde{m}(X_i,\alpha)}{d\alpha}[u_n^*]\right\|^2_{\widehat{\Sigma}^{-1}}-\left\|\frac{dm(X_i,\alpha_0)}{d\alpha}[u_n^*]\right\|^2_{\Sigma^{-1}}\right\}\right|\leq T_{n,I}+T_{n,II}+T_{n,III}=o_{P_{Z^\infty}}(1) \tag{C.21}$$

where

$$T_{n,I}=\sup_{\mathcal{N}_{osn}}\left|\frac{1}{n}\sum_{i=1}^{n}\left\{\left\|\frac{d\widetilde{m}(X_i,\alpha)}{d\alpha}[u_n^*]\right\|^2_{\widehat{\Sigma}^{-1}}-\left\|\frac{d\widetilde{m}(X_i,\alpha_0)}{d\alpha}[u_n^*]\right\|^2_{\widehat{\Sigma}^{-1}}\right\}\right|,$$

$$T_{n,II}=\left|\frac{1}{n}\sum_{i=1}^{n}\left\{\left\|\frac{d\widetilde{m}(X_i,\alpha_0)}{d\alpha}[u_n^*]\right\|^2_{\widehat{\Sigma}^{-1}}-\left\|\frac{dm(X_i,\alpha_0)}{d\alpha}[u_n^*]\right\|^2_{\widehat{\Sigma}^{-1}}\right\}\right|,$$

$$T_{n,III}=\left|\frac{1}{n}\sum_{i=1}^{n}\left\{\left\|\frac{dm(X_i,\alpha_0)}{d\alpha}[u_n^*]\right\|^2_{\widehat{\Sigma}^{-1}}-\left\|\frac{dm(X_i,\alpha_0)}{d\alpha}[u_n^*]\right\|^2_{\Sigma^{-1}}\right\}\right|.$$

Therefore, to prove equation (C.21), it suffices to show that

$$T_{n,j}=o_{P_{Z^\infty}}(1)\text{ for }j\in\{I,II,III\}.$$

Note that for $||\cdot||_{L^2(P_n)}$ with P_n being the empirical measure, $|||a||^2_{L^2(P_n)}-||b||^2_{L^2(P_n)}|\leq||a-b||^2_{L^2(P_n)}+2|\langle b,a-b\rangle_{L^2(P_n)}|$. Now, let $a\equiv\frac{d\widetilde{m}(X_i,\alpha)}{d\alpha}[u_n^*]$ and $b\equiv\frac{d\widetilde{m}(X_i,\alpha_0)}{d\alpha}[u_n^*]$. In order to show $T_{n,I}=o_{P_{Z^\infty}}(1)$, under Assumption 4.1(iii), it suffices to show

$$\sqrt{n^{-1}\sum_{i=1}^{n}\left\|\frac{d\widetilde{m}(X_i,\alpha_0)}{d\alpha}[u_n^*]\right\|^2_e}\sup_{\mathcal{N}_{osn}}\sqrt{n^{-1}\sum_{i=1}^{n}\left\|\frac{d\widetilde{m}(X_i,\alpha)}{d\alpha}[u_n^*]-\frac{d\widetilde{m}(X_i,\alpha_0)}{d\alpha}[u_n^*]\right\|^2_e}=o_{P_{Z^\infty}}(1).$$

By the property of LS projection, we have:

$$n^{-1}\sum_{i=1}^{n}\left\|\frac{d\widetilde{m}(X_i,\alpha_0)}{d\alpha}[u_n^*]\right\|^2_e\leq n^{-1}\sum_{i=1}^{n}\left\|\frac{dm(X_i,\alpha_0)}{d\alpha}[u_n^*]\right\|^2_e=O_{P_{Z^\infty}}(1)$$

due to iid data, Markov inequality, the definition of $E_{P_{Z^\infty}}\left[\left\|\frac{dm(X_i,\alpha_0)}{d\alpha}[u_n^*]\right\|^2_{\Sigma^{-1}}\right]$ and Assumption 3.1(iv). Next, by the property of LS projection, we have:

$$\begin{aligned}&\sup_{\mathcal{N}_{osn}}n^{-1}\sum_{i=1}^{n}\left\|\frac{d\widetilde{m}(X_i,\alpha)}{d\alpha}[u_n^*]-\frac{d\widetilde{m}(X_i,\alpha_0)}{d\alpha}[u_n^*]\right\|^2_e\\ \leq\ &\sup_{\mathcal{N}_{osn}}n^{-1}\sum_{i=1}^{n}\left\|\frac{dm(X_i,\alpha)}{d\alpha}[u_n^*]-\frac{dm(X_i,\alpha_0)}{d\alpha}[u_n^*]\right\|^2_e=o_{P_{Z^\infty}}(1)\end{aligned}$$

due to iid data, Markov inequality and Assumption A.7(ii). Thus we established $T_{n,I}=o_{P_{Z^\infty}}(1)$.

By similar algebra as before, in order to show $T_{n,II}=o_{P_{Z^\infty}}(1)$, given Assumption 4.1(iii), it suffices to show

$$\sqrt{n^{-1}\sum_{i=1}^{n}\left\|\frac{dm(X_i,\alpha_0)}{d\alpha}[u_n^*]\right\|^2_e}\sqrt{n^{-1}\sum_{i=1}^{n}\left\|\frac{d\widetilde{m}(X_i,\alpha_0)}{d\alpha}[u_n^*]-\frac{dm(X_i,\alpha_0)}{d\alpha}[u_n^*]\right\|^2_e}=o_{P_{Z^\infty}}(1).$$

The term $n^{-1}\sum_{i=1}^{n}\left\|\frac{dm(X_i,\alpha_0)}{d\alpha}[u_n^*]\right\|_e^2 = O_{P_{Z^\infty}}(1)$ is due to iid data, Markov inequality, the definition of $E_{P_{Z^\infty}}\left[\left\|\frac{dm(X_i,\alpha_0)}{d\alpha}[u_n^*]\right\|_{\Sigma^{-1}}^2\right]$ and Assumption 3.1(iv). The term $n^{-1}\sum_{i=1}^{n}\left\|\frac{d\widetilde{m}(X_i,\alpha_0)}{d\alpha}[u_n^*] - \frac{dm(X_i,\alpha_0)}{d\alpha}[u_n^*]\right\|_e^2 = o_{P_{Z^\infty}}(1)$ is due to iid data, Markov inequality and Assumption A.6(i). Thus $T_{n,II} = o_{P_{Z^\infty}}(1)$.

Finally, $T_{n,III} = o_{P_{Z^\infty}}(1)$ follows from the fact that $n^{-1}\sum_{i=1}^{n}\left\|\frac{dm(X_i,\alpha_0)}{d\alpha}[u_n^*]\right\|_e^2 = O_{P_{Z^\infty}}(1)$ and Assumption 4.1(iii). We thus established equation (C.21). Since

$$E_{P_{Z^\infty}}\left[n^{-1}\sum_{i=1}^{n}\left\|\frac{dm(X_i,\alpha_0)}{d\alpha}[u_n^*]\right\|_{\Sigma^{-1}}^2\right] = E_{P_X}\left[g(X,u_n^*)\Sigma(X)g(X,u_n^*)'\right] \leq C < \infty,$$

we obtain Result (1).

Result (2) immediately follows from equation (C.21) and Assumption B. *Q.E.D.*

D Supplement: Sieve Score Statistic and Score Bootstrap

In the main text we present the sieve Wald, SQLR statistics and their bootstrap versions. Here we consider sieve score (or LM) statistic and its bootstrap version. Both the sieve score test and score bootstrap only require to compute the original-sample restricted PSMD estimator of α_0, and hence are computationally attractive.

Recall that $\widehat{\alpha}_n^R$ is the original-sample restricted PSMD estimator (4.10). Let $\widehat{v}_n^{*R}$ be computed in the same way as $\widehat{v}_n^*$ in Subsection 4.2, except that we use $\widehat{\alpha}_n^R$ instead of $\widehat{\alpha}_n$. And

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

Denote

$$\begin{aligned}
\widehat{S}_n &\equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n^R)}{d\alpha}[\widehat{v}_n^{*R}/||\widehat{v}_n^{*R}||_{n,sd}]\right)' \widehat{\Sigma}_i^{-1}\widehat{m}(X_i,\widehat{\alpha}_n^R)\\
\widehat{S}_{1,n} &\equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n^R)}{d\alpha}[\widehat{v}_n^{*R}/||\widehat{v}_n^{*R}||_{n,sd}]\right)' \widehat{\Sigma}_i^{-1}\rho(Z_i,\widehat{\alpha}_n^R)
\end{aligned}$$

and

$$\begin{aligned}
\widehat{S}_n^B &\equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n^R)}{d\alpha}[\widehat{v}_n^{*R}/||\widehat{v}_n^{*R}||_{n,sd}]\right)' \widehat{\Sigma}_i^{-1}\{\widehat{m}^B(X_i,\widehat{\alpha}_n^R)-\widehat{m}(X_i,\widehat{\alpha}_n^R)\}\\
\widehat{S}_{1,n}^B &\equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n^R)}{d\alpha}[\widehat{v}_n^{*R}/||\widehat{v}_n^{*R}||_{n,sd}]\right)' \widehat{\Sigma}_i^{-1}\{(\omega_{i,n}-1)\rho(Z_i,\widehat{\alpha}_n^R)\}.
\end{aligned}$$

Then

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

which coincides with that of $\widehat{S}_{1,n}$ (once adjusted by σ_ω^2).

Following the results in Subsection 4.2 one can compute $\widehat{v}_n^{*R}$ in closed form, $\widehat{v}_n^{*R} = \bar{\psi}^{k(n)}(\cdot)'\widetilde{D}_n^{-1}\widetilde{F}_n$ where

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

And $||\widehat{v}_n^{*R}||_{n,sd}^2 = \widetilde{F}_n'\widetilde{D}_n^{-1}\widetilde{\mho}_n\widetilde{D}_n^{-1}\widetilde{F}_n$ with

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

Therefore, the bootstrap sieve score statistic $\widehat{S}^B_{1,n}$ can be expressed as

$$\begin{aligned}\widehat{S}^B_{1,n} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\frac{d\widehat{m}(X_i,\widehat{\alpha}^R_n)}{d\alpha}[\widehat{v}^{*R}_n/||\widehat{v}^{*R}_n||_{n,sd}]\right)' \widehat{\Sigma}^{-1}_i(\omega_{i,n}-1)\rho(Z_i,\widehat{\alpha}^R_n) \\ &= \left(\widetilde{F}'_n\widetilde{D}^{-1}_n\widetilde{\mho}_n\widetilde{D}^{-1}_n\widetilde{F}_n\right)^{-1/2}\widetilde{F}'_n\widetilde{D}^{-1}_n\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\frac{d\widehat{m}(X_i,\widehat{\alpha}^R_n)}{d\alpha}[\bar{\psi}^{k(n)}(\cdot)']\right)'\widehat{\Sigma}^{-1}_i(\omega_{i,n}-1)\rho(Z_i,\widehat{\alpha}^R_n).\end{aligned}$$

For the case of IID weights, this expression is similar to that proposed in Kline and Santos (2012) for parametric models, which suggests the potential higher order refinements of the bootstrap sieve score test $\left(\widehat{S}^B_{1,n}\right)^2$. We leave it to future research for bootstrap refinement.

In the rest of this section, to simplify presentation, we assume that $\widehat{m}(x,\alpha)$ is a series LS estimator (2.5) of $m(x,\alpha)$. Then we have:

$$\widehat{m}^B(x,\widehat{\alpha}^R_n)-\widehat{m}(x,\widehat{\alpha}^R_n) = \left(\sum_{j=1}^n(\omega_{j,n}-1)\rho(Z_j,\widehat{\alpha}^R_n)p^{J_n}(X_j)'\right)(P'P)^-p^{J_n}(x).$$

When $\widehat{\Sigma}=I$ then we have:

$$\widehat{S}_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\frac{d\widehat{m}(X_i,\widehat{\alpha}^R_n)}{d\alpha}[\widehat{v}^{*R}_n/||\widehat{v}^{*R}_n||_{n,sd}]\right)'\rho(Z_i,\widehat{\alpha}^R_n) = \widehat{S}_{1,n}$$

$$\widehat{S}^B_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\frac{d\widehat{m}(X_i,\widehat{\alpha}^R_n)}{d\alpha}[\widehat{v}^{*R}_n/||\widehat{v}^{*R}_n||_{n,sd}]\right)'(\omega_{i,n}-1)\rho(Z_i,\widehat{\alpha}^R_n) = \widehat{S}^B_{1,n}.$$

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

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

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

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

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

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

Assumption D.1(i) can be obtained by similar conditions to those imposed in Ai and Chen (2003). Assumption D.1(ii) can be established by controlling the entropy, as in VdV-W Chapter 2.11 and $E\left[\left\|\frac{dm(X,\alpha)}{d\alpha}[u^*_n]-\frac{dm(X,\alpha)}{d\alpha}[u]\right\|^2_e\right]=o(1)$ for all $||u^*_n-u||<\zeta_n$; this result is akin to that in lemma 1 of Chen et al. (2003). However, Assumption D.1(ii) can also be obtained by weaker conditions, yielding a $(\Upsilon(\zeta_n))^2$ that is slower than $O(n^{-1/2})$ provided that $\Upsilon(\zeta_n)M_n\delta_n=o(n^{-1/2})$. In the proof we show that $||\widehat{u}^{*R}_n-u^*_n||=o_{P_{Z^\infty}}(1)$; faster rates of convergence will relax the conditions needed to show part (ii).

Theorem D.1. *Let $\widehat{\alpha}_n^R$ be the restricted PSMD estimator (4.10), and conditions for Lemma 3.2 and Proposition B.1 hold. Let Assumptions 3.5, A.4 - A.7, 3.6(ii), 4.1, B.1 and D.1 hold and that $n\delta_n^2 \left(M_n\delta_{s,n}\right)^{2\kappa} C_n = o(1)$. Then, under the null hypothesis of $\phi(\alpha_0) = \phi_0$,*

(1) $\widehat{S}_n = \sqrt{n}\mathbb{Z}_n + o_{P_{Z^\infty}}(1) \Rightarrow N(0,1)$.

(2) Further, if conditions for Lemma A.1 and Assumptions Boot.3(ii), Boot.1 or Boot.2 hold, then:

$$\begin{aligned}
\left|\mathcal{L}_{V^\infty|Z^\infty}(\sigma_\omega^{-1}\widehat{S}_n^B \mid Z^n) - \mathcal{L}(\widehat{S}_n)\right| &= o_{P_{Z^\infty}}(1), \quad \textit{and} \\
\sup_{t\in\mathbb{R}}\left|P_{V^\infty|Z^\infty}(\sigma_\omega^{-1}\widehat{S}_n^B \le t|Z^n) - P_{Z^\infty}(\widehat{S}_n \le t)\right| &= o_{P_{V^\infty|Z^\infty}}(1) \ wpa1(P_{Z^\infty}).
\end{aligned}$$

Proof of Theorem D.1: We first note that by Lemmas **??** and 5.1, Assumptions 3.6(i) and Boot.3(i) hold. Also, by Proposition B.1 we have $\widehat{\alpha}_n^R \in \mathcal{N}_{osn}$ wpa1 under the null hypothesis of $\phi(\alpha_0) = \phi_0$. Under the null hypothesis, and Assumption 3.5, we also have (see Step 1 in the proof of Theorem 4.3):

$$\sqrt{n}\langle u_n^*, \widehat{\alpha}_n^R - \alpha_0\rangle = o_{P_{Z^\infty}}(1).$$

For **Result (1)**, we show that $\widehat{S}_n$ is asymptotically standard normal under the null hypothesis in two steps.

Step 1. We first show that $\left|\frac{||\widehat{v}_n^{*R}||_{sd}}{||\widehat{v}_n^{*R}||_{n,sd}} - 1\right| = o_{P_{Z^\infty}}(1)$ and $||\widehat{u}_n^{*R} - u_n^*|| = o_{P_{Z^\infty}}(1)$, where $\widehat{u}_n^{*R} \equiv \widehat{v}_n^{*R}/||\widehat{v}_n^{*R}||_{sd}$ and $\widehat{v}_n^{*R}$ is computed in the same way as that in Subsection 4.2, except that we use $\widehat{\alpha}_n^R$ instead of $\widehat{\alpha}_n$.

$\left|\frac{||\widehat{v}_n^{*R}||_{sd}}{||\widehat{v}_n^{*R}||_{n,sd}} - 1\right| = o_{P_{Z^\infty}}(1)$ can be established in the same way as that of Theorem 4.2(1). Also, following the proof of Theorem 4.2(1), we obtain:

$$\left\|\frac{\widehat{v}_n^{*R} - v_n^*}{||v_n^*||}\right\| = o_{P_{Z^\infty}}(1), \quad \frac{||\widehat{v}_n^{*R}||}{||v_n^*||_{sd}} = O_{P_{Z^\infty}}(1), \quad \sup_{v\in\overline{\mathbf{V}}_n}\left|\frac{\langle v_n^* - \widehat{v}_n^{*R}, v\rangle}{||v|| \times ||\widehat{v}_n^{*R}||}\right| = o_{P_{Z^\infty}}(1).$$

This and Assumption 3.1(iv) imply that $\left|\frac{\langle \widehat{v}_n^{*R}, \widehat{v}_n^{*R} - v_n^*\rangle}{||\widehat{v}_n^{*R}||_{sd}^2}\right| = o_{P_{Z^\infty}}(1)$ and $\left|\frac{\langle v_n^*, \widehat{v}_n^{*R} - v_n^*\rangle}{||\widehat{v}_n^{*R}||_{sd}^2}\right| = \frac{||v_n^*||_{sd}}{||\widehat{v}_n^{*R}||_{sd}} \times o_{P_{Z^\infty}}(1)$. Therefore,

$$\left|\frac{||v_n^*||_{sd}^2}{||\widehat{v}_n^{*R}||_{sd}^2} - 1\right| \le \left|\frac{\langle \widehat{v}_n^{*R}, \widehat{v}_n^{*R} - v_n^*\rangle}{||\widehat{v}_n^{*R}||_{sd}^2}\right| + \left|\frac{\langle v_n^*, \widehat{v}_n^{*R} - v_n^*\rangle}{||\widehat{v}_n^{*R}||_{sd}^2}\right| = o_{P_{Z^\infty}}(1).$$

and

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

Thus

$$\begin{aligned}
||\widehat{u}_n^{*R} - u_n^*|| &= \left\|\frac{\widehat{v}_n^{*R}}{||\widehat{v}_n^{*R}||_{sd}} - \frac{v_n^*}{||v_n^*||_{sd}}\right\| = \left\|\frac{\widehat{v}_n^{*R}}{||v_n^*||_{sd}}(1 + o_{P_{Z^\infty}}(1)) - \frac{v_n^*}{||v_n^*||_{sd}}\right\| \\
&= \left\|\frac{\widehat{v}_n^{*R} - v_n^*}{||v_n^*||_{sd}}\right\| + o_{P_{Z^\infty}}(\frac{||\widehat{v}_n^{*R}||}{||v_n^*||_{sd}}) = o_{P_{Z^\infty}}(1).
\end{aligned}$$

STEP 2. We show that under the null hypothesis,

$$\widehat{S}_n = \sqrt{n}\mathbb{Z}_n + o_{P_{Z^\infty}}(1) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{dm(X_i,\alpha_0)}{d\alpha}[u_n^*]\right)'\Sigma(X_i)^{-1}\rho(Z_i,\alpha_0) + o_{P_{Z^\infty}}(1). \tag{D.1}$$

By Step 1, it suffices to show that under the null hypothesis,

$$\overline{S}_n \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n^R)}{d\alpha}[\widehat{u}_n^{*R}]\right)'\widehat{\Sigma}^{-1}(X_i)\widehat{m}(X_i,\widehat{\alpha}_n^R) = \sqrt{n}\mathbb{Z}_n + o_{P_{Z^\infty}}(1).$$

Recall that $\ell_n(x,\alpha) \equiv \widehat{m}(x,\alpha_0) + \widetilde{m}(x,\alpha)$. We have:

$$\begin{aligned}
&\left|\overline{S}_n - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n^R)}{d\alpha}[\widehat{u}_n^{*R}]\right)'\widehat{\Sigma}(X_i)^{-1}\ell_n(X_i,\widehat{\alpha}_n^R)\right| \\
\leq & O_{P_{Z^\infty}}(1)\sqrt{n}\sqrt{n^{-1}\sum_{i=1}^{n}\left\|\frac{d\widehat{m}(X_i,\widehat{\alpha}_n^R)}{d\alpha}[\widehat{u}_n^{*R}]\right\|_e^2}\sqrt{n^{-1}\sum_{i=1}^{n}\|\widehat{m}(X_i,\widehat{\alpha}_n^R) - \ell_n(X_i,\widehat{\alpha}_n^R)\|_e^2},
\end{aligned}$$

where the $O_{P_{Z^\infty}}(1)$ is due to Assumption 4.1(iii). By Lemma A.2(1) and the assumption that $n\delta_n^2(M_n\delta_{s,n})^{2\kappa}C_n = o(1)$, we have:

$$\sqrt{n^{-1}\sum_{i=1}^{n}\|\widehat{m}(X_i,\widehat{\alpha}_n^R) - \ell_n(X_i,\widehat{\alpha}_n^R)\|_e^2} = o_{P_{Z^\infty}}(n^{-1/2}).$$

Also $n^{-1}\sum_{i=1}^{n}\left\|\frac{d\widehat{m}(X_i,\widehat{\alpha}_n^R)}{d\alpha}[\widehat{u}_n^{*R}]\right\|_e^2 \asymp 1$ by Step 1. Therefore

$$\overline{S}_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n^R)}{d\alpha}[\widehat{u}_n^{*R}]\right)'\widehat{\Sigma}(X_i)^{-1}\ell_n(X_i,\widehat{\alpha}_n^R) + o_{P_{Z^\infty}}(1).$$

Assumption D.1(i) implies that

$$n^{-1}\sum_{i=1}^{n}\left\|\frac{d\widehat{m}(X_i,\widehat{\alpha}_n^R)}{d\alpha}[\widehat{u}_n^{*R}] - \frac{dm(X_i,\widehat{\alpha}_n^R)}{d\alpha}[\widehat{u}_n^{*R}]\right\|_e^2 = O_{P_{Z^\infty}}(\max\left\{n^{-1/2},\epsilon_n^2\right\}).$$

And $n^{-1}\sum_{i=1}^{n}\left\|\ell_n(X_i,\widehat{\alpha}_n^R)\right\|_e^2 = O_{P_{Z^\infty}}((M_n\delta_n)^2)$ by Lemma A.2(2). These results, Assumption D.1(i) and Assumption 4.1(iii) together lead to

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{d\widehat{m}(X_i,\widehat{\alpha}_n^R)}{d\alpha}[\widehat{u}_n^{*R}]\right)'\widehat{\Sigma}(X_i)^{-1}\ell_n(X_i,\widehat{\alpha}_n^R) \\
= &\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{dm(X_i,\widehat{\alpha}_n^R)}{d\alpha}[\widehat{u}_n^{*R}]\right)'\Sigma(X_i)^{-1}\ell_n(X_i,\widehat{\alpha}_n^R) + o_{P_{Z^\infty}}(1) \\
= &\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{dm(X_i,\widehat{\alpha}_n^R)}{d\alpha}[u_n^*]\right)'\Sigma(X_i)^{-1}\ell_n(X_i,\widehat{\alpha}_n^R) + o_{P_{Z^\infty}}(1),
\end{aligned}$$

where the second equality is due to $||\widehat{u}_n^{*R} - u_n^*|| = O_{P_{Z^\infty}}(\zeta_n)$ (Assumption D.1(iii)) and Assumption

D.1(ii).

Since $\widehat{\alpha}_n^R \in \mathcal{N}_{osn}$ wpa1 under the null hypothesis, $\sqrt{n}\langle u_n^*, \widehat{\alpha}_n^R - \alpha_0\rangle = o_{P_{Z^\infty}}(1)$, and by analogous calculations to those in the proof of Lemma A.3, we obtain:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\frac{dm(X_i,\widehat{\alpha}_n^R)}{d\alpha}[u_n^*]\right)' \Sigma(X_i)^{-1}\ell_n(X_i,\widehat{\alpha}_n^R) = \sqrt{n}\mathbb{Z}_n + o_{P_{Z^\infty}}(1),$$

and hence equation (D.1) holds. By Assumption 3.6(ii) we have: $\widehat{S}_n \Rightarrow N(0,1)$ under the null hypothesis.

For **Result (2)**, we now show that $\widehat{S}_n^B$ also converges weakly (in the sense of Bootstrap Section 5) to a standard normal under the null hypothesis. It suffices to show that

$$\widehat{S}_n^B = \frac{1}{\sqrt{n}}\sum_{i=1}^n (\omega_i - 1)\left(\frac{dm(X_i,\alpha_0)}{d\alpha}[u_n^*]\right)' \Sigma(X_i)^{-1}\rho(Z_i,\alpha_0) + o_{P_{V^\infty|Z^\infty}}(1)\ wpa1(P_{Z^\infty}). \tag{D.2}$$

Note that $\ell_n^B(X_i,\widehat{\alpha}_n^R) - \ell_n(X_i,\widehat{\alpha}_n^R)) = \widehat{m}^B(X_i,\alpha_0) - \widehat{m}(X_i,\alpha_0)$, and that $n^{-1}\sum_{i=1}^n ||\widehat{m}^B(X_i,\alpha_0) - \widehat{m}(X_i,\alpha_0)||_e^2 = O_{P_{V^\infty|Z^\infty}}(J_n/n)$ wpa1(P_{Z^∞}) (see the proof of Lemma A.2). We have, by calculations similar to Step 2,

$$\left|\widehat{S}_n^B - \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\frac{dm(X_i,\widehat{\alpha}_n^R)}{d\alpha}[u_n^*]\right)' \Sigma(X_i)^{-1}\{\ell_n^B(X_i,\widehat{\alpha}_n^R) - \ell_n(X_i,\widehat{\alpha}_n^R)\}\right| = o_{P_{V^\infty|Z^\infty}}(1)\ wpa1(P_{Z^\infty}).$$

By analogous calculations to those in the proof of Lemma A.3, we obtain equation (D.2). This and Result (1) and Assumption Boot.3(ii) now imply that under the null and conditional on the data, $\sigma_\omega^{-1}\widehat{S}_n^B$ is also asymptotically standard normally distributed. The last part of Result (2) can be established in the same way as that of Theorem 5.2(1), and is omitted. *Q.E.D.*