# Inference for Point and Partially Identified Semi-Nonparametric Conditional Moment Models*

Jing Tao†
*Department of Economics, University of Wisconsin-Madison*
December 3 2014
Job Market Paper

**Abstract**

This paper considers semi-nonparametric conditional moment models where the parameters of interest include both finite-dimensional parameters and unknown functions. We mainly focus on two inferential problems in this framework. First, we provide new methods of uniform inference for the estimates of both finite- and infinite-dimensional components of the parameters and functionals of the parameters. Based on these results, we can, for instance, construct uniform confidence bands for the unknown functions and the partial derivatives of the unknown functions. Recently, uniform confidence bands for a variety of models such as conditional mean and quantiles have been introduced using strong approximation methods (Belloni, Chernozhukov and Fernández-Val, 2011, and related work). We extend the strong approximation approach to provide uniform inference in conditional moment restriction models with endogeneity. Second, for a large class of conditional moment restrictions models, we provide new results for inference when parameters are only partially identified. Under partial identification, we show how to construct pointwise confidence regions by inverting a quasi-likelihood ratio (QLR) statistic that is also employed under point identification. We provide a consistent multiplier bootstrap procedure for obtaining critical values corresponding to the QLR. Furthermore, we generalize the uniform confidence bands from point identified case to uniform confidence sets over the domain of the unknown functions by inverting a sup-QLR statistic. The new methods are applied to construct pointwise confidence intervals and uniform confidence bands for shape-invariant Engel curves.

Keywords: Conditional moment restrictions, Sieve generalized method of moments, Irregular functionals, Sieve Wald, Sieve quasi likelihood ratio, Sieve sup-Wald, Sieve sup-quasi likelihood ratio, Strong approximation, Partial identification, Multiplier bootstrap

*Latest version is available at: http://www.ssc.wisc.edu/~jtao/research.

†I am deeply grateful to my advisors Jack Porter, Bruce Hansen and Xiaoxia Shi for their advice and constant support. I thank Joachim Freyberger and Amit Gandhi for many comments and discussions. I thank Xiaohong Chen for providing relevant results and her very thoughtful and helpful suggestions and comments. I also benefited from valuable feedbacks from participants at 2013 Midwest Econometric Group Annal Meeting, 2014 Summer Meeting of the Econometric Society and the UW-Madison econometrics seminars. All errors are mine. Comments are very welcome. Email: jtao2@wisc.edu.

# 1 INTRODUCTION

It is now commonplace for economic models to be specified in terms of a set of conditional moment restrictions. Conditional moment restrictions provide a general, flexible framework for incorporating nonlinearities and non-Gaussian unobserved error distributions. In addition, theses models treat instrumental variable (IV) conditions as special cases and enable an understanding of underlying structural relations even when some of the regressors are endogenous. Since the groundbreaking work of Hansen (1982) and Hansen and Singleton (1982), parametric moment restriction models have been applied extensively and have also been extended to models that allow for semiparametric and nonparametric specifications.

This paper considers a semi-nonparametric framework and provides pointwise and uniform inference for parameters in models with conditional moment restrictions such that

$$E\left[\rho\left(Z;\beta_0,h_0(\cdot)\right)|X\right]=0 \ a.s. \ X, \tag{1.1}$$

where $\rho(\cdot;\beta,h(\cdot))$ is a $d_\rho-$vector of generalized residual functions that are known up to the parameters $\theta_0 \equiv (\beta_0, h_0)$, $\beta_0$ is a vector of finite-dimensional parameters, $h_0(\cdot)$ is a vector of unknown functions (the infinite-dimensional parameters), $Z$ is a vector of endogenous variables and $X$ is a vector of conditioning variables (instrument variables).[1] We call the model semi-nonparametric in the sense that our estimation and inference methods cover *both* parametric $\beta_0$ and nonparametric $h_0$ components.

Our primary interest is inference on functionals of $\theta_0$. Functionals of $\theta_0$ include, for example, the parametric term $\beta_0$, the unknown function evaluated at a fixed point $h_0(\bar{w})$, or linear combinations of the two components such as $\beta_0+\bigtriangledown h_0(\bar{w})$, where $\bigtriangledown h_0(\bar{w})$ is the derivative of the unknown function at a fixed point. In addition, we investigate the inference on $h_0(\cdot)$ or $\bigtriangledown h_0(\cdot)$ that is uniform over the domain of the unknown function (or any subset of that domain). In this work, *pointwise* inference indicates inference for functionals of $\theta_0$ at a fixed point, while *uniform* inference indicates inference uniformly over the domain of the functionals.

Uniform inference for models of the form (1.1) is the first main objective of interest addressed in this work. We develop inference methods that can be used to provide uniform confidence bands for unknown functions and functionals of unknown functions. Uniform inference methods are motivated by testing shape restrictions that come from economic theory. Based on these results, we can, for example, construct uniform confidence bands for unknown functions themselves, derivatives of unknown functions, the conditional average derivatives of unknown functions and other similar quantities.

Our second point of emphasis is the extension of inferential methods to Model (1.1) under

[1] Model (1.1) encompasses many important classes of econometric models. For instance, it includes the nonparametric regression (e.g., Andrews (1991) and Newey (1997)), the partially linear regression/instrumental variable (IV) regression (Robinson (1988), Ai and Chen, (2003)), the nonparametric IV regression (Newey and Powell (2003), Horowitz (2011)). The relationship to the literature will be discussed below.

partial identification. Based on a quasi-likelihood ratio (QLR) statistic, we provide pointwise and uniform inference methods for functionals of parameters for a class of conditional moment restrictions models under partial identification. Our inference methods are *robust* in the sense that we can construct confidence sets for a large class of conditional moment restriction models without knowing if the parameters are point identified or not *a priori*.

Constructing uniform confidence bands and conducting robust inference methods of parameters for Model (1.1) involve several important challenges.

First, it is difficult to establish the asymptotic theory for uniform inference for the entire support of the nonparametric functionals when the generalized residual function $\rho(\cdot, \beta_0, h_0)$ contains unknown functions of endogenous variables. Ideally, in order to perform uniform inference to construct uniform confidence bands of functionals, researchers would like to invert a valid test statistic to obtain critical values by employing modern empirical process theory to establish a limiting distribution of the statistic. However, in contrast to parametric models, the empirical processes arising in these problems do not converge weakly to Gaussian processes because they are not stochastically equicontinuous. Thus, closed-form characterization of the test statistic's asymptotic distribution is typically unavailable. Moreover, Model (1.1) can be regarded as a difficult ill-posed inverse problem. When the argument of the unknown functions are endogenous, it is difficult to propose valid uniform inference methods for both $\beta_0$ and $h_0(\cdot)$. To our best knowledge, the procedures to construct uniform confidence bands have remained an unsolved problem.[2]

To overcome the difficulties in uniform inference, under point identification, we first establish the limiting distribution of a sieve generalized method of moments (GMM) estimator for functionals of $\theta_0$ pointwisely. Our estimator is called the sieve generalized method of moments (SGMM) estimator. Then we employ empirical sieve processes to approximate the estimated nonparametric function uniformly over its domain. Although the empirical sieve processes belong to a non-Donsker class and are generally not weakly convergent, they can be strongly approximated by a sequence of Gaussian processes. As a consequence, we are able to do inference based on the the sequence of Gaussian processes. We propose sup-Wald, sup-Lagrange multiplier and sup-quasi-likelihood ratio statistics for testing restrictions uniformly over the support of nonparametric functionals. Similar to the pointwise inference problem, we show that the trinity of these three statistics hold in the sense that they can be strongly approximated by a sequence of "chi-squared processes" and are asymptotically equivalent. The critical values to be used for construction of uniform confidence bands can be simulated from the suprema of the sequence of chi-squared processes or obtained via valid multiplier bootstrap procedures.

As useful by-products of the results under point identification, we show that after rescaling at different rates, pointwisely, estimators of the finite-dimensional parameter $\beta_0$ and infinite-dimensional unknown function $h_0(\cdot)$ jointly converge to a Gaussian vector. Moreover, the parametric estimator and the estimator of the unknown function are asymptotically independent while

[2]After the first version of this paper was posted, we were aware of the related and ongoing work of Chen and Christensen (2014). They propose a procedure to construct uniform confidence bands for nonparametric IV models.

the parametric estimator achieves the semiparametric efficiency bound. For parametric and nonparametric functionals of interest, we establish that the trinity of Wald, quasi-likelihood ratio and Lagrange multiplier statistics from parametric GMM models can be extended to this more general semi-nonparametric setting. The three classes of test statistics are asymptotically equivalent and converge to a chi-squared distribution in the limit. The results are analogous to the ones obtained in a parametric GMM models (see, for example, McFadden and Newey, 1994), although we handle nonparametric functions in this study. The asymptotic independence simplifies the procedures to construct joint confidence intervals as they can be obtained from marginal confidence intervals.

The second challenge we face is that point identification can be difficult to attain for Model (1.1) (Newey and Powell (2003), Chen, Chernozhukov, Lee and Newey (2014)). In general, the rank condition is hard to verify in parametric conditional moment restriction models $E\left[\rho(Z,\beta)|X\right]$ when the generalized residual functions are nonlinear. This identification problem can be even more severe when we include unknown functions $h$ in models with conditional moment restrictions. Identification requires the instruments satisfying conditions stronger than rank conditions in the parametric case.

To consider inference with a possible lack of point identification, we propose methods for a class of conditional moment restriction models that are robust to partial identification. Under partial identification, there can be a set of parameters satisfying the moment conditions, so consistency and rate of convergence are measured by set distances based on suitable choices of norms. Our choice of norm under partial identification is based on how far $E\left[\rho(Z;\beta,h(\cdot)|X\right]$ is away from zero. With this norm, the set can be regarded as a sharp set of observationally equivalent parameters named *the identified set*. To consider inference for functional restrictions of the identified set, we focus on a class of moment condition models and functionals where there is a one-to-one mapping between the generalized residual functions and the functional restrictions. Then the set of functional restrictions can be regarded as an observationally equivalent set of functionals of the generalized residual functions. Therefore, we can utilize the generalized residual functions to distinguish parameters that satisfy both moment restrictions and functional restrictions from any other parameters. With this approach, we are able to test hypothesis of functional restrictions without knowing if they are point identified or not.

Under partial identification, the pointwise test statistic we consider is a quasi likelihood ratio statistic based on a sieve GMM criterion with a general weight matrix that does not depend on the parameters. When the parameters are point-identified, the quasi likelihood ratio statistic converges to a weighted chi-square distribution. When the parameters are not point-identified, the QLR statistic converges to the infimum of the square of a Gaussian process. Because the limiting distribution is not pivotal, we invert a multiplier bootstrap version of the statistic to obtain critical values. By inverting the test statistic, this inference procedure provides confidence regions for the functional of parameters. If the model is point identified, such confidence regions reduce to confidence intervals of functionals of $\theta_0$. For uniform inference over the support of the unknown function, we propose a sup-likelihood ratio statistic to test functional restrictions. The

test statistic may not have a limiting distribution. Instead, we employ a strong approximation approach, which provides a sequence of approximating distributions that necessarily adjusts with the sample size. This sequence of approximating distributions consists of a sequence of a "chi-square processes".[3]

It is also worth noting that the methods proposed here are computationally simple. In particular, all of our inference methods are based on a sieve generalized method of moments criterion. Once the unknown functions are replaced by their sieve approximations, the SGMM criterion effectively becomes a parametric one. Thus, the proposed methods are analogous to parametric GMM ones and are easy to compute.

We provide Monte Carlo evidence on the finite sample performance of our methods. In simulation studies we find that our methods deliver accurate coverage and relatively good power. We then apply our methods to the shape-invariant Engel curve system where total expenditure is endogenous that originated from Blundell, Chen and Kristensen (2007). By using the 1995 U.K. Family Expenditure Survey, we are able to construct confidence intervals and confidence bands for the Engel curves under point identification and confidence regions under partial identification. We formally confirm the findings of Blundell, Chen and Kristensen (2007) by revealing that lower-income people spend a larger proportion of their total expenditure on necessary goods such as food or fuel, while higher-income people spend proportionally less on necessary goods but more on leisure goods. Our empirical results are consistent with the predictions from consumption theory.

**Relationship to Literature**

This paper is related to several existing literatures. There has been extensive work on estimation and inference of semi-nonparametric models with moment restrictions under point identification. For instance, Newey and Powell (2003) propose a nonparametric two-stage least squares estimator based on series approximation and derive its consistency. Ai and Chen (2003) derive consistency and the rate of convergence of a sieve minimum distance (SMD) estimator and established the asymptotic distribution of the estimator of the parametric component $\beta_0$. The unknown functions are profiled out as infinite-dimensional nuisance parameters in Ai and Chen (2003). Hall and Horowitz (2005) propose two nonparametric estimation for nonparametric IV models based on orthogonal series and kernel methods. To deal with the ill-posed problem, Darolles, Fan, Florens and Renault (2010) propose a different consistent estimator based on Tikhonov regularization, and establish asymptotic properties of the estimated nonparametric instrumental regression function. And Horowitz and Lee (2012) construct uniform confidence bands for the nonparametric IV case

[3]Note that our results under partial identification are based on "pointwise" asymptotics in the sense that we only consider the case where the data generating process is fixed and we do not show asymptotic validity uniformly over data generating processes. The uniformity issue for the particular problems that this paper consider has not been addressed in the literature. Although it has been addressed for inference for functionals of parameters in other models with identification difficulties, for example, in parametric models with weak identification (e.g., Andrews, Moreira and Stock (2006), Andrews and Cheng (2012)) and in models defined by parametric moment inequalities (Bugni, Canay and Shi (2014)), among others.

by using properties of the unknown function such as monotonicity or smoothness to interpolate over a finite grid of points and by allowing the number of grid points to go to infinity.

The papers that are most closely related to this one are Chen and Pouzo (2009, 2012, 2014). Chen and Pouzo (2009) establish the semiparametric efficient estimation of $\beta_0$ for model (1.1) with possibly nonsmooth residuals. Their results depend on the consistency and convergence rates of the nonparametric estimation of $h_0$ based on Chen and Pouzo (2012). And Chen and Pouzo (2014) provide inference methods for functionals of both $\beta_0$ and $h_0$ based on the SMD criterion. Our asymptotic results under point identification complements the analysis in Chen and Pouzo (2014) by extending the pointwise inference to uniform inference methods for the functional of parameters. We also establish that the SGMM estimator is asymptotically equivalent to the SMD estimator; and the parametric and nonparametric estimators are asymptotically independent while the parametric estimator achieves the semiparametric efficiency bound. Note that in this paper we focus on the case where the residual functions are smooth while Chen and Pouzo (2009, 2012, 2014) allow residuals to be nonsmooth.

Our uniform inference methods are related to recent seminal work on exploring the use of strong approximation methods to derive the limiting distribution of nonparametric estimators in econometrics. Related ideas appear in estimation and inference of a variety of nonparametric (quantile) regression-type models as in Belloni, Chernozhukov and Fernández-Val (2011), Chandrasekhar, Chernozhukov, Molinari and Schrimpf (2012), Belloni, Chernozhukov, Chetverikov and Kato (2013) (henceforth, BCCK), Chernozhukov, Chetverikov, and Kato (2013) and Chernozhukov, Lee and Rosen (2013). We contribute to this literature by providing uniform inference methods for general models of the form (1.1), allowing the argument of the unknown functions to be an endogenous regressor.

Our new results also contribute to the recent literature on inference for semiparametric and nonparametric models by considering the use of likelihood ratio type statistics. For instance, Murphy and van der Vaart (2000) propose standard likelihood ratio statistics for semiparametric models and extend the classical Wilk's theorem to infinite dimensional parameter spaces. Shen and Shi (2005) consider a sieve likelihood ratio statistic and provide asymptotic distribution of sieve likelihood ratio statistics for regular functionals. Based on Shen and Shi (2005), Chen and Pouzo (2009) provide limiting distribution for a sieve quasi-likelihood ratio (QLR) statistic for the finite-dimensional parameters in semiparametric conditional moment models and Chen and Pouzo (2014) establish the pointwise limiting distribution of a sieve QLR for inference on functionals of semi/nonparametric conditional moment restrictions regardless of whether functionals are $\sqrt{n}-$estimable or not. Chen, Tamer and Torgovitsky (2011) provide methods for inference in semiparametric likelihood models with partial identification. They focus on inference on the finite-dimensional parameters $\beta$. We add new results in this literature in the following aspects. First, we provide a quasi-likelihood ratio statistics for joint inference for both finite-dimensional parameters and unknown functions in models with conditional moment restrictions; second, we show that QLR is robust to partial identification for a class of moment restrictions models; finally,

we extend the QLR statistic to a sup-QLR statistic so we can conduct inference uniformly over the support of functionals of the parameters.

The literature on nonparametric IV models has achieved point identification from conditional moments by imposing completeness conditions, e.g., Newey and Powell (2003), Chernozhukov and Hansen (2005), Hall and Horowitz (2005), Blundell, Chen and Kristensen (2007), Chernozhukov, Imbens and Newey (2007), Chen, Chernozhukov, Lee and Newey (2014). These completeness conditions can be regarded as the nonparametric analog of the classical rank conditions in parametric models. They have been a central focus in recent studies (e.g., Andrews (2011), d'Haultfoeuille, 2011)). In particular, Canay, Santos and Shaikh (2013) have examined hypothesis testing problems for completeness conditions. They conclude that no nontrivial tests for these hypothesis testing problems exist. We complement this literature by developing methods that are robust to partial identification.

An extensive literature on inference in a variety of partially identified models has been developed over the past decade, including Imbens and Manski (2004), Chernozhukov, Hong, and Tamer (2007), Andrews and Jia (2008), Beresteanu and Molinari (2008), Romano and Shaikh (2008, 2010), Stoye (2009), Andrews and Guggenberger (2009), Andrews and Soares (2010), Bontemps, Magnac, and Maurin (2010), Bugni (2010), Canay (2010), Galichon and Henry (2011), Chen, Tamer and Torgovitsky (2011), Freyberger and Horowitz (2012), Santos (2012), Andrews and Shi (2013), Chernozhukov, Lee and Rosen (2013), Canay, Bugni and Shi (2014) and other papers referenced therein. We add new results to this literature by establishing the validity of the quasi-likelihood ratio test under partial identification for a class of moment equality models. Moreover, most works in partially identified moment condition models have focused on a fully parametric setting while this paper considers the extension to moment equalities with unknown functions.

In a nonparametric IV setting, Santos (2012) has proposed pointwise inference methods for hypothesis testing under partial identification. Hong (2013) has extended the methods in Santos (2012) to conditional moment restriction models. The test statistics used in these papers have non-standard limiting behaviors and can be challenging to approximate even with bootstrap methods (see Grundl and Zhu (2014)). In this work, we take a different approach to focus on the properties of the QLR statistic, which converges to the infimum of a chi-square process under partial identification that can be reduced to a chi-squared distribution if the model is point-identified.

We are also aware of independently and concurrently work by Chen, Pouzo and Tamer (2011) (CPT, henceforth), which studies inference on Model (1.1) with partial identification based on a minimum distance criterion. Based on the presentation slides of CPT available to us, we understand that under partial identification, CPT studies inference based on a sieve minimum distance criterion while our criterions are based on a sieve GMM; CPT is primarily concerned with pointwise inference while we are interested in both pointwise and uniform inference over the support of the unknown functions. Moreover, in the partially identified case, the present paper focus on a class of moment condition models and the parameters where there is a one-to-one mapping between the generalized residual functions and the functionals of interest.

**Notation and Definitions**

For any column vector $a$, we use $a'$ to denote its transpose, $||a||_E$ to denote the Euclidean norm and for a function $a(\cdot)$ with domain $x$, we use $||a||_\infty$ to denote the sup-norm $\sup_{x\in\mathcal{X}}|a(x)|$. Let $\mathbf{H}=\mathbf{H}^1\times...\times\mathbf{H}^L$ be a separable Banach space with norm $||\cdot||_{\mathbf{H}}$. Let $\mathcal{H}=\mathcal{H}^1\times...\times\mathcal{H}^L$ be a closed, nonempty infinite-dimensional subset of $\mathbf{H}$. Let $\Theta=\mathcal{B}\times\mathcal{H}\subseteq\mathbf{R}^{d_\beta}\times\mathbf{H}$ be endowed with a (strong) norm $||\theta||_s=||\beta||_E+||h||_{\mathbf{H}}$. For two Banach spaces $\mathbf{H}_1$ and $\mathbf{H}_2$, for any mapping $\Gamma$: $\mathbf{H}_1\to\mathbf{H}_2$, let $\frac{d\Gamma(\theta_0)}{d\theta}[\delta]=\left.\frac{\partial\Gamma(\theta_0+t\delta)}{\partial t}\right|_{t=0}$ be the pathwise derivative at $\theta_0$ in the direction $\delta\in\mathbf{H}_1$. For two random variables $X_1$ and $X_2$, let $X_1\overset{d}{=}X_2$ if they have identical probability distribution. For two sequences $\{a_n\}_{n=1}^\infty$ and $\{b_n\}_{n=1}^\infty$, we use the notation $a_n\lesssim b_n$ to denote $a_n\leq cb_n$ for some constant $c>0$ that does not depend on $n$; and $a_n\simeq b_n$ means that $c_1a_n\leq b_n\leq c_2a_n$ for two constants $0<c_1\leq c_2<\infty$. If $\{a_n\}_{n=1}^\infty$ and $\{b_n\}_{n=1}^\infty$ are random sequences, we use $a_n\lesssim_p b_n$ or $a_n=O_p(b_n)$ to denote $\lim_{c\to\infty}\limsup_n\Pr(a_n/b_n>c)=0$; and $a_n=o_p(b_n)$ means for all $\varepsilon>0$, $\lim_{n\to\infty}P(a_n/b_n>\varepsilon)=0$. We use the notation $a\vee b=\max\{a,b\}$ and $a\wedge b=\min\{a,b\}$. We use $\Rightarrow$ to denote weak convergence. Let $\theta=(\beta,h)\in\Theta\equiv\mathcal{B}\times\mathcal{H}$. Let $\mathcal{H}_n=\mathcal{H}_n^1\times\cdots\times\mathcal{H}_n^L$ be a sequence of approximating spaces to $\mathcal{H}$, denoted by a sieve. Let $\Theta_n=\mathcal{B}\times\mathcal{H}_n$. Let $\Theta_n=\mathcal{B}\times\mathcal{H}_n$, $\Pi_n\theta_n=(\beta',\Pi_nh)\in\Theta_n$, $m(X,\theta)=E[\rho(Z,\theta)|X]$. In what follows, We use $X$ to be a vector of instrumental variables with support $\mathcal{X}$, where $\mathcal{X}$ is a compact subset of $\mathbf{R}^{d_x}$. We use $Y$ to be a vector of endogenous variables with support $\mathcal{Y}$, where $\mathcal{Y}$ is a subset of $\mathbf{R}^{d_y}$. Let $\{Z_i=(Y_i',X_i')\}_{i=1}^n$ be a random sample from the distribution of $Z=(Y',X')'$ with support $\mathcal{Y}\times\mathcal{X}_z$, and $\mathcal{X}_z\subseteq\mathcal{X}$. Let the domain of functional of interest as $w\in\mathcal{W}$, where $\mathcal{W}$ is a compact subspace of $\mathbf{R}^{d_w}\subseteq\mathcal{Z}$. Let $\phi(\theta):\Theta\to\mathbf{R}^{d_\phi}$ be a functional of $\theta$. Let $\phi(\theta)[w]$ be a functional evaluated at $w$.

**Structure of the Paper**

The remainder of the paper is organized as follows. Section 2 provides examples of interesting and the introduction of sieve approximations. Section 3 develops the asymptotic theory under point identification. In section 4, we relax the identification assumption and develop asymptotic inference theory under partial identification. Small sample performance is analyzed in a Monte Carlo study in Section 5. Section 6 provides an empirical example. Section 7 concludes. Proofs are in the Appendix. The online supplement contains further appendices.

## 2 MOTIVATING EXAMPLES AND SIEVE APPROXIMATIONS

In this section, we briefly introduce examples of semi-nonparametric conditional moment restriction models from the literature that we use to illustrate our methods and results. Then we briefly introduce the sieve methodology.

**Example 2.1** (Nonparametric IV model, Newey and Powell, 2003).

Consider a nonparametric IV model $Y_1 = h(Y_2)+e$, where $E[e|X] = 0$. Then given instruments $X$, the conditional moment restriction can be expressed as

$$E\left[Y_1 - h(Y_2)|X\right] = 0, \tag{2.1}$$

where $Y_1$ is a scalar and the dependent variable, $Y_2$ is the endogenous regressor, $Z = (Y_1, Y_2')$ and $X$ is the instrument. The residual function $\rho(Z, \theta) = Y_1 - h(Y_2)$. We are interested in testing and constructing pointwise confidence intervals and uniform confidence bands of the unknown function $h(\cdot)$ and its functionals. The function $h(\cdot)$ may or may not be point identified.

**Example 2.2** (Partially linear model with a known link function and an endogenous nonparametric part, Ai and Chen (2003)).

Consider a partially linear model with a known link function. Let the nonparametric part have endogenous regressors as an argument. Consider the residual function $\rho(Z, \theta) = Y_1 - G(X_1\beta + h(Y_2))$ with conditional moment restriction

$$E\left[\rho(Z, \theta)|X\right] = E\left[Y_1 - G(X_1\beta + h(Y_2))|X\right] = 0 \tag{2.2}$$

where $\theta = (\beta', h(\cdot))'$, $G(\cdot)$ is a known function, $Y_1$ is a scalar, $Z = (Y_1, X_1', Y_2')'$, $X = (X_1', X_2')'$. Suppose $\dim(X_2) = \dim(Y_2) = d$, $\dim(X_1) = d_\beta$ and $\dim(X) = d + d_\beta$. We provide pointwise and uniform joint inference methods for the finite-dimensional parameter $\beta$ and the unknown function $h$ or functionals of both $\beta$ and $h$ when $\beta$ and $h$ are both point-identified. If we restrict $G(\cdot)$ to be a strictly monotone function or an identity function, our inference methods provide valid pointwise confidence intervals and uniform confidence bands regardless of whether parameters are identified or not.

**Example 2.3** (Engel Curves, Blundell, Chen and Kristensen, 2007).

Consider an Engel curve with unknown shape where characteristic adjustments and endogeneity is allowed. Then the model is of the form of (1.1) with the residual function

$$\rho_l(Z_i, \theta) = Y_{1il} - h_l\left(Y_{2i} - g(X_{1i}'\beta_1)\right) - X_{1i}'\beta_{2,l}, \tag{2.3}$$

where $g(\cdot)$ is known, $Y_{1il}$ are observations on the budget share of good $l = 1, \ldots L \geq 1$ for each household $i$ facing the same relative prices, $Y_{2i}$ is log of total expenditure and $X_{1i}$ is a vector of household composition variables. The gross earning of the head of household is the instrument denoted $X_{2i}$. We shall discuss the details of this example in Section 6.

**Examples of Functionals of Interest**

In many applications, we are not only interested in $\theta$ itself, but also in functionals of $\theta$ denoted by $\phi(\theta)$. We roughly divide the functionals into two groups: *regular* ($\sqrt{n}-$estimable) functionals

and *irregular* (slower than $\sqrt{n}-$estimable) ones. For the rest of the paper, we use $w$ to represent the domain of functional $\phi(\theta)$, where $w \in \mathcal{W} \subseteq \mathcal{Z}$. Examples of regular functionals of interest include the parametric component $\beta$ and $E[h(w)]$, while examples of irregular functionals include the function evaluated at a single point $\phi(h)[w] = h(w)$, the partial derivative $\phi(h)[w] = \partial_w h(w)$ and the conditional average partial derivative $\phi(h)[w_{-j}] = \int \partial_{w_j} h(w) df(w_j|w_{-j})$.

**The Method of Sieves**

The method of sieves has been a popular procedure for estimating semiparametric and nonparametric models in recent years. This paper considers two classes of sieve approximations. One is for the unknown functions, the other is for the conditional moment restrictions.

First, we approximate the unknown functions by sieves. Specifically, we define a sieve space, $\mathcal{H}_n$, to be a sequence of approximating spaces to the parameter space $\mathcal{H}$ for unknown functions. A particular convenient class of sieves are linear in the parameters such that

$$\mathcal{H}_n = \left\{ h \in \mathcal{H} : h(\cdot) = \sum_{k=1}^{k_n} p_k(\cdot)'\eta_{nk} = p^{k_n}(\cdot)'\eta_n \right\} \tag{2.4}$$

where $\{p_k\}_{k=1}^{\infty}$ is a sequence of known basis functions of a Banach space. Let $\Theta_n = \mathcal{B} \times \mathcal{H}_n$. In this paper, we use slowly growing finite-dimensional sieves (i.e., $k_n \to \infty$, $k_n/n \to 0$).[4] The approximation is *dense* in the sense that for any $\theta \in \Theta$ there exists $\Pi_n\theta \in \Theta_n$ such that $\|\Pi_n\theta - \theta\| \to 0$ as $n \to \infty$. Common choices of basis functions include, for example, polynomial series expansions or splines. Discussions of choices and properties of different basis functions can be found, for example, in Chen (2007) and Hansen (2014).

Second, we follow Donald, Imbens and Newey (2003) to characterize the conditional moment restriction model (1.1) as an infinite number of appropriate unconditional moment restrictions formed from the product of the residual functions with sieve functions of the instrumental variable $X$. We use general series functions, such as polynomials or splines, to form the unconditional moments that grow in number with the sample size such that

$$E\left[\rho(Z,\theta) \otimes q^{s_n}(X)\right] = 0 \tag{2.5}$$

where $q^{s_n}(X) = (q_1(X), \ldots, q_{s_n}(X))'$, $\{q_s(\cdot)\}_{s=1}^{\infty}$ is a sequence of known basis functions that approximates any square integrable functions of $X$ as $s_n \to \infty$ slowly when $n \to \infty$, $Z' \equiv (Y', X_z')$, $Y$ is a vector of endogenous variables, $X_z$ is a subset of $X$, $X$ is a vector of conditioning variables, $\beta = (\beta_1, \ldots, \beta_{d_\beta})' \in \mathcal{B}$ being a $d_\beta-$vector of Euclidean parameters (parametric components), $h = (h_1(\cdot), \ldots, h_L(\cdot)) \in \mathcal{H}$ being a $L-$vector valued functions (nonparametric components) and $\theta \equiv (\beta', h(\cdot))$ are parameters of interest. We write $Q = (q^{s_n}(X_1), \ldots, q^{s_n}(X_n))'$, and $(Q'Q)^-$ is the

[4] Chen and Pouzo (2012) propose a sieve minimum distance estimator that allows for large dimensional sieves (i.e., $k_n/n \to$ const. $> 0$) with a general class of lower semicompact and/or convex penalties. This case will not be emphasized in this paper.

generalized inverse of the matrix $Q'Q$. We impose the following assumption on the basis functions.

**Assumption 2.1.** *For all $s_n$, $E\left[q^{s_n}(X)'q^{s_n}(X)\right]$ is finite. And for any function $a(x)$ with $E\left[a(X)^2\right] < \infty$, there are $s_n-$vectors $\pi_n$ such that $E\left[\left\{a(X) - q^{s_n}(X)'\pi_{ns}\right\}^2\right] \to 0$ as $s_n \to \infty$.*

Assumption 2.1 is the same as Assumption 1 in Donald, Imbens and Newey (2003) where they consider a parametric model with conditional moment restrictions. Assumption 2.1 is a fundamental condition on the sequence $q^{s_n}(X)$ and the distribution of $X$.

**Lemma 2.1.** *(Lemma 2.1 in Donald, Imbens and Newey, 2003) Suppose that Assumption 2.1 is satisfied and for any $\theta \in \Theta$, $E[\rho(Z,\theta)'\rho(Z,\theta)]$ is finite. If $E\left[\rho(Z,\theta)|X\right] = 0$ is satisfied, then $E\left[\rho(Z,\theta) \otimes q^{s_n}(X)\right] = 0$ for all $s_n$; if $E\left[\rho(Z,\theta)|X\right] \neq 0$, then $E\left[\rho(Z,\theta) \otimes q^{s_n}(X)\right] \neq 0$ for large enough $s_n$.*

Lemma 2.1 has been shown in Donald, Imbens and Newey (2003). This lemma is crucial for showing that an efficient estimator under the conditional moment restrictions can be derived from a sequence of unconditional moment restrictions. When $s_n$ grows with the sample size, information from the conditional moment restrictions is eventually fully accounted for.[5]

Based on Lemma 2.1, we will propose the population and sample criterion functions we use under point identification and partial identification in Section 3 and Section 4, respectively.

## 3 ASYMPTOTIC RESULTS UNDER POINT IDENTIFICATION

In this section we present some new asymptotic results for semi-nonparametric conditional moment models under point identification. We start with imposing the following point identification assumption. It will be relaxed in Section 4, when we consider the partially identified case.

**Assumption 3.1.** *$\theta_0$ is the only $\theta \in \Theta$ satisfying $E\left[\rho(Z,\theta)|X\right] = 0$ and $\theta_0 \in int(\Theta)$.*

Assumption 3.1 specifies that there exists a unique $\theta_0$ satisfying the conditional moment restriction. Under point identification, we proceed by describing the criterion function proposed for estimation, followed by results for consistency, the pointwise limiting theory, the uniform limiting theory and the inference methods under point identification. Some technical assumptions are presented in Appendix A.

We estimate the parameter vectors using a (penalized) sieve generalized method of moments (SGMM) criterion (Chen (2007)). The SGMM criterion is a semi/nonparametric version of the GMM criterion proposed by Donald, Imbens and Newey (2003) for conditional moment restrictions. We suggest to use SGMM because it is easy to use and is analogous to parametric GMM. Alternative choices of the criterion functions include, for example, the sieve minimum distance

[5] Bierens (1990) provides a different approach to characterize the conditional moment models as an infinite number of appropriate unconditional moment restrictions. Applications of Bierens'-type transformations can be seen in, for example, Santos (2012) and Andrews and Shi (2013).

estimator proposed by Ai and Chen (2003), the sieve conditional empirical likelihood estimator proposed by Otsu (2011) based on Kitamura, Tripathi and Ahn (2004)'s conditional empirical likelihood and the sieve generalized empirical likelihood proposed by Sueishi (2012) based on Newey and Smith (2004)'s generalized empirical likelihood.

Let $g_i(\theta) = \rho(Z_i, \theta) \otimes q^{s_n}(X_i)$ and $\hat{g}(\theta) = n^{-1} \sum_{i=1}^n g_i(\theta)$. We define the two-step (penalized) sieve GMM (SGMM) estimator $\hat{\theta}_n \in \Theta_n$ as the minimizer of the following criterion[6]

$$\hat{L}_n(\theta) = \hat{g}(\theta)' \left( \frac{1}{n} \sum_{i=1}^n g_i\left(\bar{\theta}_n\right) g_i\left(\bar{\theta}_n\right)' \right)^{-1} \hat{g}\left(\theta\right) + \lambda_n \text{Pen}(h) \tag{3.1}$$

where $\bar{\theta}_n$ is a preliminary estimator.[7]

The corresponding population criterion is of the form

$$L(\theta) = E\left[m(X,\theta)'\Sigma(X,\theta_0)^{-1}m(X,\theta)\right] \equiv E\left[||m(X,\theta)||^2_{\Sigma_0^{-1}}\right], \tag{3.2}$$

where $\Sigma_0(X) \equiv \Sigma(X, \theta_0) \equiv E\left[\rho(Z,\theta_0)\rho(Z,\theta_0)'|X\right]$. Clearly, by Assumption 3.1, $L(\theta)$ is uniquely minimized over $\Theta$ at $\theta_0$. Similar to the ideas in parametric GMM, we show later that the two-step sieve GMM estimator $\hat{\theta}_n$ is an efficient estimator.

Note that after the space of unknown functions $\mathcal{H}$ is approximated by its sieve space $\mathcal{H}_n$, the minimum is computed only over $\Theta_n$. Then the SGMM criterion effectively becomes a parametric one and is easy to compute.

We next introduce a norm $||\cdot||$ that will be repeatedly used later in this section. Let

$$\Theta_{os} \subset \{\theta \in \Theta : ||\theta - \theta_0||_s < K, \text{Pen}(\theta) < K\} \tag{3.3}$$

be a convex, $||\cdot||_s-$neighborhood around $\theta_0$. Let $\Theta_{osn}$ be the sieve space of $\Theta_{os}$. For any $\theta_1$, $\theta_2 \in \Theta_{os}$, recall the pathwise derivative notation defined in Section 1. We define the norm $\|\cdot\|$ as

$$\|\theta_1 - \theta_2\|^2 \equiv E\left[ \left\{ \frac{dm(X,\theta_0)}{d\theta}[\theta_1 - \theta_2] \right\}' \Sigma_0(X)^{-1} \left\{ \frac{dm(X,\theta_0)}{d\theta}[\theta_1 - \theta_2] \right\} \right]^2. \tag{3.4}$$

This norm was introduced by Ai and Chen (2003) and is an extension of the Fisher norm to conditional moment restriction models. It is motivated by the objective function of the SGMM criterion (3.2). The convergence rate and asymptotic distribution of $\hat{\theta}_n$ under point identification will be derived under the norm $||\cdot||$.

We now provide some conditions that are useful for showing consistency. These conditions are

[6]Note that the penalization needs not be used if we follow the approach of Ai and Chen (2003), Newey and Powell (2003) and Santos (2012) and solve the ill-posed inverse problem by obtaining compactness through smoothness assumption on the unknown functions. Then, (3.1) is the same as the parametric GMM criterion proposed by Donald, Imbens and Newey (2003). Common choices of Pen$(h)$ include, for example, $\text{Pen}(h) = ||h||^2_{L^2} + ||\bigtriangledown h||^2_{L^2}$.

[7]The preliminary estimator $\bar{\theta}_n$ is the minimizer of 3.1 by replacing $\left(\frac{1}{n}\sum_{i=1}^n g_i\left(\bar{\theta}_n\right) g_i\left(\bar{\theta}_n\right)'\right)^{-1}$ with a general weight matrix $\hat{W} = \left(\frac{1}{n}\sum_{i=1}^n \Sigma(X_i) \otimes q^{s_n}(X_i) q^{s_n}(X_i)'\right)^{-1}$.

similar to the ones obtained in Chen and Pouzo (2014). Our notation follows that of Chen and Pouzo (2014).

**Assumption 3.2.** *(i) The data* $\{(Y_i', X_i')_{i=1}^n\}$ *are i.i.d; (ii)* $\mathcal{Y} \times \mathcal{X}$ *is compact; (iii) the density of* $X$ *is bounded above and away from zero on* $\mathcal{X}$*.*

**Assumption 3.3.** *Let* $\Theta_n \equiv \mathcal{B} \times \mathcal{H}_n$*, where* $\mathcal{B}$ *is compact subset in* $\mathbf{R}^{d_\beta}$*, and* $\{\mathcal{H}_n\}_{n=1}^\infty$ *is a sequence of non-empty closed subsets of a separable infinite dimensional Banach space* $(\mathbf{H}, ||\cdot||_{\mathbf{H}})$ *such that for all* $\theta \in \Theta$ *there exists* $\Pi_n \theta = (\beta', \Pi_n h) \in \Theta_n$ *satisfying* $||\Pi_n \theta - \theta||_s = o(1)$*.*

**Assumption 3.4.** *(Penalty) (i)* $\lambda_n > 0$*,* $\lambda_n = o(1)$*; (ii)* $|Pen(\Pi_n h_0) - Pen(h_0)| = O(1)$ *with* $Pen(h_0) < \infty$*; (iii) Pen:* $(\mathcal{H}, ||\cdot||_{\mathbf{H}}) \to [0, \infty)$ *is lower semicompact, namely for all constants* $K$*,* $\{h \in \mathcal{H} : Pen(h) \leq K\}$ *is a compact subset in* $(\mathcal{H}, ||\cdot||_{\mathbf{H}})$*.*

**Assumption 3.5.** *(i)* $L : (\Theta, \|\cdot\|_s) \to [0, \infty)$ *is lower semicontinuous,* $L(\Pi_n \theta_0) = o(1)$*; (ii)* $\Sigma_0(X)$ *is positive definite a.s.* $X$ *and its largest and smallest eigenvalues are finite and positive.*

**Assumption 3.6.** *Uniformly over* $\Theta_{osn}$*,* $\hat{L}_n(\theta) \gtrsim L(\theta) - O_p(\bar{\varrho}_n^2)$ *for* $\bar{\varrho}_n = o_p(n^{-1/4})$*.*

Assumptions 3.1-3.6 are standard in the literature (see, e.g., Chen and Pouzo, 2014). These assumptions impose conditions on model identification, the distribution of the data, sieve space approximation, the behavior of the penalty term and the population criterion function, respectively. Assumption 3.6 can be verified by Lemma C.1 in the Supplemental Appendix. These conditions imply consistency of $\hat{\theta}_n$.

**Lemma 3.1.** *Let* $\hat{\theta}_n$ *be the two-step SGMM estimator. Let*

$$\varsigma_n \equiv \sup_{\theta \in \Theta_{osn} : \|\theta - \Pi_n \theta_0\| \neq 0} \frac{\|\theta - \Pi_n \theta_0\|_s}{\|\theta - \Pi_n \theta_0\|}$$

*be the sieve measure of local ill-posedness. Suppose that Assumptions 2.1, 3.2-3.6 and Assumptions A.1-A.5 hold. For* $\varrho_n \leq \bar{\varrho}_n$*, where* $\bar{\varrho}_n$ *is introduced in Assumption 3.6, we have*

*(i)* $||\hat{\theta}_n - \theta_0|| = O_p(\varrho_n)$*, and (ii)* $||\hat{\theta}_n - \theta_0||_s = O_p(\varsigma_n \varrho_n + \|\theta_0 - \Pi_n \theta_0\|_s)$*.*

**Remark.** Lemma 3.1 is a minor modification of the consistency results obtained in Chen and Pouzo (2012). We present it only for completeness of our discussion. Given the consistency result, with probability approaching one, the PSGMM estimator belongs to a $||\cdot||_s-$neighborhood around $\theta_0$. Common choices of $||\cdot||_s$ include $||\cdot||_{L^2}$ or $||\cdot||_\infty$ for $||\cdot||_{\mathbf{H}}$. Blundell, Chen and Kristensen (2007) and Chen and Pouzo (2012) show that $\hat{\theta}_n$ is a consistent estimator of $\theta_0$ in both $||\cdot||_{L^2}$ and $||\cdot||_\infty$ norms. Recently, Chen and Christensen (2014) give a general upper bound on the uniform convergence rate for NPIV estimators when $||\cdot||_s = ||\cdot||_{\mathbf{H}} = ||\cdot||_\infty$. □

### 3.1 Pointwise Limiting Theory

Based on the consistency results above, we now present the pointwise asymptotic distribution of functionals of $\hat{\theta}_n$ (denoted by $\phi(\hat{\theta}_n)$) and the (joint) pointwise asymptotic distribution for $\phi(\hat{\theta}_n) = (\hat{\beta}_n, \phi(\hat{h}_n))$. Lemma 3.1 allows us to focus on a shrinking neighborhood of $\theta_0$. Let $\varrho_{sn} = \varsigma_n \varrho_n + ||\theta_0 - \Pi_n \theta_0||_s$, we derive our limiting theory by focusing on the local parameter space $\mathcal{N}_{os}$ and its sieve space $\mathcal{N}_{osn}$, where for a large $K < \infty$,

$$\mathcal{N}_{os} \equiv \{\theta \in \Theta : ||\theta - \theta_0|| \leq \varrho_n \log\log n, \|\theta - \theta_0\|_s \leq \varrho_{sn} \log\log n, \text{Pen}(\theta) \leq K\} \tag{3.5}$$

and $\mathcal{N}_{osn} \equiv \Theta_n \cap \mathcal{N}_{os}$. Note that with probability approaching one, by Lemma 3.1, $\hat{\theta}_n \in \mathcal{N}_{osn} \subseteq \mathcal{N}_{os}$.

Let $\phi(\theta)$ denoted the functionals of interest on the parameter $\theta = (\beta, h(\cdot))$. These functionals are classified as *regular* or *irregular* ones. Heuristically speaking, regular functionals are the ones that can be estimated at $\sqrt{n}-$rate (e.g., the parametric component $\beta$), while irregular functionals are the ones that are estimated at slower than $\sqrt{n}-$rate (e.g., the unknown function $h(\cdot)$ evaluated at a point $w$).

To be more formal, let $||\cdot||$ be the norm defined in (3.4) and $\bar{\Delta}$ be the closed linear span of $\Theta_{os}/\{\theta_0\}$ under the norm $\|\cdot\|$ (defined in 3.4), then $(\bar{\Delta}, \|\cdot\|)$ is an infinite-dimensional Hilbert space with the following inner product:

$$\langle \delta_1, \delta_2 \rangle = E\left[\left\{\frac{dm(X,\theta_0)}{d\theta}[\delta_1]\right\}' \Sigma_0(X)^{-1} \left\{\frac{dm(X,\theta_0)}{d\theta}[\delta_2]\right\}\right]$$

for $\delta_1, \delta_2 \in \bar{\Delta}$.

Let $\theta_{0n} \in \Theta_n$ be such that $||\theta_{0n} - \theta_0|| = \min_{\theta \in \Theta_n} ||\theta - \theta_0||$. Let $\bar{\Delta}_n$ be the sieve approximation of $\bar{\Delta}$, which is a finite-dimensional space under $||\cdot||$ and $\bar{\Delta}_n$ is the closed linear span of $\Theta_{osn}/\{\theta_{0n}\}$.

We say the functional $\phi(\cdot) : \Theta \to \mathbf{R}$ is regular (at $\theta = \theta_0$) if $\frac{d\phi(\theta_0)}{d\theta}[\cdot]$ is bounded on the infinite dimensional Hilbert space $\bar{\Delta}$, i.e.,

$$\sup_{\delta \in \bar{\Delta}, \delta \neq 0} \left\{\left|\frac{d\phi(\theta_0)}{d\theta}[\delta]\right| / ||\delta||\right\} < \infty.$$

Then the Riesz representation theorem implies that there is a Riesz representer $\delta^* \in \bar{\Delta}$ of the linear functional $\frac{d\phi(\theta_0)}{d\theta}[\cdot]$ on $\left(\bar{\Delta}, ||\cdot||\right)$ such that

$$\frac{d\phi(\theta_0)}{d\theta}[\delta] = \langle \delta, \delta^* \rangle \text{ for all } \delta \in \bar{\Delta}$$

and

$$\frac{d\phi(\theta_0)}{d\theta}[\delta^*] = \|\delta^*\|^2 = \sup_{\delta \in \bar{\Delta}, \delta \neq 0} \left|\frac{d\phi(\theta_0)}{d\theta}[\delta]\right|^2 / ||\delta||^2 < \infty. \tag{3.6}$$

We say the functional $\phi(\cdot)$ is irregular (at $\theta = \theta_0$) if $\frac{d\phi(\theta_0)}{d\theta}[\cdot]$ is unbounded on $\bar{\Delta}$, i.e.,

$$\sup_{\delta \in \bar{\Delta}, \delta \neq 0} \left\{ \left| \frac{d\phi(\theta_0)}{d\theta}[\delta] \right| / ||\delta|| \right\} = \infty. \tag{3.7}$$

Then there is a Riesz representer $\delta_n^* \in \bar{\Delta}_n$ such that

$$\frac{d\phi(\theta_0)}{d\theta}[\delta_n] = \langle \delta_n^*, \delta_n \rangle \text{ for all } \delta_n \in \bar{\Delta}_n$$

and

$$\frac{d\phi(\theta_0)}{d\theta}[\delta_n^*] = \|\delta_n^*\|^2 = \sup_{\delta_n \in \bar{\Delta}_n, \delta \neq 0} \left| \frac{d\phi(\theta_0)}{d\theta}[\delta_n] \right|^2 / ||\delta_n||^2 < \infty. \tag{3.8}$$

We call $\delta_n^*$ the (empirical) Riesz representer of the functional $\frac{\partial \phi(\theta_0)}{\partial \theta}[\cdot]$ on $\bar{\Delta}_n$.

We emphasize that the sieve Riesz representation of the linear functional $\frac{\partial \phi(\theta_0)}{\partial \theta}[\cdot]$ on $\bar{\Delta}_n$ always exists no matter whether $\frac{\partial \phi(\theta_0)}{\partial \theta}[\cdot]$ is bounded on the infinite dimensional space $\bar{\Delta}$ or not because any linear functional on a finite dimensional Hilbert space is bounded. The distinctive properties of regular and irregular functionals impose different technical challenges in deriving the *joint* distribution of parametric and nonparametric components of the parameters.

The estimator of $\phi(\theta_0)$ is $\phi(\hat{\theta}_n)$, where $\hat{\theta}_n$ is the two-step SGMM estimator. Without loss of generality, we assume the basis functions used to approximate $\phi(\hat{\theta}_n)$ are the same as the basis functions used to approximate the unknown functions $h(\cdot)$.[8]

Since the parameter $\theta$ includes the unknown function $h(\cdot)$, it is difficult to derive the asymptotic distribution of $\phi(\hat{\theta}_n)$ by adopting the usual approach that is based on the first-order condition for $\hat{\theta}_n$ from minimizing the SGMM criterion. Instead, we follow the Riesz representation approach of Chen and Pouzo (2014). Specifically, we provide a representation of the functional of interest and establish the asymptotic normality of the plug-in estimator $\phi(\hat{\theta}_n)$ based on such a representation. The following condition is needed.

**Assumption 3.7.** *(i)* $\frac{d\phi(\theta_0)}{d\theta}[\delta] : \bar{\Delta} \to \mathbf{R}$ *is a linear functional and is non-zero; (ii) for* $K_n = \log\log n$ *and* $\varrho_n$ *defined in (3.5), let* $\mathcal{T}_n \equiv \left\{ t \in \mathbf{R} : |t| \lesssim K_n^2 \varrho_n \right\}$ *and* $u_n^* = \delta_n^* / ||\delta_n^*||$*, then*

$$\sup_{(\theta,t) \in \mathcal{N}_{osn} \times \mathcal{T}_n} \left| \phi(\theta + t u_n^*) - \phi(\theta_0) - \frac{d\phi(\theta_0)}{d\theta}[\theta + t u_n^* - \theta_0] \right| / \|\delta_n^*\| = o(n^{-1/2});$$

*(iii) either (a) or (b) holds: (a)* $\|\delta_n^*\| \to \infty$ *and* $\left| \frac{d\phi(\theta_0)}{d\theta}[\theta_{0n} - \theta_0] \right| / \|\delta_n^*\| = o\left(n^{-1/2}\right)$*; (b)* $\|\delta_n^*\| \to \|\delta^*\| < \infty$ *and* $\|\delta^* - \delta_n^*\| \times \|\theta_{0n} - \theta_0\| = o(n^{-1/2})$*; (iv)* $\bar{\Delta}_n$ *is dense in* $(\bar{\Delta}, \|\cdot\|)$.

Assumption 3.7 is similar to Assumption 3.5 in Chen and Pouzo. Assumption 3.7 (ii) implies that the linear expansion error of functional $\phi(\theta)$ is relatively small compared to the variance $||\delta_n^*||$

[8] For ease of exposition, we consider univariate $\phi(\theta)$ in this subsection and leave the discussion of inferences for a vector $\boldsymbol{\phi}(\theta) = (\phi_1(\theta), \ldots, \phi_J(\theta))' : \Theta \to R^J$ for next subsection.

and (iii) implies that the approximation error of sieves is relatively small compared to the variance.

Let

$$\frac{d\phi(\theta_0)}{d\theta}\left[\bar{p}^{k_n}(\cdot)\right] \equiv \left(\frac{\partial\phi(\theta_0)}{\partial\beta'}, \frac{d\phi(\theta_0)}{dh}\left[p^{k_n}(\cdot)'\right]\right)'$$

be a $(d_\beta + k_n)-$vector, where $\bar{p}^{k_n}(\cdot) = (\mathbf{1}_{d_\beta}, p^{k_n}(\cdot)')'$. Note that $\frac{d\phi(\theta_0)}{d\theta}[\bar{p}^{k_n}(\cdot)]$ can be considered as the pathwise derivative of the functional in the direction $\bar{p}^{k_n}(\cdot)$. For example, for functional $\phi(\theta_0) = x'\beta_0 + h_0(\bar{y})$, we approximate $h_0(\bar{y})$ as $p^{k_n}(\bar{y})'\gamma_n$, then $\frac{d\phi(\theta_0)}{d\theta}\left[\bar{p}^{k_n}(\cdot)\right] = (x', p^{k_n}(\bar{y})')'$. For average derivatives $\phi(\theta_0) = \int \partial_{y_j} h_0(y) df(y)$, $\frac{d\phi(\theta_0)}{d\theta}[\bar{p}^{k_n}(\cdot)] = \left(0'_{d_\beta}, \int \partial_{y_j} p^{k_n}(y)' df(y)\right)'$.

The next theorem presents asymptotic normality of the plug-in SGMM estimator $\phi(\hat{\theta}_n)$.

**Theorem 3.1.** *Suppose that Assumptions 2.1, 3.1-3.7 and A.1-A.14 hold. Then*

$$\sqrt{n}V_{\phi,n}^{-1/2}\left(\phi\left(\hat{\theta}_n\right) - \phi(\theta_0)\right) \xrightarrow{d} N(0,1),$$

*where*

$$V_{\phi,n} \equiv \frac{d\phi(\theta_0)}{d\theta}\left[\bar{p}^{k_n}(\cdot)\right]' \Omega_n \frac{d\phi(\theta_0)}{d\theta}\left[\bar{p}^{k_n}(\cdot)\right],$$

$$\Omega_n \equiv E\left[\left(\frac{dm(X,\theta_0)}{d\theta}\left[\bar{p}^{k_n}(\cdot)'\right]\right)'\Sigma_0(X)^{-1}\left(\frac{dm(X,\theta_0)}{d\theta}\left[\bar{p}^{k_n}(\cdot)'\right]\right)\right]^{-1}$$

*and* $\Sigma_0(X) = E\left[\rho(Z,\theta_0)\rho(Z,\theta_0)'|X\right]$.

**Remark.**

(1) Theorem 3.1 delivers pointwise asymptotic normality for functionals of the SGMM estimator. It includes nonparametric regression as a special case, which was studied in Newey (1997). Note that the normalization factor $V_{\phi,n}^{1/2}$ is the pointwise standard error for functionals.

(2) We obtain the asymptotic distribution by assuming the estimation bias is small relative to variance under Assumption 3.7, which is an under-smoothing condition.

(3) The limiting distribution is asymptotically equivalent to the one that is obtained from the optimally weighted SMD estimator by Chen and Pouzo (2014). If we use a general weight matrix $\frac{1}{n}\sum_{i=1}^n \Sigma(X_i)^{-1} \otimes q^{s_n}(X_i)q^{s_n}(X_i)'$ instead of the optimal weight matrix, $\Omega_n$ would be

$$\begin{aligned}
&E\left[\left(\frac{dm(X,\theta_0)}{d\theta}\left[\bar{p}^{k_n}(\cdot)'\right]\right)'\Sigma(X)^{-1}\left(\frac{dm(X,\theta_0)}{d\theta}\left[\bar{p}^{k_n}(\cdot)'\right]\right)\right]^{-1} \\
&\times E\left[\left(\frac{dm(X,\theta_0)}{d\theta}\left[\bar{p}^{k_n}(\cdot)'\right]\right)'\Sigma(X)^{-1}\Sigma_0(X)\Sigma(X)^{-1}\left(\frac{dm(X,\theta_0)}{d\theta}\left[\bar{p}^{k_n}(\cdot)'\right]\right)\right] \\
&\times E\left[\left(\frac{dm(X,\theta_0)}{d\theta}\left[\bar{p}^{k_n}(\cdot)'\right]\right)'\Sigma(X)^{-1}\left(\frac{dm(X,\theta_0)}{d\theta}\left[\bar{p}^{k_n}(\cdot)'\right]\right)\right]^{-1}.
\end{aligned}$$

The first-order asymptotic equivalence between the SGMM and the SMD estimators is analogous to the asymptotic equivalence between the parametric GMM and parametric minimum distance

estimators (McFadden and Newey (1994)). □

Based on the limiting distribution in Theorem 3.1, an interesting result is presented in Corollary 3.1 below. For example, with

$$\phi(\theta) = \lambda'\beta + \phi_h(h(\cdot)),$$

Corollary 3.1 shows that that for the two-step SGMM estimators $\hat{\theta}_n = \left(\hat{\beta}_n, \hat{h}_n(\cdot)\right)$, and functional $\phi(\hat{\theta}_n)$, the estimators $\hat{\beta}_n$ and $\phi_h(\hat{h}_n)$ become asymptotically independent if $\phi_h(\cdot)$ is an irregular functional. Moreover, $\hat{\beta}_n$ achieves the semiparametric efficient bound.

To characterize the asymptotic variance of $\hat{\beta}_n$, we introduce some notation that is standard in the literature (see, for example, Ai and Chen (2003)). For each component $\beta_j$ of $\beta$, $j = 1, \ldots, d_\beta$, let $\varpi_j^* \in \bar{\mathcal{H}}/\{h_0\}$ denote the solution to

$$\min_{\varpi_j \in \bar{\mathcal{H}} - \{h_0\}} E\left\{ \left( \frac{dm(X,\theta_0)}{d\theta_j} - \frac{dm(X,\theta_0)}{dh}[\varpi_j] \right)' \Sigma_0(X)^{-1} \left( \frac{dm(X,\theta_0)}{d\theta_j} - \frac{dm(X,\theta_0)}{dh}[\varpi_j] \right) \right\}.$$

Let

$$\frac{dm(X,\theta_0)}{dh}[\varpi^*] = \left( \frac{dm(X,\theta_0)}{dh}[\varpi_1^*], \ldots, \frac{dm(X,\theta_0)}{dh}\left[\varpi_{d_\beta}^*\right] \right),$$

and

$$D_{\varpi^*}(X) \equiv \frac{dm(X,\theta_0)}{d\beta'} - \frac{dm(X,\theta_0)}{dh}[\varpi^*].$$

We summarize the asymptotic independence result of the parametric component and functionals of the nonparametric component as follows.

**Corollary 3.1.** *(Asymptotic Independence) Suppose that Assumptions 2.1, 3.1-3.7 and A.1-A.14 hold. Suppose $\phi(\theta) = \lambda'\beta + \phi_h(h)$ and $\phi_h(\cdot)$ satisfies (3.7), then*

$$\begin{pmatrix} \sqrt{n}\lambda'\left(\hat{\beta}_n - \beta_0\right) \\ \sqrt{n}V_{\phi_{h,n}}^{-1/2}\left(\phi_h\left(\hat{h}_n\right) - \phi_h\left(h_0\right)\right) \end{pmatrix} \xrightarrow{d} N\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} V_\beta & 0 \\ 0 & 1 \end{pmatrix} \right),$$

*where*

$$\begin{aligned} V_\beta &= \lambda'\Omega_\beta\lambda \equiv \lambda'\left(E\left[D_{\varpi^*}'\Sigma_0(X)^{-1}D_{\varpi^*}\right]\right)^{-1}\lambda, \\ V_{\phi_{h,n}} &= \frac{d\phi_h(h_0)}{dh}\left[p^{k_n}(\cdot)\right]'\Omega_{h,n}\frac{d\phi_h(\theta_0)}{dh}\left[p^{k_n}(\cdot)\right], \\ \Omega_{h,n} &= \left(E\left[\frac{dm(x,\theta_0)}{dh}\left[p^{k_n}(\cdot)'\right]'\Sigma_0(X)^{-1}\frac{dm(x,\theta_0)}{dh}\left[p^{k_n}(\cdot)'\right]\right]\right)^{-1}. \end{aligned} \tag{3.9}$$

**Remark.**

(1) Our asymptotic independence result is built upon Chamberlain (1992) and Cheng and Shang (2014). Chamberlain (1992) provides the bound on the asymptotic covariance matrix for the joint distribution of the parametric terms and nonparametric terms in a semiparametric conditional

moment model without proposing efficient estimators. Recently, Cheng and Shang (2014) establish the asymptotic independence of the Euclidean estimator and the (infinite-dimensional) functional parameters for a partially linear model based on penalized estimation. We extend the result of Cheng and Shang (2014) to the general setting of conditional moment restriction models.

(2) From the results in Corollary 3.1, for the marginal distribution of $\sqrt{n}(\hat{\beta}_n - \beta_0)$, the asymptotic variance $V_\beta$ is the same as the variance matrix obtained in Ai and Chen (2003) where the unknown function $h_0$ was treated as a nuisance parameter by profiling it out. Corollary 3.1 also presents the marginal distribution of the nonparametric estimator of unknown function in a semi-nonparametric moment restriction model, which is re-scaled by the asymptotic variance. □

We close this subsection by providing consistent estimates of the variance matrices to conduct statistical inference on the parameters. Once $h \in \mathcal{H}$ is approximated by (linear) sieves $h_n \in \mathcal{H}_n$, the estimators are easy to obtain: effectively by the same procedure to get consistent variance estimates in a parametric GMM model. Furthermore, based on the asymptotic independence result, the variance estimator can be obtained by variance estimators for the marginal distributions of the parametric and nonparametric components, respectively.

In particular, the estimator for $V_\beta$ can be obtained in the following way. For each $\beta_j$, $j = 1, \ldots, d_\beta$, we estimate

$$\begin{aligned}\hat{\varpi}^*_{nj} &= \min_{\varpi_{nj} \in \mathcal{H}^j_n} \left( \frac{1}{n}\sum_{i=1}^n \left\{ \frac{d\rho(Z_i, \hat{\theta}_n)}{d\beta_j} - \frac{d\rho(Z_i, \hat{\theta}_n)}{dh}[\varpi_j] \right\} \otimes q^{s_n}(X_i) \right)' \left( \frac{1}{n}\sum_{i=1}^n g_i\left(\hat{\theta}_n\right) g_i\left(\hat{\theta}_n\right)' \right)^{-1} \\ &\quad \times \left( \frac{1}{n}\sum_{i=1}^n \left\{ \frac{d\rho(Z_i, \hat{\theta}_n)}{d\beta_j} - \frac{d\rho(Z_i, \hat{\theta}_n)}{dh}[\varpi_j] \right\} \otimes q^{s_n}(X_i) \right).\end{aligned}$$

Let the estimator of $\varpi^*_n$ be $\hat{\varpi}^*_n = \left(\hat{\varpi}^*_{n1}, \ldots, \hat{\varpi}^*_{nd_\beta}\right)$. Then the estimator of $\Omega_\beta^{-1}$ is

$$\begin{aligned}\hat{\Omega}^{-1}_{\beta,n} &= \left( \frac{1}{n}\sum_{i=1}^n \left\{ \frac{d\rho(Z_i, \hat{\theta}_n)}{d\beta'} - \frac{d\rho(Z_i, \hat{\theta}_n)}{dh}[\hat{\varpi}^*] \right\} \otimes q^{s_n}(X_i) \right)' \left( \frac{1}{n}\sum_{i=1}^n g_i\left(\hat{\theta}_n\right) g_i\left(\hat{\theta}_n\right)' \right)^{-1} \\ &\quad \times \left( \frac{1}{n}\sum_{i=1}^n \left\{ \frac{d\rho(Z_i, \hat{\theta}_n)}{d\beta'} - \frac{d\rho(Z_i, \hat{\theta}_n)}{dh}[\hat{\varpi}^*] \right\} \otimes q^{s_n}(X_i) \right). \qquad (3.10)\end{aligned}$$

For the nonparametric component, the estimator of $\Omega_{h,n}^{-1}$ is

$$\begin{aligned}\hat{\Omega}^{-1}_{h,n} &= \left( \frac{1}{n}\sum_{i=1}^n \frac{d\rho(Z_i, \hat{\theta}_n)}{dh}\left[p^{k_n}(\cdot)'\right] \otimes q^{s_n}(X_i) \right)' \left( \frac{1}{n}\sum_{i=1}^n g_i\left(\hat{\theta}_n\right) g_i\left(\hat{\theta}_n\right)' \right)^{-1} \\ &\quad \times \left( \frac{1}{n}\sum_{i=1}^n \frac{d\rho(Z_i, \hat{\theta}_n)}{dh}\left[p^{k_n}(\cdot)\right] \otimes q^{s_n}(X_i) \right). \qquad (3.11)\end{aligned}$$

For any functionals of $\theta$ regardless of whether $\phi(\cdot)$ is regular or not, the estimators of $\Omega_n^{-1}$ and

$V_{\phi,n}$ are

$$\begin{aligned}\hat{\Omega}_n^{-1} &= \left(\frac{1}{n}\sum_{i=1}^{n}\frac{d\rho(Z_i,\hat{\theta}_n)}{d\theta}\left[\bar{p}^{k_n}(\cdot)'\right]\otimes q^{s_n}(X_i)\right)'\left(\frac{1}{n}\sum_{i=1}^{n}g_i\left(\hat{\theta}_n\right)g_i\left(\hat{\theta}_n\right)'\right)^{-1} \\ &\quad\times\left(\frac{1}{n}\sum_{i=1}^{n}\frac{d\rho(Z_i,\hat{\theta}_n)}{dh}\left[p^{k_n}(\cdot)\right]\otimes q^{s_n}(X_i)\right) \qquad (3.12)\end{aligned}$$

and

$$\hat{V}_{\phi,n}=\frac{d\phi(\hat{\theta}_n)}{d\theta}\left[\bar{p}^{k_n}(\cdot)\right]'\hat{\Omega}_n\frac{d\phi(\hat{\theta}_n)}{d\theta}\left[\bar{p}^{k_n}(\cdot)\right], \qquad (3.13)$$

respectively.

**Theorem 3.2.** *(Consistency of Variance) Suppose that Assumptions of Theorem 3.1 hold. Then* $\hat{\Omega}_{\beta,n}\xrightarrow{p}\Omega_\beta$, $\hat{\Omega}_{h,n}\xrightarrow{p}\Omega_{h,n}$, $\hat{\Omega}_n\xrightarrow{p}\Omega_n$ *and* $\hat{V}_{\phi,n}\xrightarrow{p}V_{\phi,n}$.

Although we have shown in Theorem 3.1 that the SGMM estimator is asymptotically equivalent to the optimally weighted SMD estimator in Chen and Pouzo (2014), the estimator of the variance for the SGMM estimator is calculated in a different way. For example, suppose $\rho(Z,\theta)$ is a scalar, we can estimate $\hat{\Sigma}_0(X)$ for the SMD estimator by

$$\hat{\Sigma}_0(X)=p^{s_n}(X)'(P'P)^{-1}\sum_{j=1}^{n}p^{s_n}(X_j)\rho^2(Z_j,\bar{\theta}_n^{SMD})$$

where $\bar{\theta}_n^{SMD}$ is a first-stage preliminary estimator. Asymptotically, $\hat{\Sigma}_0(X)$ is a consistent estimator for $\Sigma_0(X)$. However, in finite sample, SGMM estimator and SMD estimator may lead to different estimates of $\hat{\Sigma}_0(X)$. □

### 3.2 Uniform Limiting Theory

The results of this subsection are motivated by an interest in performing uniform inference over the domain of the unknown function $h_0(\cdot)$ or functionals of $h_0(\cdot)$. Researchers may be interested in uniform inference over the arguments of functionals of interest rather than pointwise results. For example, they may be interested in the following hypothesis: $\mathbb{H}_0:\bigtriangledown h(w)=0$ v.s. $\mathbb{H}_1:\bigtriangledown h(w)>0$ for all $w\in\mathcal{W}$. This section establishes a limiting theory for inference uniformly over the domain of functional. These results can be used to construct uniform confidence bands. Without loss of generality, we always write the domain of functional $\phi(\theta)$ as $w$, where $w$ is specified by each functional of interest.[9]

Furthermore, to distinguish the uniform analysis from pointwise one, we write the functional of interest as $\phi(\theta)[w]$ ( $\phi(\theta)[w]$ means functional of $\theta$ evaluated at $w$) instead of $\phi(\theta)$ to emphasize

[9]For instance, for the nonparametric IV model in Example 2.1, if we are interested in inference on $\bigtriangledown h(y_2)$ uniformly over $y_2\in\mathcal{Y}_2$, we write $w=y_2$. For the Engel Curves in Example 2.3, let $g(\cdot)$ be an identity function, if we are interested in inference on $h(y_2-x_1'\beta_1)$, we write $w=y_2-x_1'\beta_1$.

that we are considering inference uniformly over $w$. In contrast to parametric models, constructing uniform confidence bands is a difficult problem in semi-nonparametric models. Intuitively, one wishes to obtain the asymptotic distribution of a scaled version of

$$\sup_w \left| \frac{\phi(\hat{\theta}_n)[w] - \phi(\theta_0)[w]}{\hat{\sigma}_{\phi,n}(w)} \right|, \tag{3.14}$$

then the uniform inference could be conducted based on such process. However, as is pointed out in Chernozhukov, Lee and Rosen (2013), even for nonparametric regression model (a special case of model (1.1)), the left-hand side of (3.14) is not asymptotically equicontinuous, hence it does not have an asymptotic distribution. One may fail to derive valid inference methods in this case.

Given the lack of asymptotic distribution and applicability of empirical process methods for (3.14), we turn to another method of distributional approximation. In particular, we follow the strong approximation[10] literature to develop an approximation of the series process by a sequence of zero-mean Gaussian processes. The key idea is that when sample size increases, the accuracy of the strong approximation increases. Hence, we can do inference based on the the sequence of approximating Gaussian processes.

Our strong approximation results extend the previous literature (e.g., Belloni, Chernozhukov and Fernández-Val, 2011; Belloni, Chernozhukov, Chetverikov and Kato (2013)) in two respects. First, we consider a general model where the residual functions in the conditional moment restrictions can be nonlinear in a flexible way. Second, we allow the argument of the unknown functions to be endogenous.

**Theorem 3.3.** *Suppose that Assumptions 2.1, 3.1-3.6, Assumptions A.1-A.19 hold uniformly over $w \in \mathcal{W}$. Let $b_n$ be a sequence of positive numbers such that $b_n \to \infty$. Suppose that $b_n^6 (k_n + d_\beta)^2 \xi_{\rho,k_n}^2 \log^2 n/n \to 0$, then we have for some $\mathcal{N}_{d_\beta+k_n} \sim N(0, I_{d_\beta+k_n})$,*

$$\frac{\sqrt{n}\left(\phi(\hat{\theta}_n)[w] - \phi(\theta_0)[w]\right)}{\left\|\Omega_n^{1/2} A_n(w)\right\|_E} =_d \frac{A_n(w)' \Omega_n^{1/2}}{\left\|\Omega_n^{1/2} A_n(w)\right\|_E} \mathcal{N}_{d_\beta+k_n} + o_p(b_n^{-1}) \text{ in } \ell^\infty(\mathcal{W}), \tag{3.15}$$

*where $\Omega_n$ is defined in Theorem 3.1, $A_n(w) \equiv \left(\frac{d\phi(\theta_0)}{d\theta}\left[\bar{p}^{k_n}(\cdot)\right]\right)[w]$.*[11]

**Remark.** Theorem 3.3 can be regarded as a (uniform) functional central limit theorem for two-step SGMM estimators. It establishes that uniformly over $w$, the estimate of the functional $\phi(\theta_0)[w]$ can be strongly approximated by a sequence of Gaussian processes. If $k_n$ is a constant that does not vary when the sample size increases, the right hand side of (3.15) reduces to a standard multivariate normal distribution with a $(d_\beta + k_n) \times (d_\beta + k_n)$ identity matrix. In next subsection, we propose inference methods based on Theorem 3.3. □

[10] see Chapter 10 of Pollard (2001) or Appendix A in Chernozhukov, Lee and Rosen (2013) for a formal definition.

[11] With some abuse of notation, we use $\delta(w)$ or $\delta_n(w)$ to denote the direction of $\delta \in \bar{\Delta}$ or $\delta_n(w) \in \bar{\Delta}_n$ with argument $w$ for functional and $\phi(\theta)\left(\frac{\partial\phi(\theta_0)}{\partial\theta}[\delta]\right)[w]$ to denote function $\frac{\partial\phi(\theta_0)}{\partial\theta}[\delta]$ with argument $w$.

### 3.3 Inference Methods

This subsection outlines a large sample theory of hypothesis testing for SGMM estimators under point identification. We show that the trinity of Wald, quasi-likelihood ratio and Lagrange multiplier tests from parametric GMM models (Newey and McFadden, 1994) can be extended to the more general semi-nonparametric GMM models. Furthermore, we propose sup-Wald, sup-quasi-likelihood ratio and sup-Lagrange multiplier tests for uniform inference and establish properties of this trinity of tests. Our inference methods can be used to construct confidence intervals/regions and confidence bands/uniform confidence regions for functionals of the parameters (including the parameters themselves).

#### 3.3.1 Pointwise Inference Methods

We start with a univariate t-statistic and then extend it to a Wald statistic for multivariate tests. For the hypothesis $\mathbb{H}_0 : \phi(\theta) = \phi(\theta_0)$, with $\phi : \Theta \to \mathbf{R}$, the t-statistic is defined by

$$t_n = \hat{\sigma}_{\phi,n}^{-1} \left( \phi(\hat{\theta}_n) - \phi(\theta_0) \right),$$

where $\phi\left(\hat{\theta}_n\right)$ is a functional of the estimator $\hat{\theta}_n$ such that

$$\hat{\sigma}_{\phi,n}^2 \equiv \frac{d\phi\left(\hat{\theta}_n\right)}{d\theta} \left[\bar{p}^{k_n}(\cdot)\right]' \hat{\Omega}_n \frac{d\phi\left(\hat{\theta}_n\right)}{d\theta} \left[\bar{p}^{k_n}(\cdot)\right] / n,$$

and $\hat{\Omega}_n$ is defined in (3.12). The next result is a direct application of Theorem 3.1, which establishes that $t_n$ converges to a standard normal distribution under the null. Suppose, for example, we are interested in the unknown function at fixed point $w$, then $t_n$ can be employed to construct confidence intervals for $\hat{h}_n$ at a fixed point.

**Corollary 3.2.** *Suppose that Assumptions of Theorem 3.1 are satisfied. Then under the null, we have*

$$t_n = \hat{\sigma}_{\phi,n}^{-1} \left( \phi(\hat{\theta}_n) - \phi(\theta_0) \right) \xrightarrow{d} N(0,1).$$

**Remark.** Corollary 3.2 implies a way to construct confidence intervals as

$$\left[ \phi\left(\hat{\theta}_n\right) - c(1-\tau)\hat{\sigma}_{\phi,n},\ \phi\left(\hat{\theta}_n\right) - c(1-\tau)\hat{\sigma}_{\phi,n} \right],$$

where $c(1-\tau)$ is the $(1-\tau)$th quantile of the standard normal distribution. For implementation, the procedure is analogous to the one for parametric GMM model. □

If there are multiple restrictions on $\theta$, the joint hypothesis is $\mathbb{H}_0 : \boldsymbol{\phi}(\theta) = \boldsymbol{\phi}(\theta_0)$, $\boldsymbol{\phi} : \Theta \to \mathbf{R}^J$. A generalization of the $t-$statistic is a weighted quadratic form, known as the Wald statistic and denoted by $\text{Wald}_n$:

$$\text{Wald}_n = \left(\boldsymbol{\phi}(\hat{\theta}_n) - \boldsymbol{\phi}(\theta_0)\right)' \hat{V}_{\boldsymbol{\phi},n}^{-1} \left(\boldsymbol{\phi}(\hat{\theta}_n) - \boldsymbol{\phi}(\theta_0)\right)$$

where $\hat{V}_{\boldsymbol{\phi},n}$ is an estimate of $V_{\boldsymbol{\phi},n}$ such that

$$\hat{V}_{\boldsymbol{\phi},n} = \frac{d\boldsymbol{\phi}(\hat{\theta}_n)}{d\theta} \left[\bar{p}^{k_n}(\cdot)\right]' \hat{\Omega}_n \frac{d\boldsymbol{\phi}(\hat{\theta}_n)}{d\theta} \left[\bar{p}^{k_n}(\cdot)\right] \tag{3.16}$$

for $\hat{\Omega}_n$ in (3.12). The Wald statistic is useful for joint hypothesis such that $\mathbb{H}_0 : \bigtriangledown h(w^1) = \bigtriangledown h(w^2) = \beta_1 = \beta_2 = 0$ for two fixed points $w^1, w^2 \in \mathcal{W}$.

Let $\boldsymbol{\phi}(\theta) = (\phi_1(\theta), \ldots, \phi_J(\theta))$. Without loss of generality, we assume that $\left\{\frac{d\phi_j(\theta_0)}{d\theta}[\delta]\right\}_{j=1}^{J}$ are linearly independent. Otherwise we can conduct a linear transformation for the joint hypothesis.

**Assumption 3.8.** *(i) For $\boldsymbol{\phi}(\theta) = (\phi_1(\theta), \ldots, \phi_J(\theta))$, $\frac{d\phi_j(\theta_0)}{d\theta}[\cdot]$ is a linear functional on $\bar{\Delta}$ that satisfies Assumption 3.7 for $j = 1, \ldots, J$; (ii)*

$$\frac{d\boldsymbol{\phi}(\theta_0)}{d\theta}[\delta] \equiv \left(\frac{d\phi_1(\theta_0)}{d\theta}[\delta], \ldots, \frac{d\phi_J(\theta_0)}{d\theta}[\delta]\right)'$$

*is linearly independent.*

Two alternative choices to the Wald statistic are the quasi-likelihood ratio statistic and the Lagrange Multiplier statistic. Let $\hat{L}_n(\theta)$ be the second-stage SGMM criterion function defined in (3.1). Let the estimates under $\mathbb{H}_0 : \boldsymbol{\phi}(\theta) = \boldsymbol{\phi}(\theta_0) = r_0$ be

$$\tilde{\theta}_n = \arg \min_{\theta \in \Theta_n \cap \{\boldsymbol{\phi}(\theta) = r_0\}} \hat{L}_n(\theta). \tag{3.17}$$

and those unconstrained estimator be

$$\hat{\theta}_n = \arg \min_{\theta \in \Theta_n} \hat{L}_n(\theta).$$

The quasi-likelihood ratio statistic (sometimes is called a minimum distance statistic) is the difference

$$\text{QLR}_n = n\left\{\hat{L}_n(\tilde{\theta}_n) - \hat{L}_n\left(\hat{\theta}_n\right)\right\}. \tag{3.18}$$

Finally, the two-step SGMM criterion with optimal weight matrix implies that the Lagrange Multiplier statistic (score statistic) is

$$\text{LM}_n = \frac{n}{4}\left\{\frac{d\hat{L}_n\left(\tilde{\theta}_n\right)}{d\theta}\left[\bar{p}^{k_n}(\cdot)'\right]' \left(\frac{d\boldsymbol{\phi}(\tilde{\theta}_n)}{d\theta}\left[\bar{p}^{k_n}(\cdot)\right]\right) \tilde{V}_{\boldsymbol{\phi},n}^{-1} \left(\frac{d\boldsymbol{\phi}(\tilde{\theta}_n)}{d\theta}\left[\bar{p}^{k_n}(\cdot)\right]\right)' \frac{d\hat{L}_n\left(\tilde{\theta}_n\right)}{d\theta}\left[\bar{p}^{k_n}(\cdot)'\right]\right\}$$

where $\tilde{\theta}_n$ is the constrained estimator defined in (3.17),

$$\tilde{V}_{\boldsymbol{\phi},n} = \frac{d\boldsymbol{\phi}(\tilde{\theta}_n)}{d\theta}\left[\bar{p}^{k_n}(\cdot)\right]' \tilde{\Omega}_n \frac{d\boldsymbol{\phi}(\tilde{\theta}_n)}{d\theta}\left[\bar{p}^{k_n}(\cdot)\right]. \tag{3.19}$$

and $\tilde{\Omega}_n$ is exactly analogous to $\hat{\Omega}_n$ in (3.12) by replacing $\hat{\theta}_n$ with $\tilde{\theta}_n$ in the expression.

Theorem 3.4 develops the asymptotic distribution for the three classes of test statistics. We develop a semi-nonparametric Wilks' phenomenon (Wilks, 1938), i.e., the asymptotic distribution under the null is free of nuisance parameters, by showing that all of the three statistics converge to a chi-square distribution with degree of freedom $J$ under regularity conditions.

**Theorem 3.4.** *(Pointwise Trinity) Suppose that Assumptions of Theorem 3.1 and Assumption 3.8 are satisfied. Then*

*(i) Under $\mathbb{H}_0$, $Wald_n$, $QLR_n$ and $LM_n$ all converge in distribution to a $\chi^2_J$ distribution.*

*(ii) For $\varrho_n$, $\varrho_{sn}$ defined in Lemma 3.1 and $K_n = \log\log n$, let $\omega_n \in \bar{\Delta}_n$ such that $||\omega_n||_s \leq \sqrt{n}||\delta^*_n||^{-1}K_n\varrho_{sn}$, $||\omega_n|| \leq \sqrt{n}||\delta^*_n||^{-1}K_n\varrho_n$ for all $n$. Let $\frac{\partial\boldsymbol{\phi}(\theta_0)}{\partial\theta}[\omega_n] = c_n = c(1+o(1))$, Then under the following local alternative sequence,*

$$\{\mathring{\theta}_n \in \mathcal{N}_{osn} : \mathring{\theta}_n = \theta_0 + \frac{||\delta^*_n||}{\sqrt{n}} \times \omega_n\}, \tag{3.20}$$

*$Wald_n$, $QLR_n$ and $LM_n$ all converge in distribution to a $\chi^2_J(c'c)$ distribution.*

**Remark.**

(1) To the best of our knowledge, Theorem 3.4 is the first semi-nonparametric version of the trinity results for SGMM models. It establishes that the three major classes of statistics are asymptotically equivalent (at least to a first-order asymptotic approximation) for SGMM estimates. This pattern of first-order asymptotic equivalence is analogous to the trinity results for GMM estimates in the parametric framework.

(2) Although there is no clear statistical reason to choose between the three statistics based on Theorem 3.4, in practice, it is often computationally easier to use one of the trinity tests rather than another. The computational advantages of each test mirror their computational advantages in standard parametric settings.

The Wald statistic is based on the length of the vector $\boldsymbol{\phi}(\hat{\theta}_n) - \boldsymbol{\phi}(\theta_0)$, i.e., the discrepancy between the unconstrained estimator and the hypothesized value $\boldsymbol{\phi}(\theta_0)$. It is particularly useful when the variance matrix is easy to compute. On the other hand, if the hypothesis is non-linear and the constrained estimator is available, a better approach to construct the test statistic can be to directly use the SGMM criterion function via the $\text{QLR}_n$ statistic. Newey and West (1987) was the first paper to propose such an idea in a parametric setting. The $\text{QLR}_n$ statistic generalizes their approach to a semi-nonparametric model. It is especially useful when estimating the variance for studentization is difficult. In semi-nonparametric models, the asymptotic variance of the estimate may not be in closed form; furthermore, inverting the Fisher information matrix can be difficult

when the dimension of the matrix is high. Thus, it could be more convenient to invert $\text{QLR}_n$ to construct confidence intervals, which are invariant to nonsingular transformations of the moment conditions. Similar to the $\text{QLR}_n$, to get the $\text{LM}_n$ statistic, we need to calculate the constrained estimator. Compared to the $\text{QLR}_n$, the advantage of $\text{LM}_n$ is that it does not require one to compute the unconstraint estimator and may have some computational advantages in certain applications.

(3) The treatment of the trinity of the tests have provides a modest extension of the results in Chen and Pouzo (2014) to the SGMM setting. Chen and Pouzo (2014) have established the trinity of test by using SMD estimator. Our approach is especially convenient for testing restrictions that depend on both parametric and nonparametric components of $\theta$. Based on the asymptotic independence result in Corollary 3.1, we can ignore the estimate of covariance terms and construct the confidence intervals based on marginal distributions of parametric and nonparametric estimates, respectively. □

### 3.3.2 Uniform Inference Methods

To consider hypotheses such that $\mathbb{H}_0 : \phi(\theta)[w] = \phi(\theta_0)[w]$ for all $w \in \mathcal{W}$ with $\phi(\theta_0)[w]$ being a given function with argument $w$. we begin by augmenting the notation to write the test statistics as the ones indexed by $w$ (for example, $t_n$ by $t_n(w)$, $\text{Wald}_n$ by $\text{Wald}_n(w)$).

We start with a test statistic based on the following $t_n-$statistic process

$$\left\{ t_n(w) = \frac{\phi(\hat{\theta}_n)[w] - \phi(\theta_0)[w]}{\hat{\sigma}_{\phi,n}(w)}, w \in \mathcal{W} \right\}. \tag{3.21}$$

As we argued in Section 3.2, this process may not have a limit distribution uniformly over $w \in \mathcal{W}$. Alternatively, we find a (studentized) Gaussian process to approximate the process in (3.21) as

$$\left\{ t_n^*(w) = \frac{A_n(w)'\Omega_n^{1/2}\mathcal{N}_{d_\beta+k_n}/\sqrt{n}}{\sigma_{\phi,n}(w)}, w \in \mathcal{W} \right\}$$

where $\mathcal{N}_{d_\beta+k_n}$ is a $(d_\beta + k_n)-$vector of i.i.d.random variables that are drawn from a standard multivariate normal distribution $N(0,1)$ and $A_n(w) \equiv \left( \frac{d\phi(\theta_0)}{d\theta} \left[ \bar{p}^{k_n}(\cdot) \right] \right)[w]$. We are interested in constructing the following confidence bands

$$\left[ \dot{l}(w),\ \ddot{l}(w) \right] = \left[ \phi(\hat{\theta}_n)[w] - c_n(1-\tau)\hat{\sigma}_{\phi,n}(w), \phi(\hat{\theta}_n)[w] + c_n(1-\tau)\hat{\sigma}_{\phi,n}(w) \right],\ w \in \mathcal{W}$$

where we set $c_n(1-\tau)$ be the $(1-\tau)$th quantile of

$$\sup_{w\in\mathcal{W}} \left| \hat{t}_n^*(w) \right| \equiv \sup_{w\in\mathcal{W}} \left| \frac{\hat{A}_n(w)'\hat{\Omega}_n^{1/2}\mathcal{N}_{d_\beta+k_n}/\sqrt{n}}{\hat{\sigma}_{\phi,n}(w)} \right|, w \in \mathcal{W}.$$

Thus, $c_n(1-\tau)$ can be simulated numerically. In Theorem 3.5, we show that $\phi(w) \in \left[ \dot{l}(w),\ \ddot{l}(w) \right]$

for all $w \in \mathcal{W}$ with probability $1-\tau$.

**Theorem 3.5.** *(Uniform Confidence Bands for Functionals). Suppose that Assumptions of Theorem 3.3 hold. Then*

*(i)*

$$\sup_{w\in\mathcal{W}} |t_n(w)| \overset{d}{=} \sup_{w\in\mathcal{W}} |t_n^*(w)| + o_p(1),$$

*(ii)*

$$\Pr\left\{ \sup_{w\in\mathcal{W}} |t_n(w)| \leq c_n\left(1-\tau\right) \right\} = 1-\tau+o(1).$$

**Remark.** (1) The proof strategy for Theorem 3.5 is similar to that proposed in Belloni, Chernozhukov, Chetverikov and Kato (2013), although we need to handle endogeneity in our model. Since the limit distribution may not exist, their insight is to use distributions provided by a strong approximation.We show that the test has asymptotically correct size, even though the strong approximation approach cannot help us to obtain a fixed limiting distribution.

(2) One-sided confidence band can be defined by, for example,

$$\left[\dot{l}(w),\ \ddot{l}(w)\right] \equiv (-\infty, \phi(\hat{\theta}_n)[w] + c_n(1-\tau)\hat{\sigma}_n(w)],\ \forall w \in \mathcal{W},$$

with modifications to consider one-sided critical value for a given level $1-\tau$.□

For multivariate constraints, we propose the following three corresponding test statistics and show that they are asymptotically equivalent when sample size increases. The three test statistics are the uniform version of the three main statistics we use for pointwise inference.

We will approximate the three main statistic processes by the following "chi-square coupling"

$$T_n^*(w) = \mathcal{N}'_{d_\beta+k_n} \Omega_n^{1/2} \mathbf{A}_n(w) V_{\boldsymbol{\phi}}^{-1} \mathbf{A}_n(w)' \Omega_n^{1/2} \mathcal{N}_{d_\beta+k_n}, \tag{3.22}$$

where $\mathbf{A}_n(w) \equiv \frac{d\boldsymbol{\phi}(\theta_0)}{d\theta}[\bar{p}^{k_n}(\cdot)]$ is a $(d_\beta+k_n)\times J$ matrix.

**Theorem 3.6.** *(Uniform Trinity) Suppose that Assumptions of 3.3 and Assumption 3.8 (ii) hold. We have*

*(i)*

$$\sup - Wald_n \equiv \sup_{w\in\mathcal{W}} \{ Wald_n(w)\} \overset{d}{=} \sup_w \{T_n^*(w)\} + o_p(1);$$

$$\sup - QLR_n \equiv \sup_{w\in\mathcal{W}} \{ QLR_n(w)\} \overset{d}{=} \sup_w \{T_n^*(w)\} + o_p(1)$$

*and*

$$\sup - LM_n \equiv \sup_{w\in\mathcal{W}} \{ LM_n(w)\} \overset{d}{=} \sup_w \{T_n^*(w)\} + o_p(1).$$

*(ii) Let*

$$\hat{T}_n^*(w) = \mathcal{N}'_{d_\beta+k_n} \hat{\Omega}_n^{1/2} \hat{\mathbf{A}}_n(w) \hat{V}_{\boldsymbol{\phi}}^{-1} \hat{\mathbf{A}}_n(w)' \hat{\Omega}_n^{1/2} \mathcal{N}_{d_\beta+k_n},$$

$\hat{\mathbf{A}}_n[w] = \left(\frac{\partial \boldsymbol{\phi}(\hat{\theta}_n)}{\partial \beta'}, \frac{\partial \boldsymbol{\phi}(\hat{\theta}_n)}{\partial h}\left[p^{k_n}(\cdot)'\right]\right)'[w]$, *and* $c_n^0(1-\tau)$ *be the conditional* $(1-\tau)$*th quantile of* $\sup_{w\in\mathcal{W}}\left\{\hat{T}_n^*(w)\right\}$ *given the data. We have*

$$\Pr\left\{\text{sup-}Wald_n \leq c_n^0(1-\tau)\right\} = 1-\tau+o(1).$$

*(iii) Let*

$$\tilde{T}_n^*(w) = \mathcal{N}'_{d_\beta+k_n}\tilde{\Omega}_n^{1/2}\tilde{\mathbf{A}}_n(w)\tilde{V}_{\boldsymbol{\phi},n}^{-1}\tilde{\mathbf{A}}_n(w)'\tilde{\Omega}^{1/2}\mathcal{N}_{d_\beta+k_n},$$

$\tilde{\mathbf{A}}_n(w) = \frac{\partial \boldsymbol{\phi}(\tilde{\theta}_n)}{\partial h}\left[\bar{p}^{k_n}(\cdot)\right][w]$, $c_n^{00}(1-\tau)$ *be the conditional* $(1-\tau)$*th quantile of* $\sup_{w\in\mathcal{W}}\left\{\tilde{T}_n^*(w)\right\}$ *given the data. We have*

$$P\left\{\sup - LM_n \leq c_n^{00}(1-\tau)\right\} = 1-\tau+o(1).$$

**Remark.** (1) Theorem 3.6 shows that the uniform confidence regions are asymptotically similar. We establish the "uniform trinity" results for the three main classic test statistics. A relevant result is presented in Chen and Pouzo (2014). They establish the inference methods for functionals of increasing dimension, where the dimension of $\boldsymbol{\phi}(\theta_0)$, $J$, can grow with sample size $n$. By restricting the growth rate of $J$ (i.e., $J = J(n)$ cannot grow faster than $n^{1/4}$), they show that the limiting distribution of a weighted sieve Wald or a weighted sieve QLR is a standard normal. Our results in Theorem 3.6 do not restrict the growth rate of $J$, but then the test statistics do not have closed-form limiting distributions. Even though the limit distributions may not exist, we can use approximations provided by a sequence of "chi-squared processes" to obtain critical values.

Note that $\Omega_n^{1/2}$ and $\mathbf{A}_n(w)$ in (3.22) are unknown. Based on results in Theorem 3.6, we can set the critical values for $\sup - \text{Wald}_n$ as the $(1-\tau)$th quantile of $\sup_{w\in\mathcal{W}}\left\{\hat{T}_n^*(w)\right\}$ and the critical values for $\sup - \text{LM}_n$ as the $(1-\tau)$th quantile of $\sup_{w\in\mathcal{W}}\left\{\tilde{T}_n^*(w)\right\}$, respectively. For implementation, there are two main approaches to obtain critical values for uniform inference. One is to directly obtain critical values from the the "chi-squared processes". For the sup-Wald statistic, we first obtain the *unconstrained* estimator $\hat{\theta}_n$, the plug-in estimator $\hat{\mathbf{A}}_n$ and the variance estimator $\hat{V}_{\phi,n}^{-1}$ defined in (3.16), then simulate the quantile of $\sup_{w\in\mathcal{W}}\left\{\hat{T}^*(w)\right\}$ given the data by taking draws on the Gaussian part of the chi-square processes keeping the other terms fixed at their estimated values. For the sup-$\text{LM}_n$ statistic, in contrast, we first obtain the *constrained* estimator $\tilde{\theta}_n$, the plug-in estimator $\tilde{\mathbf{A}}_n$ and the variance estimator $\tilde{V}_{\phi,n}^{-1}$ defined in (3.19), then simulate the quantile of $\sup_{w\in\mathcal{W}}\left\{\tilde{T}_n^*(w)\right\}$ given the data.

(2) Alternatively, sup-$\text{QLR}_n$ does not require to estimate variance-covariance matrix. However, to implement $\sup - \text{QLR}_n$, we need to implement a multiplier bootstrap procedure to obtain critical values. For brevity of the paper, we put the details of such multiplier bootstrap method under point identification in Appendix B. □

# 4 ASYMPTOTIC RESULTS UNDER PARTIAL IDENTIFICATION

In this section we discuss inference methods under partial identification by relaxing the point identification assumption (Assumption 3.1). That is, the conditional moment restrictions are now allowed to be satisfied at more than one value of $\theta$. The set of parameter values that satisfy the conditional moment restrictions is called *the identified set*

$$\Theta_0 = \left\{\theta = \left(\beta', h(\cdot)\right) : \ E\left[\rho\left(Z;\beta,h(\cdot)\right)|X\right] = 0\right\} \ a.s.\ X. \tag{4.1}$$

In the partially identified setting, we will use a GMM criterion with a general weight matrix. In the point identified case, the optimal weight matrix corresponds to the variance-covariance of the moments at the true value $\theta_0$. In the partially identified case, there is no longer a unique value of $\theta$ at which one would naturally evaluate the variance-covariance matrix. Hence we proceed with a general weight matrix and consider a one-step SGMM criterion

$$\bar{L}_n(\theta) = \hat{g}(\theta)'\hat{W}\hat{g}(\theta) \tag{4.2}$$

where $\theta \in \Theta_n$, $\Theta_n$ is the sieve space for $\Theta$, $\hat{W}$ is a positive semi-definite matrix such that $\hat{W} = \left(\frac{1}{n}\sum_{i=1}^n \Sigma(X_i) \otimes q(X_i)^{s_n} q(X_i)^{s_n'}\right)^{-1}$, $\Sigma(X)$ is a positive definite matrix that does *not* depend on $\theta$.[12] The corresponding population criterion function is defined by

$$\bar{L}(\theta) = \sqrt{m(X,\theta)'\Sigma(X)^{-1}m(X,\theta)}.$$

**Assumption 4.1.** *(i) The identified set $\Theta_0$ is a nonempty, closed, bounded strict subset of $\Theta$ under $||\cdot||_s$; (ii) $\bar{L} : \Theta \to [0,\infty)$ is lower semicontinuous on $\Theta$ under $||\theta||_s = |\beta|_e + ||h||_{\mathbf{H}}$*

**Assumption 4.2.** *(i) $\Sigma(X)$ is positive definite a.s. $X$, its largest and smallest eigenvalues are finite positive; (ii) each element of $\rho(Z,\theta)$ satisfies an envelop condition over $\theta \in \Theta_n$; (iii) $\bar{L}_n(\theta) \gtrsim \bar{L}(\theta) - O_p(\varrho_{pn})$ with for $\varrho_{pn} = o_p(n^{-1/4})$.*

Conditions in Assumption 4.1 and Assumption 4.2 are modifications of Assumptions 3.3, 3.5 and 3.6 under partial identification. They provide conditions for set consistency when point identification fails.

We establish the consistency of $\hat{\Theta}_0$ for $\Theta_0$ as follows. Let the family of Hausdorff norms be defined by

$$d_H\left(\Theta_1,\Theta_2,||\cdot||\right) \equiv \max\left\{d(\Theta_1,\Theta_2), d(\Theta_2,\Theta_1)\right\}, \text{ with } d(\Theta_1,\Theta_2) \equiv \sup_{\theta_1\in\Theta_1}\inf_{\theta_2\in\Theta_2}||\theta_1-\theta_2||.$$

[12] A more general form of the criterion function $\bar{L}_n(\theta) = \hat{g}(\theta)'\hat{W}\hat{g}(\theta) + \lambda_n \text{Pen}(h)$ with non-compact $\mathcal{H}$ is presented in Appendix B. For simplicity and ease of exposition in this section, we set $\lambda_n = 0$ and focus on the case where $\mathcal{H}$ is compact under $||\cdot||_{\mathbf{H}}$.

Set consistency can be constructed under the family of Hausdorff norms, which is based on different choices of $||\cdot||$. Unlike the parametric case, different choices of norms imply significantly different rates of convergences. We will show that set consistency can be constructed under "strong" norms $||\cdot||_s$ based on $||\cdot||_{\mathbf{H}}$, where $||\cdot||_{\mathbf{H}}$ can be, for example, $||\cdot||_{L^2}$ or $||\cdot||_\infty$. Elements in the identified set can be distinguished under $||\cdot||_s$.

As is well known in the semiparametric literature under point identification, in order for the parametric term to be $\sqrt{n}$ consistent, the unknown function must be estimated at a rate that is at least $o_p(n^{-1/4})$. However, in general, convergence rates obtained from $||\cdot||_s$ are slower than $o_p(n^{-1/4})$. For point identified models in Section 3, we establish the convergence rate and limiting theory under the norm defined in (3.4) that is based on the derivatives of the conditional moment functions evaluated at $\theta_0$. In the partially identified model, there may not be a single value $\theta_0$ at which the derivatives of the conditional moments should be evaluated. Instead, the parameters satisfying the conditional moment restrictions are allowed to lie in a set $\Theta_0$. For this reason, we consider a different norm than the one given in (3.4).

More specifically, we establish the rate of convergence for estimators of $\Theta_0$ and inference methods based on a different pseudo-metric called $||\cdot||_{wp}$ ($wp$ is an abbreviation for "weak norm" under partial identification), which is defined by

$$\|\theta_1 - \theta_2\|_{wp} = \sqrt{E\left[\{E\left[\rho(Z,\theta_1) - \rho(Z,\theta_2)|X\right]\}' \Sigma(X)^{-1} \{E\left[\rho(Z,\theta_1) - \rho(Z,\theta_2)|X\right]\}\right]}$$

for $\theta_1, \theta_2 \in \Theta$.

For all $\theta \in \Theta$ and any $\theta_0 \in \Theta_0$, since $m(X,\theta_0) = E\left[\rho(Z,\theta_0)|X\right] = 0$,

$$\begin{aligned} ||\theta - \theta_0||_{wp} &= \sqrt{E\left[E\left[\rho(Z,\theta) - \rho(Z,\theta_0)|X\right]' \Sigma(X)^{-1} E\left[\rho(Z,\theta) - \rho(Z,\theta_0)|X\right]\right]} \\ &= \sqrt{E\left[m(X,\theta)' \Sigma(X)^{-1} m(X,\theta)\right]}. \end{aligned}$$

Note that for any $\theta_0^1$, $\theta_0^2 \in \Theta_0$, with $\theta_0^1 \neq \theta_0^2$, we have $\left\|\theta_0^1 - \theta_0^2\right\|_{wp} = 0$ and for any $\theta \notin \Theta_0$, $\left\|\theta - \theta_0^1\right\|_{wp} = \left\|\theta - \theta_0^2\right\|_{wp}$.

An important insight is that although consistency is based on the Hausdorff norm with $||\cdot||_s$, the elements of $\Theta_0$ form an equivalence class under $||\cdot||_{wp}$. In this sense $\Theta_0$ can be treated as a singleton under $||\cdot||_{wp}$, so it is convenient to describe convergence rate of $\hat{\Theta}_0$ based on $||\cdot||_{wp}$.[13]

Let $\hat{\Theta}_0$ be a collection of $\hat{\theta}_n = \left(\hat{\beta}_n, \hat{h}_n\right) \in \Theta_n$ that is a set of the minimizers of $\bar{L}_n(\theta)$. The following Lemma formalizes the results of consistency and the convergence rate of $\hat{\Theta}_0$ for $\Theta_0$.

**Lemma 4.1.** *(Set Consistency and Rate of Convergence). Let Assumptions 2.1, 3.2, 3.3, 4.1, 4.2 and A.20 hold. We have*

$$d_H\left(\hat{\Theta}_0, \Theta_0, ||\cdot||_s\right) = o_p(1);\ d_H\left(\hat{\Theta}_0, \Theta_0, ||\cdot||_{wp}\right) = o_p(n^{-1/4}).$$

[13]Similar ideas to treat identified set as an equivalence class have been explored in Liu and Shao (2003), Chen, Tamer and Torgovitsky (2011) for likelihood models and Santos (2011) for a nonparametric IV model.

**Remark.** Lemma 4.1 provides consistency results for $\hat{\Theta}_0$ to $\Theta_0$ under $||\cdot||_s$ (such as $||\cdot||_\infty$ and $||\cdot||_{L^2}$) and the rate of convergence of $\hat{\Theta}_0$ to $\Theta_0$ under norm $||\cdot||_{wp}$. It implies that we can focus our attention on the neighborhood of the identified set when considering inference. For any $\theta_0 \in \Theta_0$, let be $\sigma_n$ be the convergence rate of $d_H\left(\hat{\Theta}_0, \Theta_0, ||\cdot||_{wp}\right)$. By Lemma 4.1, we can define the neighborhood of $\Theta_0$ as

$$\mathcal{B}(\theta_0) = \mathcal{B}(\Theta_0) \equiv \{\theta \in \Theta : ||\theta - \theta_0||_{wp} \leq \sigma_n \log\log n\} \tag{4.3}$$

and the corresponding sieve approximation of $\mathcal{B}(\theta_0)$ is defined by

$$\mathcal{B}_n(\theta_0) \equiv \{\theta_n \in \Theta_n : ||\theta_n - \theta_0||_{wp} \leq \sigma_n \log\log n\}.$$

Let $\theta_{0n} = \arg\min_{\theta \in \Theta_n} ||\theta - \theta_0||_{wp}$.

### 4.1 Pointwise Inference Method

Given consistency,we can focus on the neighborhood $\mathcal{B}(\theta_0)$ and its sieve approximation $\mathcal{B}_n(\theta_0)$ for all $\theta_0 \in \Theta_0$ to develop our testing results.

Suppose we are interested in the following vector of functionals of the parameter $\boldsymbol{\phi}(\theta) = (\phi_1(\theta), \ldots, \phi_J(\theta))' : \Theta \to \mathbf{R}^J$. Let the null set be $\mathbb{R} = \{\theta : \boldsymbol{\phi}(\theta) = \boldsymbol{r}\}$. The hypothesis we consider are of the form

$$\mathbb{H}_0 : \Theta_0 \cap \mathbb{R} \neq \emptyset, \ \mathbb{H}_1 : \Theta_0 \cap \mathbb{R} = \emptyset, \tag{4.4}$$

where $\mathbb{R}$ is a set of functions that satisfy a property we wish to test for. When $\theta_0$ is point identified, the null hypothesis and the alternative simplify to

$$\mathbb{H}_0 : \boldsymbol{\phi}(\theta) = \boldsymbol{\phi}(\theta_0), \ \mathbb{H}_1 : \boldsymbol{\phi}(\theta) \neq \boldsymbol{\phi}(\theta_0).$$

We denote $\Theta_0^r = \Theta_0 \cap \mathbb{R}$.

To test hypotheses on functionals of $\theta$, we utilize the information we obtain from the conditional moment restrictions. However, if the residual functions are identical for two different parameters, the information provided by the conditional moment restrictions (via the residual functions) would be the same for each parameter value. In this case, the information from the conditional moment restrictions would not allow us to distinguish between these two values of the parameter. Thus, we impose the following Assumption 4.3 to guarantee that there is a one-to-one mapping from parameters of interest to residual functions.

**Assumption 4.3.** *Assume that (i) $\forall \theta^1, \theta^2 \in \Theta$, if $\rho(Z, \theta^1) = \rho(Z, \theta^2)$, then $\phi(\theta^1) = \phi(\theta^2)$; (ii) there exists a mapping $\mathbf{F}(\rho(\cdot, \theta)) = \phi(\theta)$ that is pathwise differentiable at any (fixed) point $\theta_0 \in \Theta_0$ so that $\frac{d\mathbf{F}(\rho(\cdot,\theta_0))}{d\rho}[\rho - \rho_0]$ exists for $\theta \in \mathcal{B}(\theta_0)$.*[14]

[14]Note that the residual function $\rho(\cdot, \theta)$ is always indexed by $\theta$. For simplicity, we sometimes write $\rho(\theta_0)$ or $\rho_0$ to represent $\rho(\cdot, \theta_0)$ and $\rho(\theta)$ to represent $\rho(\cdot, \theta)$.

Importantly, Assumption 4.3 implies that functionals of the parameters $\theta$ can be treated as functionals of the residual functions in our analysis. For the case where $\theta^1 \neq \theta^2$ but $\rho(Z,\theta^1) = \rho(Z,\theta^2)$, a.s. $Z$, Assumption 4.3 guarantees that $\mathbf{F}(\rho(\cdot,\theta^1)) = \mathbf{F}(\rho(\cdot,\theta^2))$ because $\phi(\theta^1) = \phi(\theta^2)$. So $\mathbf{F}(\rho(\cdot,\theta))$ is a well-defined function. Another way to view Assumption 4.3 is to consider the situation where for $\phi(\theta^1) \neq \phi(\theta^2)$, $\phi(\theta^1) \in \Theta_0 \cap \mathbb{R}$ and $\phi(\theta^2) \notin \Theta_0 \cap \mathbb{R}$. Assumption 4.3 ensures that $\rho(Z,\theta^1) \neq \rho(Z,\theta^2)$ a.s. $Z$, which means that there is a chance that the conditional moment restrictions provide different information on these parameters so that we are able to learn whether one parameter is in $\Theta_0 \cap \mathbb{R}$ and the other is not. In this sense, testing the functional restriction $\boldsymbol{\phi}(\theta) = \boldsymbol{r}$ is equivalent to test $\mathbf{F}(\rho) = \mathbf{r}$.

An assumption that directly implies Assumption 4.3 is the following Assumption 4.3'. Although we only need Assumption 4.3 to satisfy when we do inference and Assumption 4.3' is stronger, we argue that in some applications Assumption 4.3' is easier to verify.

**Assumption 4.3'.** Assume that (i) $\forall \theta^1, \theta^2 \in \Theta$, if $\rho(Z,\theta^1) = \rho(Z,\theta^2)$, then $\theta^1 = \theta^2$ a.s. $Z$; (ii) $\rho(Z,\theta)$ is a smooth function of $\theta$ for any point $\theta_0 \in \Theta_0$ such that $\frac{d\rho(Z,\theta_0)}{d\theta}[\theta - \theta_0]$ exists for $\theta \in \mathcal{B}(\theta_0)$.

While Assumption 4.3 (or 4.3') imposes some limits on the models we consider, we see that these conditions are not too restrictive and many examples of interest satisfy Assumption 4.3 or 4.3'. In order to fix ideas, we illustrate Assumption 4.3 and Assumption 4.3' through the following examples.

**Example 4.1** (Parametric Linear/Nonlinear IV). Consider the following model

$$\begin{aligned} Y_1 &= g(Y_2,\beta) + e_i,\ E[e|Y_2] \neq 0,\ E[e|X] = 0, \\ \rho(Z,\beta) &= Y_1 - g(Y_2,\beta), \end{aligned}$$

where $g(\cdot)$ is known, $\beta \in \mathbf{R}^{d_\beta}$, and $Z = (Y, W')$ and $X$ is a vector of IVs. To satisfy Assumption 4.3', we require that $\forall \beta^1, \beta^2 \in \mathbf{R}^{d_\beta}$, if $g(Y_2,\beta^1) = g(Y_2,\beta^2)$ a.s. $Y_2$, then $\beta^1 = \beta^2$. For example, Assumption 4.3' is satisfied when $g(Y_2,\beta) = Y_2'\beta$ or $g(Y_2,\beta) = \frac{\exp(Y_2'\beta)}{1+exp(Y_2'\beta)}$. □

**Example 4.2** (Partially Linear Model, Example 2.2 continued). The model we consider is

$$E[\rho(Z,\theta)|X] = E[Y_1 - G(X_1\beta + h(Y_2))|X]. \tag{4.5}$$

If $G(\cdot)$ is an identity function or a strictly monotone function, then Assumption 4.3' is satisfied. A direct implication is that the nonparametric IV model (Example 2.1) satisfies Assumption 4.3' as well. □

**Example 4.3** (Engel Curves, Example 2.3 continued). Suppose $g(\cdot)$ is an identity function in

Example 2.3, then for goods $l = 1, \ldots, L$, the residual function is

$$\rho_l(Z, \theta) = Y_{1l} - h_l\left(Y_2 - X_1'\beta_1\right) - X_1'\beta_{2,l}.$$

with

$$E\left[\rho_l(Z, \theta)|X\right] = E\left[Y_{1l} - h_l\left(Y_2 - X_1'\beta_1\right) - X_1'\beta_{2,l}|X\right] = 0.$$

The unknown function $h_l(\cdot)$ and $\beta_1$, $\beta_{2,l}$ may not be point identified. Suppose the functional of interest is $\phi(\theta) = h_l(0)$: the unknown function evaluated at zero. Then for two unknown functions, $h_l^1(\cdot)$ and $h_l^2(\cdot)$, if $h_l^1(0) \neq h_l^2(0)$, we have $\rho_l(Z, \theta^1) \neq \rho_l(Z, \theta^2)$ and Assumption 4.3 (i) is satisfied. Also notice that since $\phi(\cdot)$ is a linear functional of $\theta$, $\mathbf{F}(\cdot)$ is a linear functional of $\rho$. Thus, Assumption 4.3 (ii) is satisfied.

On the other hand, suppose, for example, we are interested in $\phi(\theta) = \beta_1$ and suppose there are two points $\beta_1^1 \neq \beta_1^2$ such that

$$\rho_l(Z, \theta^1) = \rho_l(Z, \theta^2) \ a.s.Z \tag{4.6}$$

then $h_l(Y_1 - X_1'\beta_1^1) = h_l(Y_1 - X_1'\beta_1^2)$ $a.s.Y_1, X_1,$. Assumption 4.3' captures the idea that in such a case, we can not hope to consider a hypothesis where $\beta_1^1$ fell into the null space and $\beta_1^2$ is in the alternative space. To satisfy Assumption 4.3', we would need (4.6) to imply $\beta_1^1 = \beta_1^2$. One sufficient condition is to assume $h_l(\cdot)$ is strictly monotone and $X_1$ is full-rank for certain functions without assuming $h_l(\cdot)$ is identified. However, this simplification would not hold in general (e.g. $h_l(\cdot) = |\cdot|$). In such cases Assumption 4.3' would fail. □

It follows that for the purpose of testing hypothesis in (4.4), we can define an equivalence class of functions based on the space of residual functions such that

$$\bar{\mathbf{V}} = cl\left\{\nu(Z, \theta) = \left(\rho(Z, \theta) - \rho(Z, \theta_0)\right), \theta \in \mathcal{B}(\theta_0)\right\}$$

and

$$\bar{\mathbf{V}}_n = cl\left\{\nu_n(Z, \theta) = \left(\rho(Z, \theta) - \rho(Z, \theta_{0n})\right) : E_0[\nu(Z)] = 0, \ \theta \in \mathcal{B}_n(\theta_0)\right\},$$

where "$cl$" represents closed linear span.

Let $L_0^2(P_z) = \left\{v(Z, \theta) : E_0\left[v(Z)\right] = 0, \ E_0\left[\left(\nu(Z)\right)^2\right] < \infty\right\}$ be a well-defined Hilbert space. Since $\bar{\mathbf{V}}$ is a subspace of $L_0^2(P_z)$ , $(\bar{\mathbf{V}}, ||\cdot||_{wp})$ is a Hilbert space with the inner product

$$\langle \nu_1, \nu_2 \rangle_{wp} \quad = \quad E\left[E\left[\nu(Z, \theta_1)|X\right]' \Sigma(X)^{-1} E\left[\nu(Z, \theta_2)|X\right]\right]$$

and

$$\|\nu_1 - \nu_2\|_{wp} = \sqrt{E\left[E\left[\nu(Z, \theta_1) - \nu(Z, \theta_2)|X\right]' \Sigma(X)^{-1} E\left[\nu(Z, \theta_1) - \nu(Z, \theta_2)|X\right]\right]}.$$

Note that $||\cdot||_{wp}$ is a "strong" norm on $\bar{\mathbf{V}}$ in the sense that for $\nu \in \bar{\mathbf{V}}$, if $||\nu||_{wp} \neq 0$, then $\nu \neq 0$.

**Lemma 4.2.** *Suppose Assumption 4.1-4.3 hold. Then for $\theta_0 \in \Theta_0$ and for $w \in \mathcal{W}$, there exists a*

*Riesz representer* $\nu^*(\cdot,\theta_0) \in \bar{\mathbf{V}}$ *such that*

$$\left(\frac{d\mathbf{F}(\rho_0)}{d\rho}[\rho-\rho_0]\right)[w] = E\left[E\left[\nu^*(Z,\theta_0,w)|X\right]'\Sigma(X)^{-1}E\left[\rho(Z,\theta)-\rho(Z,\theta_0)|X\right]\right]$$

*and*

$$\left(\frac{d\boldsymbol{\phi}(\theta_0)}{d\theta}[\theta-\theta_0]\right)[w] \;=\; E\left[E\left[\nu^*(Z,\theta_0,w)|X\right]'\Sigma(X)^{-1}E\left[\frac{d\rho(Z,\theta_0)}{d\theta}[\theta-\theta_0]|X\right]\right].$$

Similar to our discussion in Section 3, $\|\nu^*\|$ can be infinity if the functional is irregular. If the later is the case, we consider the approximation on the (finite-dimensional) sieve space such that there exists a $\nu_n \in \bar{\mathbf{V}}_n$ with

$$\left(\frac{d\boldsymbol{\phi}(\theta_{0n})}{d\theta}[\theta-\theta_{0n}]\right)[w] \;=\; E\left[E\left[\nu_n^*(Z,\theta_0,w)|X\right]'\Sigma(X)^{-1}E\left[\frac{d\rho(Z,\theta_{0n})}{d\theta}[\theta-\theta_{0n}]|X\right]\right]. \quad (4.7)$$

We are now using the representations to show the properties of our tests. We impose the following assumption on our functionals of interests, which is a modification of Assumption 3.8 under partial identification.

**Assumption 4.4.** *The following hold uniformly over* $\theta \in \Theta_0^r$*: (i) For* $\boldsymbol{\phi}(\theta) = (\phi_1(\theta),\ldots,\phi_J(\theta))$*,* $\frac{\partial\phi_j(\theta_0)}{\partial\theta}[\cdot]$ *is a linear functional in the direction* $\theta-\theta_0$ *for* $j=1,\ldots,J$ *and is linearly independent across* $j$*; (ii) for all* $w \in \mathcal{W}$*,*

$$\left|\left(\frac{d\boldsymbol{\phi}(\theta_0)}{d\theta}[\theta_{0n}-\theta_0]\right)[w]\right| / \left\|\nu_n^*(\cdot,\theta_0,w)\right\|_{wp} = o\left(n^{-1/2}\right)$$

*and*

$$\left|\{\boldsymbol{\phi}(\theta)[w]-\boldsymbol{\phi}(\theta_0)[w]-\left(\frac{d\boldsymbol{\phi}(\theta_0)}{d\theta}[\theta-\theta_0]\right)[w]\right| / \left\|\nu_n^*(\cdot,\theta_0,w)\right\|_{wp} = o\left(n^{-1/2}\right)$$

*for all* $\theta \in \mathcal{B}_n(\theta_0)$*.*

Assumption 4.4 is similar to Assumption 4.1 in Chen, Tamer and Torgovitsky (2011). It controls the nonlinearity bias of $\phi(\cdot)$ and sieve approximation error of $\theta_{0n}$. It also imposes an under-smoothing condition.

We suggest to employ the quasi-likelihood ratio statistic we used under point identification to construct (pointwise) confidence regions of parameters of interest. For the criterion function $\bar{L}_n(\theta)$ defined in (4.2), the quasi-likelihood ratio statistic is

$$\mathrm{QLR}_n(\mathbf{r}) = n\left(\inf_{\theta\in\Theta_n\cap\{\boldsymbol{\phi}(\theta)=\mathbf{r}\}}\bar{L}_n(\theta) - \inf_{\theta\in\Theta_n}\bar{L}_n(\theta)\right). \quad (4.8)$$

For $\theta \in \Theta_0^r$ and $w \in \mathcal{W}$, let the sample variance be

$$\|\nu_n^*(\theta, w)\|_{sd}^2 = \text{Var}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n E\left[\nu_n^*(Z_i, \theta, w)|X_i\right]' \Sigma(X_i)^{-1} \rho(Z_i, \theta)\right)$$

and the studentized Riesz representer be

$$\mu_n^*(\theta, w) = \nu_n^*(\theta, w)/||\nu_n^*(\theta, w)||_{sd}.$$

We next establish that the quasi-likelihood ratio statistic has a tight but not pivotal limiting distribution under the null hypothesis. And the QLR statistic does not have a drifting term in the limit like the ones in Santos (2012) or Hong (2013).

**Theorem 4.1.** *For any $r \in \mathbf{R}^J$. Suppose that Assumptions 2.1, 3.2, 3.3, 4.1-4.4, Assumptions A.20-A.25 hold. Then for fixed $w \in \mathcal{W}$,*

$$\begin{aligned} QLR_n(\mathbf{r}) &= \inf_{\theta \in \Theta_0 \cap R} \left\{ \frac{1}{||\mu_n^*(\theta, w)||_{wp}} \times \frac{1}{\sqrt{n}} \sum_{i=1}^n E\left[\mu_n^*(Z_i, \theta, w)|X_i\right]' \Sigma(X_i)^{-1} \rho(Z_i, \theta) \right\}^2 + o_{p^\star}(1) \\ &\Rightarrow \inf_{\theta \in \Theta_0 \cap R} \left\{ \frac{1}{||\mu_n^*(\theta, w)||_{wp}} \times \mathbf{G}(\cdot, \theta) \right\}^2, \end{aligned}$$

*where we denote $\mathbf{G}(\cdot, \theta)$ to be a tight centered Gaussian process indexed by $\theta$.*

**Remark.** Note that if point identification happens to hold, the limiting distribution in Theorem 4.1 reduces to a weighted chi-squared distribution with degrees of freedom $J$. It can not be reduced to a standard chi-squared with identity weight because the test is presented for an arbitrary weight matrix (which would not be optimal in general). To consistently estimate the asymptotic distribution of the QLR statistic, we propose a computationally simple bootstrap procedure to obtain the critical values for the asymptotic distribution in next theorem. □

Once the asymptotic properties in Theorem 4.1 are established, the multiplier bootstrap can be verified immediately. When point identification fails and the limiting distribution of the test statistic is not pivotal, it is not new in the literature to use bootstrap methods to calculate the critical values of the test statistic. For example, Hansen (1996) proposes a multiplier bootstrap procedure for a class of parametric models where a nuisance parameter is not identified under the null. Chen, Tamer and Torgovitsky (2011) propose sieve LR bootstrap for partially identified semiparametric likelihood models. And Chen, Pouzo and Tamer (2011) propose a sieve bootstrap procedure for partially identified semi/nonparametric conditional moment restriction models based on a minimum distance criterion.

We define the multiplier bootstrap draw of the SGMM estimator $\hat{\theta}_n^\star$ as a solution to the following

criterion weighted by $\{\zeta_i\}_{i=1}^n$

$$\bar{L}_n^\star(\theta) = \left(\frac{1}{n}\sum_{i=1}^n \zeta_i \rho(Z_i,\theta)\otimes q^{s_n}(X_i)\right)' \left(\frac{1}{n}\sum_{i=1}^n \Sigma(X_i)\otimes q^{s_n}(X_i)q^{s_n}(X_i)'\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^n \zeta_i \rho(Z_i,\theta)\otimes q^{s_n}(X_i)\right). \tag{4.9}$$

The multiplier bootstrap we consider consists of i.i.d. positive random weights applied to every observation.[15] The bootstrap weights satisfies the following condition.

**Assumption 4.5.** *Let $\{\zeta_i\}_{i=1}^n$ is an i.i.d. sequence, independent of $\{X_i, Z_i\}_{i=1}^n$ and satisfying $E[\zeta_i] = 1$, $E[\zeta_i^2] = v_0^2 < \infty$ and $\int_0^\infty \sqrt{\Pr(|\zeta - 1| \geq \varepsilon)}d\varepsilon < \infty$.*

In particular, we select $\{\zeta_i\}_{i=1}^n$ to be i.i.d. draws from the standard exponential distribution with $E[\zeta_i] = 1$, $\mathrm{Var}(\zeta_i) = 1$. The choice of such weights is only for ease of exposition.

For each draw of such weights, for $\hat{\mathbf{r}} = \boldsymbol{\phi}(\hat{\theta}_n)$, let

$$\mathrm{QLR}_n^\star(\hat{\mathbf{r}}) = n\left(\inf_{\theta\in\Theta_n\cap\{\boldsymbol{\phi}(\theta)=\boldsymbol{\phi}(\hat{\theta}_n)\}} \bar{L}_n^\star(\theta) - \inf_{\theta\in\Theta_n}\bar{L}_n^\star(\theta)\right)$$

be the bootstrap sieve QLR statistic. To validate the use of the multiplier bootstrap, we provide the following theorem.

**Theorem 4.2.** *(Multiplier Bootstrap) Let $\bar{L}_n^\star(\theta)$ be defined by (4.9). Let $\hat{\theta}_n$ be the minimizer of $\bar{L}_n(\theta)$ over $\Theta_n$ defined in (4.8). Suppose the Assumptions of Theorem 4.1 and Assumption 4.5 hold. Then for fixed $w \in \mathcal{W}$,*

$$\begin{aligned} & QLR_n^\star(\hat{\mathbf{r}}) \\ = & \inf_{\theta\in\Theta_0\cap R}\left\{\frac{1}{||\mu_n^*(\theta,w)||_{wp}}\times\frac{1}{\sqrt{n}}\sum_{i=1}^n(\zeta_i-1)E[\mu_n^*(Z_i,\theta,w)|X_i]'\Sigma(X_i)^{-1}\rho(Z_i,\theta)\right\}^2 + o_{p^\star}(1) \\ \Rightarrow & \inf_{\theta\in\Theta_0\cap R}\left\{\frac{1}{||\mu_n^*(\theta,w)||_{wp}}\times\mathbf{G}(\cdot,\theta)\right\}^2 \end{aligned}$$

*where $\mathbf{G}(\cdot,\theta)$ is a tight centered Gaussian process indexed by $\theta$.*

**Remark.** (1) We can apply Theorem 4.2 to construct confidence sets for $\boldsymbol{\phi}(\theta)$ such that

$$\mathcal{C}_n = \{\mathbf{r} : \mathrm{QLR}_n(\mathbf{r}) \leq \hat{c}_n(\mathbf{r}, 1-\tau)\}$$

where $\hat{c}_n(\mathbf{r}, 1-\tau)$ is $(1-\tau)$th quantile using the multiplier bootstrap such that

$$\hat{c}_n(\mathbf{r}, 1-\tau) = \inf\left\{u : \frac{1}{B}\sum_{b=1}^B 1\left\{\mathrm{QLR}_{n,b}^\star(\hat{\mathbf{r}}) \leq u\right\} \geq 1-\tau\right\}.$$

[15] In contrast, the random vector of observation is weighted by multinomial $(n, n^{-1}, \ldots, n^{-1})$ for the nonparametric bootstrap. The weights are exchangeable but not independent.

(2) Our limiting theories in Theorem 4.1 and Theorem 4.2 hold pointwisely when we fix a distribution $P_0$. When $\theta_0$ lies on the boundary of the parameter space, although our pointwise results still holds, it is not clear if they hold uniformly over a sequence of $P_n$.

(3) One of the reasons we focus on the multiplier bootstrap in this paper is that the i.i.d. behavior of the weights simplifies the the proof of Theorem 4.2. While this result does not rule out the possibility that the bootstrap may still work in our case, we leave such exploration for future work. □

### 4.2 Uniform Inference Method

Next we return to the case considered in Section 3.2. We want to provide methods of inference that are, for instance, uniform over the arguments $w$ of functionals of interest under partial identification. Generally speaking, we provide an inference procedure to construct uniform confidence set for $\phi(\theta)[w]$ uniformly over $w$. Similar to the point identified model, the sequence of empirical sieve processes are indexed by $k_n$. Hence, they may not be stochastically equicontinuous. Due to the lack of asymptotic equicontinuity, we employ strong approximations and approximate the test statistic process by a sequence of Gaussian processes that can be used to construct a uniform confidence set. For the purpose of considering inference uniformly over $w$, we strengthen our restrictions on the functionals we consider as follows.

**Assumption 4.6.** *Assume that (i) $\forall \theta_1, \theta_2 \in \Theta$, if $\rho(Z, \theta_1) = \rho(Z, \theta_2)$, then $\phi(\theta_1) = \phi(\theta_2)$; (ii) There exists a mapping $\mathbf{F}(\rho(\cdot, \theta)) = \phi(\theta)$ that is differentiable at any point $\theta_0 \in \Theta_0$ so $\left(\frac{d\mathbf{F}(\rho(\cdot,\theta_0))}{d\theta}[\rho - \rho_0]\right)[w]$ exists for all $w \in \mathcal{W}$ and $\theta \in \mathcal{B}(\theta_0)$.*

Assumption 4.6 is stronger than Assumption 4.3, however, the uniform limiting theory is also a stronger result. And note that Assumption 4.3' still implies Assumption 4.6.

Next we show that the entire $\sup - \mathrm{QLR}_n$ process can be uniformly close to the suprema of a sequence of Chi-square processes of the stated form. The $\sup - \mathrm{QLR}_n$ statistic under partial identification is

$$\sup_w \{\mathrm{QLR}_n(w)\} = \sup_w \left\{ n \left[ \inf_{\theta \in \Theta_n \cap \{\boldsymbol{\phi}(\theta)[w] = \mathbf{r}(w)\}} \bar{L}_n(\theta) - \inf_{\theta \in \Theta_n} \bar{L}_n(\theta) \right] \right\}.$$

Let the class of score functions be

$$\mathcal{S}_n \equiv \left\{ S_n(\cdot, \theta, w) = E\left[\mu_n^*(Z_i, \theta, w) | X_i\right]' \Sigma(X_i)^{-1} \rho(Z_i, \theta) : \theta \in \Theta_0 \cap R, w \in \mathcal{W} \right\}. \tag{4.10}$$

We provide the following results for uniform inference under partial identification.

**Theorem 4.3.** *Suppose that the restricted identified set $\Theta_0 \cap R$ is a compact space under $\|\cdot\|_s$. Suppose that Assumptions 2.1, 3.2, 3.3, 4.1-4.2, Assumption 4.6, 4.4 hold and suppose that Assumptions A.20-A.26 hold for all $w \in \mathcal{W}$, then*

*(i)*

$$\begin{aligned}
&\sup_{w\in\mathcal{W}}\left\{QLR_n(r(w))\right\}\\
=\;&\sup_{w\in\mathcal{W}}\left\{\inf_{\theta\in\Theta_0\cap\{\boldsymbol{\phi}(\theta)[w]=\mathbf{r}(w)\}}\left(\frac{1}{||\mu_n^*(\theta,w)||_{wp}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}S_n(\cdot,\theta,w)\right)^2\right\}+o_p(1)\\
\overset{d}{=}\;&\sup_{w\in\mathcal{W}}\left\{\inf_{\theta\in\Theta_0\cap\{\boldsymbol{\phi}(\theta)[w]=\mathbf{r}(w)\}}\left(\frac{1}{||\mu_n^*(\theta,w)||_{wp}}\times\mathbf{G}\left[S_n(\cdot,\theta,w)\right]\right)^2\right\}+o_p(1).
\end{aligned}$$

*where* $\mathbf{G}\left[S_n(\cdot,\theta,w)\right]$ *is a sequence of Gaussian processes with continuous paths almost surely. It has covariance function* $E\left[S_n(t_1)S_n(t_2)\right]-E\left[S_n(t_1)\right]E\left[S_n(t_2)\right]$ *that is uniformly non-degenerate in* $k_n$ *and is uniformly Hölder on* $\Theta_0^r\times\mathcal{W}$ *with* $t=(\theta,w)\in\Theta_0^r\times\mathcal{W}\equiv\mathcal{T}$.

*(ii) Furthermore, suppose Assumption 4.5 holds. For* $\hat{\mathbf{r}}(w)=\boldsymbol{\phi}(\hat{\theta}_n)[w]$. *The bootstrap process has*

$$\begin{aligned}
&\sup_{w}\left\{QLR_n^\star(\hat{\mathbf{r}}(w))\right\}\\
=\;&\sup_{w}\left\{\inf_{\theta\in\Theta_0\cap\{\boldsymbol{\phi}(\theta)[w]=\hat{\mathbf{r}}(w)\}}\left(\frac{1}{||\mu_n^*(\theta,w)||_{wp}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[(\zeta_i-1)\,S_n(\cdot,\theta,w)\right]\right)^2\right\}+o_{p^\star}(1)\\
\overset{d}{=}\;&\sup_{w}\inf_{\theta\in\Theta_0\cap\{\phi(w)=\mathbf{r}(w)\}}\left(\frac{1}{||u_n^*(\theta,w)||_{wp}}\tilde{\mathbf{G}}\left[S_n(\cdot,\theta,w)\right]\right)^2+o_{p^\star}(1),
\end{aligned}$$

*where* $\tilde{\mathbf{G}}\left[S_n(\cdot,\theta,w)\right]$ *is a sequence of Gaussian processes with the same distributions as the process* $\mathbf{G}\left[S_n(\cdot,\theta,w)\right]$.

**Remark.** Similar to the pointwise inference case, in order to get a critical value to control the size of the test, we can employ the $(1-\tau)$th quantile of $\mathrm{QLR}_n^\star\left(\hat{\mathbf{r}}(w)\right)$ by defining

$$\hat{c}_n(1-\tau)\equiv\inf\left\{\alpha:P\left(\sup_w\left\{\mathrm{QLR}_n^\star\left(\hat{\mathbf{r}}(w)\right)\right\}\leq\alpha\right)\geq1-\tau\right\}.$$

## 5 MONTE CARLO

In this section, we investigate the finite sample properties of the proposed inference methods. We consider two simulation experiments. The first simulation design is based on Horowitz (2012). And the second one is based on Santos (2012). There are 1000 Monte Carlo replications in each experiment.

**Experiment 1:**

Consider the following data generating process (DGP):

$$\begin{aligned} \text{DGP1}: \ Y_i &= \beta X_{1i} + \sum_{j=1}^{100} (-1)^{j+1} j^{-2} \sin(j\alpha\pi Y_{1i}) + 0.3e_i, \\ Y_{1i} &= \Phi(u_{1i} + u_{2i}), \ X_{1i} \sim \text{Uniform}(0,1), \\ X_{2i} &= \Phi(u_{1i}), \ e_i = \lambda u_{2i} + (1-\lambda) u_{3i}, \end{aligned}$$

where $\Phi(\cdot)$ is the CDF of normal distribution, and $u_{1i}$, $u_{2i}$ and $u_{3i}$ are generated from independent standard normal distributions. The parameter $\alpha$ controls the wave length of the sine function and the parameter $\lambda$ controls the degree of endogeneity. We set our sample size $n$ to be 500 and 1000.

Let $\rho(Z_i, \theta) = Y_i - \beta X_{1i} - \sum_{j=1}^{100}(-1)^{j+1} j^{-2} \sin(j\alpha\pi Y_{1i})$, $h_0(Y_{1i}) = \sum_{j=1}^{100}(-1)^{j+1} j^{-2} \sin(j\alpha\pi Y_{1i}))$ with

$$E\left[ Y - \beta_0 X_1 - \sum_{j=1}^{100} (-1)^{j+1} j^{-2} \sin(j\alpha\pi Y_1) | X \right] = 0.$$

We follow Horowitz (2012) to assume the model is point identified, although it is worth noting that the $\text{QLR}_n$ is robust to partial identification. For each simulation, we use $\text{Pen}(h) = ||h||^2_{L^2} + ||\bigtriangledown h||^2_{L^2}$ and $\lambda_n = 0.0005$.

The basis functions (for the instrument $X_2$ and the unknown function) we choose are third order polynomial splines. We set the order of the basis functions to be $k_n = 7$ for $p^{k_n}(\cdot)$ ( the basis function for the unknown function) and $s_n = 9$ for $q^{s_n}(\cdot)$ (the basis function for the instrument).[16] We set $\beta_0 = 0$ or $\beta_0 = 1$ to consider both nonparametric IV model and partially-linear IV model. We consider three different null hypotheses. The first one is to test the unknown function at different points for a nonparametric IV model (when we set $\beta_0 = 0$). The second and third ones are to test the joint hypotheses on both the parametric component and the unknown function for partially linear IV model (when we set $\beta_0 = 1$). We first want to test the joint hypothesis of $\beta_0$ and the unknown function evaluated at different points. Then we consider testing the joint hypothesis of $\beta_0$ and the derivatives of the unknown function evaluated at different points.

Table 1 reports the simulated size of $t_n$, $\text{Wald}_n$ and $\text{QLR}_n$ for pointwise hypothesis tests. We set $\alpha = 3$, $\lambda = 0.2, 0.8$. For the nonparametric IV model ($\beta_0 = 0$), we set $\phi(h) = h(y_1)$ for the $25th$, $50th$, $75th$ quantiles of $Y_1$ (these points are fixed for each simulation). We compare the sizes of $t_n$ and $\text{QLR}_n$. We set the nominal sizes to be $\tau = 0.05$ and $\tau = 0.1$. The sizes of the tests are close to the nominal ones, and are not sensitive to the choices of different statistics or different degrees of endogeneity evaluated by $\lambda$. For the partially-linear IV model ($\beta_0 = 1$), for the two joint hypotheses we consider, we compare the sizes of $\text{Wald}_n$ and $\text{QLR}_n$. The performances of $\text{QLR}_n$ are relatively better than the ones of $\text{Wald}_n$. They are close to the nominal sizes and are not sensitive to the degrees of endogeneity.

Table 2 reports uniform inference results under point identification for the nonparametric IV

[16] We have also tried some other different combinations of $k_n$ and $s_n$ and got similar results. Liu and Tao (2014) propose a simple Mallows' criterion to select the combination of $k_n$ and $s_n$ simultaneously.

model ($\beta_0 = 0$) and the partially-linear IV model ($\beta_0 = 1$). The critical values for uniform tests are obtained by multiplier bootstrap with 500 replications for each.

We report the empirical coverages of the confidence bands with nominal level of 90% and 95%, respectively. The confidence bands have empirical coverages close to the nominal levels and are not sensitive to the choices of $\alpha$ or $\lambda$. In general, performances are improved when sample sizes increase.

**Experiment 2:**

In the second experiment, we assume that

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

and define $Y_1 = 2\left(\Phi(Y_1^*/3) - 0.5\right)$, $X = 2\left(\Phi(X^*/3) - 0.5\right)$, $e = \varepsilon^*$, where $\Phi(\cdot)$ is the c.d.f. of a standard normal distribution.

Consider the following relationship

$$DGP2: \; Y \;\; = \;\; 2\sin(Y_1\pi) + e,$$

so $\rho(Z,\theta) = \rho(Z,h) = Y - 2\sin(Y_1\pi)$ with $E\left[\rho(Z,h)|X\right] = 0$.

Santos (2012) argues that the unknown function $h(Y_1) = 2\sin(Y_1\pi)$ for this DGP may not be point identified. We follow Santos (2012) to define the parameter space as a compact space such that $\Theta = \mathcal{H} \equiv cl\left(\left\{\theta : \|\theta\|_{2,2} \leq B\right\}\right)$. The choice of $B$ measures the space of $\Theta$. Following the setup in Section 3.4 of Santos (2012), we consider three different choices of $B$ ($B$ can be 50, 100 or 1000). The null hypothesis is $\mathbb{H}_0 : \mathcal{H}_0 \cap \{\sin(0) = 0\}$.

Table 3 presents the simulated size of the multiplier bootstrap QLR-test we suggest and the J-test from Santos (2012) as a function of nominal size $\tau$. The basis functions are chosen to be cubic-splines with $k_n = 6$ and $s_n = 8$.[17] To implement the J-test, we also need to choose norm constraints $B_n$ for bootstrap. In contrast, our QLR-test does not require such choice of $B_n$. Overall, when $B_n = 50$, J-test provides good size control, however, when $B_n = 100$ or $B_n = 1000$, there are size distortions in most specifications. On the other hand, the QLR test has good size control when the nominal size $\tau$ varies from 0.01, 0.05 to 0.1.

Figure 1 gives the QLR test's rejection probability for the null hypothesis $\mathbb{H}_0 : \mathcal{H}_0 \cap \{\sin(0) = \gamma\}$ as a function of $\gamma \in [-0.5, 0.5]$. We choose $B$ to be 100 or 1000. Notice that a larger choice of norm constraint seems to decrease the power of the test for alternatives far away from the null. Compared with the power provided in Figure 1 of Santos (2012), the power performance of the QLR-test is better than the J-test. For example, when $B = 100$ and $\gamma = 0.4$, the power of the QLR-test is above 0.8 while the power of the J-test is below 0.5.

[17] For the J-test, in Table 3, we present the number provided in Santos (2012).

## 6 EMPIRICAL APPLICATION

In this section we study a shape-invariant Engel curve system with endogenous total expenditure (Blundell, Chen and Kristensen, 2007; BCK, henceforth). Engel curves describe the relationship between consumer expenditure on a particular good or service and the consumer's total resources holding prices fixed. It can be regarded as the Marshallian demand function conditioning on the prices of all goods fixed (Lewbel, 2008). Much of the evidence in the literature has let to the recommendation to estimate Engel curves in a nonparametric way without imposing a functional form *a priori* (see, for example, Banks, Blundell and Lewbel, 1997; Imbens and Newey, 2009; Horowitz, 2011).

BCK have argued that the total expenditure could be jointly determined with individual demands. Given the potential endogeneity of the total expenditure variable, we follow BCK's suggestion to use exogenous sources of income as suitable instrumental variables for total expenditure.[18]In our analysis, we use the gross earnings of the household head as an instrument for total expenditure.

The data we use is the 1995 British Family Expenditure Survey (FES). The data is a subset of married and cohabiting couples. The age of the household head is between 20 and 50. Households where the head of household is unemployed are excluded. The data also excludes those couples with three or more children. The income of the head for each household (IV) is measured by the amount he earned in the chosen year before taxes. BCK provide the asymptotic distribution for parametric estimates $(\hat{\beta}_1, \hat{\beta}_2)$. In the current paper, our focus is to provide pointwise and uniform confidence bands for the Engel curves by employing a sieve GMM method.

Suppose each household $i$ faces the same relative prices. Let $Y_{il}$ be the budget share of good $l = 1, \ldots, L$ for each household $i$. Let $Y_{1i}$ be the log of total expenditure and $X_{1i}$ be a vector of household demographic variables. Let $\{(Y_{1il}, Y_{2i}, X_{1i})\}_{i=1}^n$ be i.i.d. observations. Blundell, Browning and Crawford (2003) have argued that the model that is consistent with consumer optimization theory should have the form

$$Y_{1il} \quad = \quad h_l\left(Y_{2i} - X_{1i}'\beta_1\right) + X_{1i}'\beta_{2,l} + e_{il}.$$

where $h_l, l = 1, \ldots, L$ are unknown. We allow for the possibility that $E\left[e_{il}|Y_{2i}\right] \neq 0$, i.e., the total expenditure is endogenous. Let the gross earnings of the head of household be the instrument and be denoted $X_{2i}$. We assume $E\left[e_{il}|X_{1i}, X_{2i}\right] = 0,\ l = 1, \ldots, L$. Thus, $\rho_l\left(Z_i, \theta\right) = Y_{1il} - h_l\left(Y_{2i} - X_{1i}'\beta_1\right) - X_{1i}'\beta_{2l}$ with $E\left[\rho_l(Z_i, \theta)|X_i\right] = 0$ for $l = 1, \ldots, L$.

The two-step SGMM criterion allows us to choose the preliminary estimator $\bar{\theta}_n$ in a flexible way. We use the profiling approach suggested by BCK to get the $\bar{\theta}_n$. In the first step, we fixed $\beta$,

[18]Imbens and Newey (2009) have proposed an alternative ways to estimate Engel curves in a nonseparable model by considering a triangular simultaneous equations model and using control function approach.

approximate $h_{l,n}(\beta) = p^{k_n}(\cdot)'\gamma_n^l(\beta)$, and compute $\bar{h}_{l,n}(\beta)$ by two-stage least squares such that

$$\bar{\gamma}_n^l(\beta) = \left(P(\beta_1)'Q(Q'Q)^{-1}Q'P(\beta_1) + \lambda_n \text{Pen}(h_n)\right)^- P(\beta_1)'Q(Q'Q)^-Q'Y_{1l}(\beta_{2,l}).$$

Then $\bar{h}_{l,n}(\beta) = p^{k_n}(\cdot)'\bar{\gamma}_n^l(\beta)$. Then we plug $\bar{h}_n(\beta;\cdot) = \left(\bar{h}_{1,n}(\beta;\cdot),\ldots,\bar{h}_{L,n}(\beta;\cdot)\right)'$ into the GMM criterion and calculate

$$\min_{\beta\in\mathcal{B}} \sum_{l=1}^{L} \left(Y_{1l}(\beta_{2,l}) - P(\beta_1)'\bar{\gamma}_n^l\right)' Q(Q'Q)^-Q' \left(Y_{1l}(\beta_{2,l}) - P(\beta_1)'\bar{\gamma}_n^l\right).$$

In next step, we plug $\bar{\theta}_n = (\bar{\beta}, p^{k_n}(\cdot)'\bar{\gamma}_n(\bar{\beta}))$ into the criterion and obtain $\hat{\theta}_n$ by minimizing over the criterion function with optimal weight matrix

$$\min_{(\beta,h_n)\in\Theta_n} \hat{g}(\theta)' \left(\frac{1}{n}\sum_{i=1}^{n} g_i\left(\bar{\theta}_n\right) g_i\left(\bar{\theta}_n\right)\right)^{-1} \hat{g}(\theta) + \lambda_n \text{Pen}(h).$$

In Figure 2, we report the estimates of Engel curves for different categories of nondurable goods and services, along with their pointwise and uniform confidence bands. To construct the pointwise confidence bands, we let $\hat{w}_{il} = y_{1i} - x_{1i}'\hat{\beta}_1$, for $l = 1,\ldots,L$,

$$\begin{aligned}
\hat{\sigma}_{n,l} &= p^{k_n}(y_2)'\hat{\Omega}_h p^{k_n}(y_2)/n, \\
\hat{\Omega}_n &= \left(\left(\frac{1}{n}\sum_{i=1}^{n} p^{k_n}(\hat{w}_{il}) \otimes q^{s_n}(x_i)\right)' \left(\frac{1}{n}\sum_{i=1}^{n} g_i\left(\hat{\theta}_n\right) g_i\left(\hat{\theta}_n\right)'\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} p^{k_n}(\hat{w}_{il}) \otimes q^{s_n}(x_i)\right)\right)^{-1}.
\end{aligned}$$

Then the 95% pointwise confidence bands are simply

$$\left[\hat{h}_{n,l}(y_2) - 1.96\hat{\sigma}_{n,l},\ \hat{h}_{n,l}(y_2) + 1.96\hat{\sigma}_{n,l}\right].$$

To construct uniform confidence bands, for $b = 1,\ldots,2000$, we compute the sup $-t_n$ statistic by

$$\sup_{y_2\in\mathcal{Y}_2} \left|\hat{t}_n^{*,b}(y_2)\right| = \sup_{y_2\in\mathcal{Y}_2} \left|\frac{p^{k_n}(y_2)'\hat{\Omega}_n^{1/2}\mathcal{N}_{d_\beta+k_n}^b}{\sqrt{n}\hat{\sigma}_{n,l}}\right|.$$

Then we form a 95%- uniform confidence bands as

$$\left[\hat{h}_{n,l}(y_2) - c_n(0.95)\hat{\sigma}_{n,l},\ \hat{h}_{n,l}(y_2) + c_n(95)\hat{\sigma}_{n,l}\right], y_2 \in \mathcal{Y}_2,$$

where $c_n(0.95)$ is the 95% sample quantile of $\left\{\sup_{y_2\in\mathcal{Y}_2}\left|\hat{t}_n^{*,b}(w)\right| : 1 \le b \le B\right\}$. For the support of $y_1$, $\mathcal{Y}_1$, we restrict $\mathcal{Y}_2 = [4.75, 6.178]$, where 4.75 is the 5% quantile of the distribution of $y_2$ and 6.578 is the 99% quantile of the distribution of $y_1$.

Each Engel curve's pointwise confidence intervals and uniform confidence bands are reported

in Figure 2. From Figure 2, we find that the shape of the curves are similar to the ones obtained in BCK. The confidence bands are tight along the estimates of the curves. When households' total expenditure increases, they tend to spend proportionally less on necessary goods such as food, fuel, alcohol, motor and more on goods and services like food at restaurants and leisure goods. For some categories of the goods such as alcohol or food at restaurants, from the uniform confidence bands we obtain, we cannot reject the hypothesis that the curves are constant ones. We also find that the confidence bands and confidence intervals tend to be more narrow in the middle when the data has more observations and wider at the two ends of the curves when the data has less observations.

Figure 3 reports the pointwise confidence intervals by QLR and the uniform confidence intervals by sup-QLR by using the non-optimal weight matrix, respectively. The two tests (discussed in Section 4) are robust to partial identification and do not require estimates of the Engel curves. We report the bands for food, fuel and motor by using different choices of $(s_n, k_n)$. We choose 100 points over $[4.75, 6.178]$ evenly and consider 200 bootstrap repetitions for each. The shape of the confidence bands are not very sensitive to our choices of the orders of basis functions and are, in general, wider than the ones obtained in Figure 2. We cannot reject the hypothesis that the curves are linear ones.

## 7 CONCLUSION

This paper studies the problem of pointwise and uniform inference for semi-nonparametric conditional moment restriction models. Our parameter of interest contains both a parametric component and a nonparametric component of the parameter. Under point identification, we first provide pointwise asymptotic results for functionals of sieve GMM estimators regardless of whether the functionals are $\sqrt{n}-$estimable or not. Then we extend the pointwise asymptotic results to the entire support of the functionals and develop a uniform limiting theory for functionals of interest.

We provide formal conditions that justify a strong approximation of functionals of sieve GMM estimators to a sequence of Gaussian processes uniformly over the support of the functionals. This approximation essentially provides a functional central limit theorem for functionals of sieve GMM estimators. We propose a uniform version of the three main classes of test statistics for hypotheses on restrictions uniformly over the support of functionals. We show that sup-Wald, sup-QLR and sup-LM are asymptotically equivalent. These results are useful to construct uniform confidence bands for functionals of the parameters.

We then relax the point identification assumption and consider models that allow for partial identification. We first provide consistency and nonparametric convergence rates for set estimates of the identified set based on a PSGMM criterion. To do inference on restrictions of functionals of the parameters in the identified set, we focus on a general class of conditional moment restriction models. We show that, based on a non-optimally weighted SGMM criterion function, the sieve QLR inference is robust to partial identification. The limiting distribution of the sieve QLR

under partial identification is the infimum of the square of a weighted Gaussian process. We then provide a valid multiplier bootstrap procedure to obtain critical values and invert the test statistic to get confidence regions. We further show that the sup-QLR statistic is also robust to partial identification if we consider hypotheses uniformly over the support of functionals.

The inference methods we propose are easy to compute and are analogous to implementations in parametric GMM models. Numerical evidence shows that our methods are promising for applications. In our empirical application, we provide confidence intervals and confidence bands for the Engel curve systems for different categories of nondurable goods and services by using 1995 British Family Expenditure Survey.

Finally, we point out some possible extensions that we do not consider in this paper. First, little work has been done to discuss the choices of number of instruments and regressors. In a recent work, Liu and Tao (2014) have proposed a simple Mallows' criterion to select the number of instruments and the number of regressors in NPIV model. However, it is not clear how to choose them in general semi-nonparametric conditional restriction models with nonlinear or even nonsmooth residual functions. Second, there are still some important theoretical questions that remain to be answered. It is not clear how the estimation and inference procedures would adjust if only weak IVs are available. And we should consider the uniformity issue over the data generating processes of our test statistics under partial identification. We leave these extensions for future works.

## References


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

[2] Andrews, D. W. K. (1991): "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Models," *Econometrica*, 59, 307-345.

[3] Andrews, D. W. K. (2011): "Examples of $L^2-$Complete and Boundedly-Complete Distributions," Yale University Cowles Foundation Discussion Paper 1801.

[4] Andrews, D. W. K. and P. J. Barwick (2012): "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure," *Econometrica*, 80, 2805-2826.

[5] Andrews, D. W. K. and X. Cheng (2012): "Estimation and Inference with Weak, Semi-strong, and Strong Identification," *Econometrica*, 80, 2153-2211.

[6] Andrews, D. W. K. and P. Guggenberger (2009): Validity of Subsampling and "Plug-in Asymptotic" Inference for Parameters Defined by Moment Inequalities," *Econometric Theory*, 25, 669-709.

[7] Andrews, D.W. K., M. J. Moreira and J. H. Stock (2006): "Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression," *Econometrica,* 74, 715-752.

[8] Andrews, D.W. K. and Shi (2013): "Inference Based on Conditional Moment Inequalities," *Econometrica*, 81, 609-666.

[9] Andrews, D. W. K. and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78, 119-158.

[10] Banks, J., R. Blundell, and A. Lewbel (1997): "Quadratic Engel Curves and Consumer Demand," Review of Economics and Statistics, 79, 527-539.

[11] Belloni, A. and V. Chernozhukov (2011): "$\ell_1-$Penalized Quantile Regression in High-Dimensional Sparse Models," *The Annals of Statistics*, 39, 82-130.

[12] Belloni, A., V. Chernozhukov, D. Chetverikov, and K. Kato (2013): "On the Asymptotic Theory for Least Squares Series: Pointwise and Uniform Results," *arXiv: 1212.0442v2.*

[13] Belloni, A., V. Chernozhukov, and Fernandez-Val (2011): "Conditional Quantile Process Based on Series or Many Regressors," *arXiv: 1105.6154v1.*

[14] Bierens, H. H. (1990): "A Consistent Conditional Moment Test of Functional Form," *Econometrica*, 58, 1443-1458.

[15] Blundell, R., M. Browning, and I. Crawford (2003): "Nonparametric Engel Curves and Revealed Preference," *Econometrica*, 71, 205-240.

[16] Blundell, R., X. Chen and D. Kristensen (2007): "Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves," *Econometrica*, 75, 1613-1669.

[17] Bugni, F. A., I. A. Canay and S. Shi (2014): "Inference for Functions of Partially Identified Parameters in Moment Inequality Models," working paper.

[18] Canay, I. A. (2010): "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," *Journal of Econometrics*, 156, 408-425.

[19] Canay, I. A., A. Santos and A. M. Shaikh (2013): "On the Testability of Identification in Some Nonparametric Models with Endogeneity," *Econometrica*, 81, 2535-2559.

[20] Chamberlain, G. (1987): "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," *Journal of Econometrics* 34, 305-334.

[21] Chamberlain, G. (1992): "Efficiency Bounds for Semiparametric Regression," *Econometrica*, 60, 567-596.

[22] Chandrasekhar, A. V. Chernozhukov, F. Molinari and P. Schrimpf (2012): "Inference for Best Linear Approximations to Set Identified Functions," *arXivv:1212.5627.*

[23] Chen, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of Econometrics 6B*, 5549-5632, ed. by J. J. Heckman and E. E. Leamer. Amsterdam: North Holland.

[24] Chen, X., V. Chernozhukov, S. Lee, and W. K. Newey (2014): "Local Identification of Nonparametric and Semiparametric Models," *Econometrica*, 82, 785-809.

[25] Chen, X. and T. M. Christensen (2013): "Optimal Uniform Convergence Rates for Sieve Nonparametric Instrumental Variables Regression, " *arXiv: 1311.0412*.

[26] ——- (2014): "Optimal Sup-norm Rate, Adaptive Estimation, and Inference on NPIV", work in progress.

[27] Chen, X., Z. Liao, and Y. Sun (2014): "Sieve Inference on Possibly Misspecified Semi-Nonparametric Time series Models," *Journal of Econometrics*, 178, 639-658.

[28] Chen, X., O. Linton and I. van Keilegom (2003): "Estimation of Semiparametric Models When the Criterion Functions Is Not Smooth," *Econometrica*, 71, 1583-1600.

[29] Chen, X. and D. Pouzo (2009): "Efficient Estimation of Semiparametric Conditional Moment Models with Possibly Nonsmooth Residuals," *Journal of Econometrics*, 152, 46-60.

[30] Chen, X. and D. Pouzo (2012): "Estimation of Nonparametric Conditional Moment Models With Possibly Nonsmooth Moments," *Econometrica*, 80, 277-321.

[31] Chen, X. and D. Pouzo (2014): "Sieve Wald and QLR Inferences on Semi/nonparametric Conditional Moment Models," forthcoming in *Econometrica*.

[32] Chen, X., D. Pouzo and E. Tamer (2011): "Inference on Partially Identified Semi/Nonparametric Conditional Moment Restrictions Models," working paper.

[33] Chen, X., E. Tamer and A. Torgovitsky (2012): "Sensitivity Analysis in Semiparametric Likelihood Models," working paper.

[34] Cheng, G. and Z. Shang (2014): "Joint Asymptotics for Semi-Nonparametric Models under Penalization," *arXiv*: 1311.2628.

[35] Chernozhukov, V., D. Chetverikov, and K. Kato (2013): "Gaussian Approximation of Suprema of Empirical Processes," *The Annals of Statistics*, forthcoming.

[36] Chernozhukov, V., and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica,* 73, 245-261.

[37] Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and Confidence Regions for Parametric Sets in Econometric Models," *Econometrica*, 75, 1243-1284.

[38] Chernozhukov, V., G. W. Imbens and W. K. Newey (2007): "Instrumental Variable Estimation of Nonseparable Models," *Journal of Econometrics*, 139, 4-14.

[39] Chernozhukov, V., S. Lee and A. M. Rosen (2013): "Intersection Bounds: Estimation and Inference," *Econometrica*, 667-737.

[40] Darolles, S., Y. Fan, J. Florens, and E. Renault (2011): "Nonparametric Instrumental Regression," *Econometrica*, 79, 1541-1566.

[41] Donald, S., G., G. W. Imbens and W. K. Newey (2003): "Empirical Likelihood Estimation and Consistent Tests with Conditional Moment Restrictions," *Journal of Econometrics*, 117, 55-93.

[42] Freyberger, J., and J. Horowitz (2012): "Identification and Shape Restrictions in Nonparametric Instrumental Variables Estimation," working paper.

[43] Gagliardini, P. and O. Scaillet (2012): "Nonparametric Instrumental Variable Estimation of Structural Quantile Effects," *Econometrica*, 80, 1533-1562.

[44] Galichon, A. and M. Henry (2011): "Set Identification in Models with Multiple Equilibria," *Review of Economic Studies*, 78, 1264-1298.

[45] Giné, E. and Koltchinskii,V. (2006): "Concentration Inequalities and Asymptotic Results for Ratio Type Empirical Processes," *The Annals of Probability*, 34, 1143-1216.

[46] Grundl, S. and Y. Zhu (2014): "Nonparametric Tests in Moment Equality Models with an Application to Infer Risk Aversion in First-Price Auctions," working paper.

[47] Hall, P. and J. Horowitz (2005): "Nonparametric Methods for Inference in the Presence of Instrumental Variables," *The Annals of Statistics*, 33, 2904-2929.

[48] Hall, P. and J. Horowitz (2005): "Nonparametric Methods for Inference in the Presence of Instrumental Variables," *The Annals of Statistics*, 33, 2904-2929.

[49] Hansen, B. E. (1996): "Inference When a Nuisance Parameter is Not Identified Under the Null Hypothesis," *Econometrica*, 64, 413-430.

[50] Hansen, B. E. (2014): "Nonparametric Sieve Regression: Least Squares, Averaging Least Squares, and Cross-Validation," in *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics*, 215-248, edited by J. Racine, L. Su and A. Ullah. Oxford: Oxford University Press.

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

[52] Hansen, L. P. and K. J. Singleton (1982): "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50, 1269-1286.

[53] Hong, S. (2013): "Inference in Semiparametric Conditional Moment Model with Partial Identification," working paper.

[54] Horowitz, J. (2011): "Applied Nonparametric Instrumental Variables Estimation," *Econometrica*, 79, 347-394.

[55] Horowitz, J. (2012): "Specification Testing in Nonparametric Instrumental Variable Estimation," *Journal of Econometrics*, 167, 383-396.

[56] Horowitz, J. and S. Lee (2012): "Uniform Confidence Bands for Functions Estimated Nonparametrically with Instrumental Variables," *Journal of Econometrics*, 168, 175-188.

[57] Huang, J. Z. (1998): "Projection Estimation in Multiple Regression With Applications to Functional ANOVA Models," *The Annuals of Statistics*, 26, 242-272.

[58] Huang, J. Z. (2003): "Local Asymptotics for Polynomial Spline Regression," *The Annals of Statistics*, 31, 1600-1635.

[59] Imbens, G. W., and C. F. Manski (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845-1857.

[60] Imbens, G. W., and W. K. Newey (2009): "Identification and Estimation of Triangular Simultaneous Equation Models Without Additivity," *Econometrica*, 77, 1481-1512.

[61] Kahn, S. and E. Tamer (2010): "Irregular Identification, Support Conditions, and Inverse Weight Estimation," *Econometrica*, 78, 2021-2042.

[62] Kitamura, Y., G., Tripathi, G. and H. Ahn (2004): "Empirical Likelihood Based Inference in Conditional Moment Restriction Models," *Econometrica*, 72, 1667-1714.

[63] Lewbel, A. (2008): "Engel Curve," *The New Palgrave Dictionary of Economics,* 2nd edition, edited by Steven N. Durlauf and Lawrence E. Blume. Palgrave Macmillan.

[64] Liu, C. and J. Tao (2014): "Model Selection and Model Averaging in Nonparametric Instrumental Variables Models," working paper.

[65] Liu, X. and Y. Shao (2003): "Asymptotics for Likelihood Ratio Tests under Loss of Identifiability," *The Annals of Statistics*, 807-832.

[66] Murphy, S. A. and van der Vaart, A. W. (2000): "On Profile Likelihood (with Discussion)," *Journal of the American Statistic Association*, 95, 449-485.

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

[68] Newey, W. K. (2013): "Nonparametric Instrumental Variables Estimation," *American Economic Review*: *Papers and Proceedings*, 103, 550-556.

[69] Newey, W. K. and D. L. McFadden (1994): "Large Sample Estimation and Hypothesis testing," in *Handbook of Econometrics* 4, 2111-2245, ed. by R. Engle and D. McFadden. Amsterdam: North Holland.

[70] Newey, W. K. and J. L. Powell (2003): "Instrumental Variables Estimation of Nonparametric Models," *Econometrica*, 71, 1557-1569.

[71] Newey, W. K., J. L. Powell and F. Vella (1999): "Nonparametric Estimation of Triangular Simultaneous Equation Models," *Econometrica*, 67, 565-603.

[72] Newey, W.K. and R. J. Smith (2004): "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72, 219-255.

[73] Newey, W. K. and K. D. West (1987): "Hypothesis Testing with Efficient Method of Moments Estimation," *International Economic Review*, 28, 777-787.

[74] Nickl, R. and B. M. Pötscher (2007): "Bracketing Metric Entropy Rates and Empirical Central Limit Theorems for Function Classes of Besov- and Sobolev-Type," *Journal of Theoretical Probability*, 20,177-199.

[75] Otsu, T. (2011): "Empirical Likelihood Estimation of Conditional Moment Restriction Models with Unknown Functions," *Econometric Theory*, 27, 8-46.

[76] Pollard, D. (2002): *A User's Guide to Measure Theoretic Probability.* Cambridge Series in Statistics and Probabilistic Mathematics.

[77] Robinson, P. (1988): "Root-n-Consistent Semiparametric Regression," *Econometrica*, 56, 931-954.

[78] Romano, J. P., and A. M. Shaikh (2008): "Inference for Identifiable Parameters in Partially Identified Econometric Models," *Journal of Statistical Planning and Inference*, 138, 2786-2807.

[79] Romano, J. P., and A. M. Shaikh (2010): "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica*, 78, 169-211.

[80] Santos, A. (2011): "Instrumental Variable Methods for Recovering Continuous Linear Functionals," *Journal of Econometrics*, 161, 129-146.

[81] Santos, A. (2012): "Inference in Nonparametric Instrumental Variables with Partial Identification," *Econometrica*, 80, 213-275.

[82] Shen, X. (1997): "On Methods of Sieves and Penalization, " *The Annals of Statistics*, 25, 580-615.

[83] Shen, X. and J. Shi (2005): "Sieve Likelihood Ratio Inference on General Parametric Space," Science in China Series A: Mathematics, 48(1), 67-78.

[84] Stoye, J. (2009): "More on Confidence Intervals for Partially Identified Parameters," *Econometrica*, 1299-1315.

[85] Sueishi, N. (2012): "Efficient Estimation via Conditional Moment Restrictions Containing Unknown Functions," working paper, Kyoto University.

[86] van der Vaart, A., and J. Wellner (1996): *Weak Convergence and Empirical Process: with Applications to Statistics*. New York: Springer-Verlag.

[87] Wilks, S. (1938): "The Large-sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses," *Annals of Mathematical Statistics*, 9(1): 60-62.in the main text.

# APPENDIX

## Notation and Definitions

The following notation and definitions will be used in the appendix, including some that go beyond the ones defined in the main text.

$$
\begin{aligned}
m(X,\theta) &\equiv E\left[\rho(Z,\theta)|X\right] \\
\rho(\theta) &\equiv \rho(Z,\theta) \\
\Sigma_0(X) &\equiv \Sigma(X,\theta_0) \equiv E\left[\rho(Z,\theta_0)\rho(Z,\theta_0)'|X\right] \\
\Psi(X,\theta) &\equiv \Sigma(X,\theta)^{-1/2} m(X,\theta) \\
\Omega_n &\equiv E\left[\frac{dm(X,\theta)}{d\theta}\left[\bar{p}^{k_n}(\cdot)'\right]' \Sigma_o(X)^{-1} \frac{dm(X,\theta)}{d\theta}\left[\bar{p}^{k_n}(\cdot)'\right]\right]^{-1} \\
\|\delta_n^*\| &\equiv A_n(w)'\Omega_n A_n(w) \\
u_n^* &= \delta_n^* / \|\delta_n^*\| \\
\frac{d\Psi(X,\theta)}{d\theta}\left[\delta_n^*\right] &\equiv \Sigma(X,\theta)^{-1/2} \frac{dm(X,\theta)}{d\theta}\left[\delta_n^*\right] \\
\alpha_n(w) &\equiv \Omega_n A_n(w) / ||\Omega_n A_n(w)|| \\
\mu_n^* &= \nu_n^* / ||\nu_n^*||_{sd} \\
h\left(\Theta_1,\Theta_2\right) &\equiv \sup_{\theta_1\in\Theta_1} \inf_{\theta_2\in\Theta_2} \|\theta_1-\theta_2\| \\
d_H\left(\Theta_1,\Theta_2,||\cdot||\right) &\equiv \max\left\{h(\Theta_1,\Theta_2), h\left(\Theta_2,\Theta_1\right)\right\} \\
Q(X,\theta) &\equiv \Sigma(X,\theta)^{1/2} \otimes q(X)', \\
Q(\theta) &\equiv \left(Q(X_1,\theta)',\ldots,Q(X_n,\theta)'\right)' \\
\Psi(X,\theta) &\equiv \Sigma(X,\theta)^{-1/2} m(X,\theta) \\
\hat{\Psi}(x,\theta) &\equiv Q(x,\theta)\left(Q(\theta)'Q(\theta)\right)^{-1} \sum_{j=1}^{n} Q(X_j,\theta)'\Sigma(X_j,\theta)^{-1/2}\rho(Z_j,\theta) \\
\tilde{\Psi}(x,\theta) &\equiv Q(x,\theta)\left(Q(\theta)'Q(\theta)\right)^{-1} \sum_{j=1}^{n} Q(X_j,\theta)'\Sigma(X_j,\theta)^{-1/2} m(X_j,\theta) \\
\hat{\Psi}_\theta(X,\theta) &\equiv Q\left(X_i,\theta\right)\left(Q(\theta)'Q(\theta)\right)^{-1} \sum_{j=1}^{n} Q(X_j,\theta)\Sigma(X_j,\theta)^{-1/2} \frac{d\rho(Z_j,\theta)}{d\theta}\left[p^{k_n}(\cdot)\right]. \\
\frac{d\hat{\Psi}(x,\theta)}{d\theta}\left[\delta_n^*\right] &\equiv Q(x,\theta)\left(Q(\theta)'Q(\theta)\right)^{-1} \sum_{j=1}^{n} Q(X_j,\theta)'\Sigma(X_j,\theta)^{-1/2} \frac{d\Psi(X_j,\theta)}{d\theta}\left[\delta_n^*\right].
\end{aligned}
$$

For a matrix $Q$, we use $\lambda_{\min}(Q)$ and $\lambda_{\max}(Q)$ to denote the minimal and maximal eigenvalues of $Q$, respectively. We use $\mathbb{E}_n[g] = \mathbb{E}_n\left[g(x_i)\right] = \frac{1}{n}\sum_{i=1}^n g(x_i)$ and $\mathbb{G}_n\left[g\right] = \mathbb{G}_n\left[g(x_i)\right] = \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(g(x_i) - E\left[g(X_i)\right]\right)$.

For a function class $\mathcal{F}$ equipped with an envelope function $\sup_{f\in\mathcal{F}}|f(x)| \leq F(x)$. We use $N\left(\varepsilon,\mathcal{F},L^2(Q)\right)$ to denote the covering number of $\mathcal{F}$-the minimal number of $L^2(Q)-$balls of radius $\varepsilon$ to cover the function set $\mathcal{F}$. We use $N\left(\varepsilon\left\|F\right\|_{Q,2},L^2\left(Q\right),\mathcal{F}\right)$ to denote the covering number relative to the envelop function $F(x)$.

The entropy is defined by the logarithm of the covering number. Let

$$J(\delta, \mathcal{F}) = \sup_Q \int_0^\delta \sqrt{1 + \log N\left(\varepsilon ||F||_{Q,2}, \mathcal{F}, L^2(Q)\right)} d\varepsilon$$

where the supremum is taken over probability measures $Q$ with $||F||_{Q,2} > 0$. Similarly, we use $N_{[]}(\varepsilon, \mathcal{F}, ||\cdot||)$ to denote the bracketing number of size $\varepsilon$ for $\mathcal{F}$ under the norm $||\cdot||$.

The Hölder space of order $\eta > 0$, denoted by $\Lambda^\eta(\mathcal{W})$ , is a space of functions $g : \mathcal{W} \to \mathbf{R}$ such that the first $\eta$th derivatives are bounded, and the $\underline{\eta}$th derivative satisfies

$$\max_{\sum_{i=1}^{d_x} a_i = \underline{\eta}} |\bigtriangledown^{\mathbf{a}} g(w) - \bigtriangledown^{\mathbf{a}} g(w')| \lesssim ||w - w'||_e^{\eta - \underline{\eta}}.$$

The Hölder norm is defined by

$$\|g\|_{\Lambda^\eta} = \sup_w |g(w)| + \max_{a_1 + a_2 + \ldots, a_{d_x} = \underline{\eta}} \sup_{x \neq x'} \frac{|\partial^{a_1 + a_2 + \ldots, + a_{d_w}} g(w)|}{\partial w_1^{a_1} \ldots \partial w_{d_w}^{a_{d_w}}} < \infty.$$

And the Hölder ball with radius c is defined by $\Lambda_c^\eta(\mathcal{W}) \equiv \{g \in \Lambda_c^\eta(\mathcal{W}) : ||g||_{\Lambda^\eta} \leq c < \infty\}$.

# A ASSUMPTIONS

For a large constant $K$, let $\Theta_n^K = \{\theta \in \Theta_n : \lambda_n \text{Pen}(h) \leq \lambda_n K\}$ such that both $\Pi_n \theta_0$ and $\hat{\theta}_n$ belongs to $\Theta_n^K$ w.p.a.1. Let $\Theta_{os} \subset \{\theta \in \Theta : \lambda_n \text{Pen}(\theta) < \lambda_n K, \|\theta - \theta_0\|_s < K\}$. Let $\Theta_{osn}$ be the sieve space of $\Theta_{os}$. Let $\bar{\varrho}_n^2 = \frac{s_n}{n} + b_{m,s_n}^2$, $\varrho_n^2 = \max\left\{s_n/n, b_{m,s_n}^2\right\} = o_p(n^{-1/2})$. Let $K_n = \log\log n$.

**Assumption A.1.** *(i) For each $s_n$ there is a constant $\xi_{s_n}$ and matrix $B_1$ such that $\tilde{q}^{s_n}(x) = B_1 q^{s_n}(x)$ for all $x \in \mathcal{X}$, $\sup_{x \in \mathcal{X}} \|\tilde{q}^{s_n}(x)\| \leq \xi_{s_n}$, $E\left[\tilde{q}^{s_n}(x)\tilde{q}^{s_n}(x)'\right]$ has smallest eigenvalue bounded away from zero and $\sqrt{s_n} \lesssim \xi_s$; (ii) for each $k_n$ there is a constant $\xi_{k_n}$ and matrix $B_2$ such that $\tilde{p}^{k_n}(y) = B_2 p^{k_n}(y)$ for all $y \in \mathcal{Y}$, $\sup_{y \in \mathcal{Y}} \left\|\tilde{p}^{k_n}(y)\right\| \leq \xi_{k_n}$, $E\left[\tilde{p}^{k_n}(y)\tilde{p}^{k_n}(y)'\right]$ has smallest eigenvalue bounded away from zero; (iii) $s_n \log(s_n)/n = o(1)$ for $q^{s_n}(x)$ a polynomial spline.*

Assumption A.1 is a normalization that is standard in the literature (see, e.g., Newey (1997)). Explicit formula for $\xi_{k_n}$ and $\xi_{s_n}$ are specific for different basis functions. For example, the polynomial series satisfies $\xi_{k_n} \lesssim k_n$ and the Fourier series satisfies $\xi_{k_n} \lesssim \sqrt{k_n}$. For convenience, in the main context, we state $q^{s_n}(x) = \tilde{q}^{s_n}(x)$, $\tilde{p}^{k_n}(y_2) = p^{k_n}(y_2)$, $E\left[\tilde{p}^{k_n}(y_2)\tilde{p}^{k_n}(y_2)\right] = I_{k_n}$ and $E\left[\tilde{q}^{s_n}(x)\tilde{q}^{s_n}(x)'\right] = I_{s_n}$.

**Assumption A.2.** *$\hat{L}_n(\theta_{0n}) \lesssim L(\theta_{0n}) + o_p(n^{-1})$; (ii) there exists an open $||\cdot||_s$−neighborhood of $\theta_0$, $\Theta_{os}$, such that (a) $||\theta - \theta_0||^2 \lesssim L(\theta) \lesssim ||\theta - \theta_0||^2$ holds for al $\theta \in \Theta_{os}$; (b) $\Theta_{os}$ is convex; (c) $m(\cdot, \theta)$ is continuously pathwise differentiable with respect to $\theta \in \Theta_{os}$.*

**Assumption A.3.** *$\Theta$ is convex at $\theta_0$ in the sense that for any $\theta \in \Theta$, $(1-t)\theta_0 + t\theta \in \Theta$ for small $t > 0$. We also assume that for almost all $Z$, $\rho(Z, (1-t)\theta_0 + t\theta)$ is continuously differentiable at $t = 0$.*

**Assumption A.4.** *(Penalty) We have either $\lambda_n = o(n^{-1})$ or $\lambda_n \sup_{h^1, h^2 \in \mathcal{N}_{os}} \left|Pen\left(h^1\right) - Pen(h^2)\right| = o(n^{-1})$.*

Assumption A.4 allow us to ignore the effect of the penalty term in first-order limiting theory. To derive the asymptotic distribution of functionals of $\theta_0$, we give some more conditions on $\phi(\cdot)$.

For a class of functions $\mathcal{F}$, let $N_{[]}(\varepsilon, \mathcal{F}, ||\cdot||_{L^2})$ be the $L^2-$covering number with bracketing of $\mathcal{F}$. For $s = 1, \ldots, s_n$, let

$$\mathcal{F}_{s,1} = \left\{\rho(\cdot,\theta)q_s(\cdot) : \theta \in \Theta_n^K\right\}$$

and

$$\mathcal{F}_{s,2} = \left\{\rho(\cdot,\theta)q_s(\cdot) : \theta \in \Theta_{osn}\right\}.$$

**Assumption A.5.** *(i) Uniformly over $\theta \in \Theta_n^K$, there are $s_n-$vectors $\pi_n$ such that*

$$E\left[\left\{m(X,\theta) - q^{s_n}(X)'\pi_n\right\}^2\right] = O(b_{m,s_n}^2) = o(1)$$

*as $s_n \to \infty$. (ii) There exists a sequence of measurable functions $\{\bar{\rho}_n(Z)\}_{n=1}^{\infty}$ such that $E\left[\bar{\rho}_n(Z)^2|X\right] < \infty$ and $\sup_{\theta\in\Theta_n^{K_0}} |\rho(Z,\theta)| \leq \bar{\rho}_n(Z)$. (iii) With $C_n$ such that $\frac{s_n}{n}C_n = o(1)$, we have*

$$\max_{1\leq s\leq s_n} \int_0^1 \sqrt{1 + \log N_{[]}(\varepsilon, \mathcal{F}_{s,1}, ||\cdot||_{L^2})}d\varepsilon \leq \sqrt{C_n} < \infty$$

*and*

$$\max_{1\leq s\leq s_n} \int_0^1 \sqrt{1 + \log N_{[]}(\varepsilon, \mathcal{F}_{s,2}, ||\cdot||_{L^2})}d\varepsilon < \infty.$$

*(iv) For $\omega \in (0,1]$, and $K : \mathcal{X} \to \mathbf{R}$ with $E\left[|K(X)|^2\right] < \infty$, $\forall\epsilon_n > 0$, $\forall\theta_n' \in \mathcal{N}_{osn} \cup \{\theta_0\}$ and all $n$, for $r > 1$,*

$$E\left[\sup_{\theta_n\in\mathcal{N}_{osn}:||\theta_n-\theta_n'||_s\leq\epsilon_n} \|\rho(Z,\theta_n) - \rho(Z,\theta_n')\|_E^2 |X = x\right] \leq K(x)^2\epsilon_n^{2\omega},$$

$$E\left[\sup_{\theta_n\in\mathcal{N}_{osn}:||\theta_n-\theta_n'||_s\leq\epsilon_n} \|\rho(Z,\theta_n)\rho(Z,\theta_n)' - \rho(Z,\theta_n')\rho(Z,\theta_n')'\|_E^r |X = x\right] \leq K(x)^r\epsilon_n^{r_\omega}.$$

Assumption A.5 is similar to Assumption C.2 in Chen and Pouzo (2012), which is used to obtain the consistency and convergence rates of PSGMM estimators.

**Assumption A.6.** *(Lindeberg condition) Let $M_{n,i} = \frac{dm(X_i,\theta_0)}{d\theta}[u_n^*]'\Sigma_0(X_i)^{-1}\rho(Z_i,\theta_0)$. For $\varepsilon > 0$,*

$$\lim_{n\to\infty} \sup E\left[(M_{n,i})^2 1\left\{|\varepsilon| n^{-1/2}M_{n,i}\right\} > 1\right] = 0.$$

**Assumption A.7.** *(i) There is a $\delta_n^* \in \Theta_n\backslash\{\theta_0\}$ such that $\varrho_n \times \|\delta_n^* - \delta^*\| = o(n^{-1/2})$. (ii)*

$$\sup_{\theta_n\in\mathcal{N}_{0sn}} \left\|\frac{1}{n}\sum_{i=1}^n \frac{d\rho(Z_i,\theta_n)}{d\theta}[u_n^*(w)] \otimes q^{s_n}(X_i)\right\|_E = O_p(1)$$

*and*

$$\sup_{\theta_n,\in\mathcal{N}_{0sn}} \left\|\frac{1}{n}\sum_{i=1}^n \frac{d^2\rho(Z_i,\theta_n)}{d\theta}[u_n^*(w)] \otimes q^{s_n}(X_i)\right\|_E = O_p(1).$$

**Assumption A.8.** *(Assumption A.5 (ii) and (iii) in Chen and Pouzo (2014)). (i) $\varrho_n \simeq \sqrt{s_n/n} = \max\left\{\sqrt{s_n/n}, b_{m,s_n}\right\} = o(n^{-1/4})$; (ii) $\forall\epsilon_n > 0$, $\max\left\{(K_n\varrho_n)^2,\ (K_n\varrho_{sn})^{2\omega}\right\} = (K_n\varrho_{sn})^{2\omega}$;*

*(iii) let* $\mathcal{F}_3 \equiv \{\rho(\cdot,\theta) - \rho(\cdot,\theta_0) : \theta \in \mathcal{N}_{osn}\}$, *then*

$$1 \leq \sqrt{C_n} \equiv \int_0^1 \sqrt{1 + \log\left(N_{[]}(\mathcal{F}_3, \varepsilon(K_n\delta_{s,n})^{\omega}, ||\cdot||_{L^2}\right)} d\varepsilon < \infty$$

*and* $\max\left\{(K_n\varrho_{s,n})^{\omega}\sqrt{C_n}, K_n\right\} n\varrho_n^2 (K_n\varrho_{s,n})^{\omega}\sqrt{C_n} \to 0$.

**Assumption A.9.** *Uniformly over* $\bar{\theta}_n \in \mathcal{N}_{osn} \cup \{\theta_0\}$,

*(i)* $E\left[\left\|\frac{d\hat{\Psi}(X_i,\bar{\theta}_n)}{d\theta}[\delta_n^*(w)] - \frac{d\hat{\Psi}(X_i,\bar{\theta}_n)}{d\theta}[\delta_n^*(w)]\right\|_E^2\right] = O_p\left((K_n\varrho_n)^{-2} n^{-1}\right)$;

*(ii)*

$$E\left[\left\|\tilde{\Psi}(X_i,\bar{\theta}) - \Psi(X_i,\bar{\theta})\right\|_E^2\right] = O_p\left((K_n\varrho_n)^{-2} n^{-1}\right).$$

*(iii) Let* $\{a_{1n}\}_{n=1}^{\infty}$ *and* $\{a_{2n}\}_{n=1}^{\infty}$ *be real valued positive sequences such that* $a_{1n} = o(1)$ *and* $a_{2n} = o(1)$. *Suppose there is a continuous mapping* $F : \mathbf{R}_+ \to \mathbf{R}_+$ *such that*

$$\max\left\{F(a_{1n}), n^{-1/4}\right\}\varrho_n \log\log n = o(n^{-1/2})$$

*and*

$$\sup_{\mathcal{N}_{osn}} \sup_{\|u_n^*(w)-u_n(w)\|\leq a_{1n}} \frac{1}{n}\sum \left\|\frac{d\Psi(X_i,\bar{\theta}_n)}{d\theta}[u_n^*(w)] - \frac{d\Psi(X_i,\bar{\theta}_n)}{d\theta}[u_n(w)]\right\|_E^2 = O_p\left(\max\{F(a_{1n})^2, n^{-1/2}\}\right);$$

$$\sup_{\mathcal{N}_{osn}} \sup_{\delta_n\in\bar{\Delta}_n:||\delta_n||=1} \frac{1}{n}\sum_{i=1}^{n} \left\|\frac{d\hat{\Psi}(X_i,\bar{\theta}_n)}{d\theta}[u_n(w)] - \frac{d\Psi(X_i,\bar{\theta}_n)}{d\theta}[u_n(w)]\right\|_E^2 = O_p\left(\max\{a_{2n}^2, n^{-1/2}\}\right).$$

**Assumption A.10.** *(i)* $d_\rho s_n \geq k_n$ *and* $k_n/n = o(1)$. *(ii)* $k_n \ln(n)\xi_{k_n}^2 n^{-1/2} = o(1)$. *(iii)* $n^{-1/4+1/m}\xi_{s_n} = o(1)$; *(iv)* $n^{-1/2}s_n^{1/2}\xi_{s_n} = o(n^{-1/4})$.

**Assumption A.11.** *(i)* $\sup_{\theta\in\mathcal{N}_{osn}} \sup_{(\delta_n^*,w)\in\bar{\Delta}_n\times\mathcal{W}} \left\|\frac{d\phi(\theta)}{d\theta}[\delta_n^*(w)] - \frac{d\phi(\theta_0)}{d\theta}[\delta_n^*(w)]\right\|_E^2 (K_n\varrho_n)^2 = o(n^{-1})$;

*(ii)*

$$\sup_{\theta\in\mathcal{N}_{osn}} \left\|\frac{d\phi(\theta)}{d\theta}\left[\bar{p}^{k_n}(\cdot)\right] - \frac{d\phi(\theta_0)}{d\theta}\left[\bar{p}^{k_n}(\cdot)\right]\right\|_E \lesssim \xi_{\phi,n}K_n\varrho_n.$$

**Assumption A.12.** $\sup_{\delta_1(w),\delta_2(w)\in\bar{\Delta}_n} |\langle\delta_1(w),\delta_2(w)\rangle_n - \langle\delta_1(w),\delta_2(w)\rangle| = o_p(K_n\varrho_n)$.

**Assumption A.13.** *For each* $k_n$ *and any* $\theta \in \mathcal{N}_{osn}$, $\delta \in \bar{\Delta}_{k_n} = \bar{\Delta}_n \to \frac{d\hat{\Psi}(\cdot,\theta)}{d\theta}[\delta] \in L^2(f_X)$ *is a linear functional.*

**Assumption A.14.** *(i)* $m(X,\theta)$ *is twice pathwise differentiable in* $\theta \in \mathcal{N}_{osn}$ *and uniformly over the direction* $\delta_n^*(w)$ *for* $w \in \mathcal{W}$. *Furthermore,*

$$(K_n\varrho_n)^2 \times E\left[\sup_{(\theta,w)\in\mathcal{N}_{osn}\times\mathcal{W}} \left|\frac{d^2m(X,\theta)}{d\theta d\theta}[u_n^*(w), u_n^*(w)]\right|^2\right] = o(1);$$

*(ii)*

$$E\left[\sup_{(\theta,w)\in\mathcal{N}_{osn}\times\mathcal{W}}\left(\frac{dm(X,\theta)}{d\theta}\left[u_n^*(w)\right]-\frac{m(X,\theta_0)}{d\theta}\left[u_n^*(w)\right]\right)'\right.$$
$$\left.\times\Sigma_0(X)^{-1}\left(\frac{dm(X,\theta)}{d\theta}\left[u_n^*(w)\right]-\frac{m(X,\theta_0)}{d\theta}\left[u_n^*(w)\right]\right)\right]=o(n^{-1/2});$$

*(iii) for all* $\theta_n\in\mathcal{N}_{osn}$, $\bar{\theta}\in\mathcal{N}_{0s}$ *and* $w\in\mathcal{W}$,

$$E\left[\left(\frac{dm(X,\theta_0)}{d\theta}\left[u_n^*(w)\right]\right)'\Sigma_0(X)^{-1}\left(\frac{dm(X,\bar{\theta})}{d\theta}\left[\theta_n-\theta_0\right]-\frac{dm(X,\bar{\theta})}{d\theta}\left[\theta_n-\theta_0\right]\right)\right]=o(n^{-1/2}).$$

Assumption A.1-A.14 is similar to the standard ones in the literature (see, e.g., Chen and Pouzo (2009) and Chen and Pouzo (2014) ). In some assumptions, we require the conditions to hold uniformly over $w\in\mathcal{W}$. However, they are stronger than the ones required for pointwise inference, where we only need them to hold for a fixed $w\in\mathcal{W}$.

**Assumption A.15.** *Let* $\xi_{k,\phi}=\sup_{w\in\mathcal{W}}||A_n(w)||$ *and* $\xi_{k_n}^L\equiv\sup_{w,w'\in\mathcal{W}:w\neq w'}\frac{||A_n(w)-A_n(w')||}{||w-w'||}$. *Loadings on the coefficient satisfies (i)* $\sup_{w\in\mathcal{W}}1/||A_n(w)||\lesssim 1$; *(ii)* $\log\xi_{k_n}^L\lesssim\log k_n$.

**Assumption A.16.** *(i) There is a non-zero linear functional mapping from* $\bar{\Delta}$ *to* $\boldsymbol{R}$ *such that for all* $w\in\mathcal{W}$: $\delta\to\left(\frac{d\phi(\theta_0)}{d\theta}\left[\delta\right]\right)[w]$; *(ii) for* $K_n=\log\log n$ *and* $\varrho_n$ *defined in (3.5), let* $\mathcal{T}_n\equiv\left\{t\in\mathbf{R}:|t|\lesssim K_n^2\varrho_n\right\}$, *then*

$$\sup_{(\theta,t,w)\in\mathcal{N}_{osn}\times\mathcal{T}_n\times\mathcal{W}}\left|\phi(\theta)[w]-\phi(\theta_0)[w]-\left(\frac{\partial\phi(\theta_0)}{\partial\theta}\left[\theta-\theta_0\right]\right)[w]\right|/\left\|\delta_n^*(w)\right\|=o(n^{-1/2});$$

*(iii) either (a) or (b) holds: (a)* $\|\delta_n^*(w)\|\to\infty$ *and* $\left|\left(\frac{\partial\phi(\theta_0)}{\partial\theta}\left[\theta_{0n}-\theta_0\right]\right)[w]\right|/\left\|\delta_n^*(w)\right\|=o\left(n^{-1/2}\right)$; *(b)* $\|\delta_n^*(w)\|\to\|\delta^*(w)\|<\infty$ *and* $\|\delta^*(w)-\delta_n^*(w)\|\times\|\theta_{0n}-\theta_0\|=o(n^{-1/2})$.

**Assumption A.17.** *(i)* $\max_{1\leq j\leq d_\rho}\sup_{\theta\in\mathcal{N}_{0sn}}\left|\frac{d\rho_j(X,\theta)}{d\theta}\left[\bar{p}^{k_n}(\cdot)'\right]\right|\leq\xi_{\rho,k_n}<\infty$; *(ii) for all* $\theta\in\mathcal{N}_{osn}$, $E\left[\left\|\frac{dm(X,\theta)}{d\theta}\left[\bar{p}^{k_n}(\cdot)'\right]\right\|^2\right]\leq c(X)k_n$ *and* $E\left[c(X)^2\right]\leq$ *const.* $<\infty$. *(iii) the right and side of* $\frac{dm(X,\theta)}{d\theta}\left[\delta_n^*(w)\right]=E\left[\left.\frac{d\rho(Z,\theta+t\delta_n^*(w))}{dt}\right|_{t=0}|X\right]$ *is uniformly continuous in* $w$ *for all* $\theta\in\mathcal{N}_{osn}$.

**Assumption A.18.** *For some* $p\geq 8$, $\rho(Z,\theta)$ *satisfies the pth order envelop condition in* $\theta\in\Theta_n$.

**Assumption A.19.** *Let* $m>2$. *Suppose that* $E\left[|\rho(Z,\theta_0)|X|^m\right]\lesssim 1$; $\left(\xi_{k_n}^L\right)^{2m/(m-2)}\log n/n\lesssim 1$, $\xi_{s_n}\left(\xi_{k_n}^L+\sqrt{s_n}\right)\to 0$, $O_p\left(\rho_{s,n}\right)=o_p(b_n^{-1})$, $\xi_{\phi,n}K_n\varrho_{s,n}=o_p(b_n^{-1})$, $n^{1/m}\varrho_{s,n}+(v_n\vee 1)\sqrt{\frac{\xi_{s_n}^2\log n}{n}}=o_p\left(b_n^{-1}\right)$, $O_p\left(\max\{a_{1n},n^{-1/4}\}K_n\varrho_n\right)=o_p(b_n^{-1})$ *and* $\xi_{\rho,k_n}n^{1/m}=o_p(b_n^{-1})$.

The following conditions are used for the proof of Theorems and Lemmas in Section 4. They are modified from the conditions we used in Section 3 to fit the partial identification setting.

**Assumption A.20.** *The penalty function* $Pen:\mathcal{H}\to[0,\infty)$ *satisfies the following conditions: (i) Suppose that* $Pen(\cdot)$ *is a measurable function with* $\sup_{h\in\mathcal{H}_0}Pen(h)<\infty$; *(ii) the set* $\{h\in\mathcal{H}:Pen(h)\leq M\}$ *is compact under* $||\cdot||_s$ *for all* $M\in[0,\infty)$; *(iii)* $\lambda_n>0$ *such that* $\lambda_n\sup_{h\in\mathcal{H}_0}|Pen(h_n)-Pen(h)|=O(\lambda_n)=o(1)$;

*(iv)* $\lambda_n = O_p(\varrho_{pn}) = o_p(1)$, $\sup_{(\beta,h)\in\Theta_0} \bar{L}(\theta, \Pi_n h) = O_p(\lambda_n) = o_p(1)$, $E\left[\sup_{\theta\in\Theta} ||\rho(Z,\theta)||_E^4 | X\right] < \infty$; *(v)* $\sup_\Theta ||\theta - \Pi_n\theta||_s = O(c_{1n})$, $\max\{c_{1n}, \varrho_{pn}, \lambda_n\} = o_p(n^{-1/4})$.

**Assumption A.21.** *(Uniform Approximation Error) Uniformly over $\theta_0 \in \Theta_0^r$ and $\theta_{0n} \in \mathcal{B}_{0n}(\theta_0)$ and $w \in \mathcal{W}$,* $E\left[\left\|E\left[u_n^*(\theta_{0n}, w)|X_i\right]' \Sigma(X_i)\rho(Z_i, \theta_{0n}) - E\left[u_n^*(\theta_0, w)|X_i\right]' \Sigma(X_i)\rho(Z_i, \theta_0)\right\|_e^2\right] = o(n^{-1/2})$.

**Assumption A.22.** *For $\omega \in (0,1]$, and $K : \mathcal{X} \to \mathbf{R}$ with $E\left[|K(X)|^2\right] < \infty$, $\forall \epsilon_n > 0$, $\forall \theta \in \mathcal{B}_n(\theta_0) \cup \Theta_0$ and all $n$, for $r > 1$,*

$$E\left[\sup_{\theta^1,\theta^2\in\mathcal{B}_n(\theta_0)\cup\Theta_0 : ||\theta^1-\theta^2||\le\varrho_n} \left\|\rho(Z,\theta^1) - \rho(Z,\theta^2)\right\|_E^2 | X = x\right] \le K(x)^2 \varrho_n^{2\omega},$$

$$E\left[\sup_{\theta^1,\theta^2\in\mathcal{B}_n(\theta_0)\cup\Theta_0 : ||\theta^1-\theta^2||\le\epsilon_n} \left\|\rho(Z,\theta^1)\rho(Z,\theta^1)' - \rho(Z,\theta^2)\rho(Z,\theta^2)'\right\|_E^r | X = x\right] \le K(x)^r \epsilon_n^{r\omega}.$$

**Assumption A.23.** *For $\sigma_n$ defined in (4.3), uniformly over $\theta \in \mathcal{B}_n(\theta_0)$ and $w \in \mathcal{W}$, let $\mathcal{T}_n \equiv \{t \in \mathbf{R} : |t| \lesssim \sigma_n \log\log n\}$, then uniformly over $\theta \in \mathcal{B}_n(\theta_0)$,$t \in \mathcal{T}_n$ and $w \in \mathcal{W}$,*

$$\frac{\left|\mathbf{F}(\rho(\cdot\theta) + t\mu_n^*(\cdot,\theta_0,w))[w] - \mathbf{F}(\rho(\cdot,\theta_0)) - \left(\frac{d\mathbf{F}(\rho_0)}{d\rho}\left[\rho + t\mu_n^*(\cdot,\theta_0,w) - \rho_0\right]\right)[w]\right|}{\left\|\nu_n^*(\cdot,\theta_0,w)\right\|_{wp}}$$
$$= o(n^{-1/2}).$$

**Assumption A.24.** *(i)*

$$\sup_{\theta\in\mathcal{B}_n(\theta_0)\cup\Theta_0} \sup_{(\nu_n^*,w)\in\bar{\Delta}_n\times\mathcal{W}} \left\|\frac{\partial\phi(\theta)}{\partial\theta}\left[\nu_n^*(w)\right] - \frac{\partial\phi(\theta_0)}{\partial\theta}\left[\nu_n^*(w)\right]\right\|_E^2 (K_n\varrho_n)^2 = o(n^{-1});$$

*(ii)*

$$\sup_{\theta\in\mathcal{B}_n(\theta_0)\cup\Theta_0} \left\|\frac{d\phi(\theta)}{d\theta}\left[\bar{p}^{k_n}(\cdot)\right] - \frac{d\phi(\theta_0)}{d\theta}\left[\bar{p}^{k_n}(\cdot)\right]\right\|_E \lesssim \xi_{\phi,n} K_n \varrho_n.$$

**Assumption A.25.** *For fixed $w \in \mathcal{W}$, the empirical process $v_n(\theta_0)$ with $l-$component*

$$v_{n,l}(\theta_0) = \mathbb{G}_n\left\{E\left[\nu_{n,l}^*(Z_i,\theta_0)|X_i\right]' \Sigma(X_i)^{-1}\rho_l(Z_i,\theta_0)\right\},\ l = 1,\ldots,d_\rho,$$

*is asymptotically equicontinuous uniformly over $\theta_0 \in \Theta_0^r$ so that for any $\varepsilon > 0$,*

$$\lim_{\delta\to 0}\lim\sup_{n\to\infty} \Pr\left(\sup_{\theta_0\in\Theta_0^r : ||\theta_0^1-\theta_0^2||_s\le\epsilon_n} \left\|v_n(\theta_0^1) - v_n(\theta_0^2)\right\| > \varepsilon\right) = 0.$$

**Assumption A.26.** *For $S_n(\cdot)$ defined in (4.10) and $t = (\theta_0, w)$, $\left\|\mu_n^*(Z_i,\theta_0,w)\right\|_E \lesssim \Upsilon_{k_n}$,*

$$\left\|\left(S_n(t^1) - S_n(t^2)\right)\right\|_{P,2} \lesssim \left\|t^1 - t^2\right\|^c$$

*for some $0 < c \le 1/2$ in $L^2-$norm.*

# B MATHEMATICAL PROOFS

## B.1 Proofs of Sections 3

### B.1.1 Proofs of Section 3.0-3.1

Proof of Lemma 3.1 is presented in the Supplemental Appendix.

**Proof of Theorem 3.1.**

For any $\theta \in \mathcal{N}_{osn}$, let the local perturbation be $\theta(\epsilon_n) = \theta \pm \epsilon_n u_n^*$ form some $\epsilon_n = o(n^{-1/2})$, we have $\theta(\epsilon_n) \in \mathcal{N}_{osn}$. Since $\hat{\theta}_n \in \mathcal{N}_{osn}$ w.p.a.1, it implies that $\hat{\theta}_n(\epsilon_n) = \hat{\theta}_n \pm \epsilon_n u_n^* \in \mathcal{N}_{0sn}$ w.p.a.1. Furthermore, the definition of $\hat{\theta}_n$ implies that

$$-O_p\left(\epsilon_n^2\right) \lesssim L_n\left(\hat{\theta}_n \pm \epsilon_n u_n^*\right) - L_n\left(\hat{\theta}_n\right) + \lambda\left(\text{Pen}\left(\hat{\theta}_n \pm \epsilon_n u_n^*\right) - \text{Pen}\left(\hat{\theta}_n\right)\right).$$

For the penalty term, by Assumption 3.4 and A.4, a second order Taylor expansion yields

$$\lambda_n\left(\frac{dP(\hat{\theta}_n)}{d\theta}\left[\epsilon_n u_n^*\right] + \frac{1}{2}\frac{d^2P\left(\theta(s)\right)}{d\theta d\theta}\left[\epsilon_n u_n^*, \epsilon_n u_n^*\right]\right) = O_p\left(\lambda_n \epsilon_n\right) = o_p(n^{-1})$$

uniformly over $\theta(\epsilon_n) = \hat{\theta}_n + \epsilon_n u_n^* \in \mathcal{N}_{osn}$.

For some $s \in [0,1]$, a Taylor expansion and results in Lemma B.1 imply that

$$\begin{aligned}
&\left.\frac{d\hat{L}_n\left(\theta(\epsilon_n)\right)}{d\epsilon_n}\right|_{\epsilon_n=0} \\
=\ & 2\left(\frac{1}{n}\sum_{i=1}^n \frac{d\rho(Z_i,\hat{\theta}_n)}{d\theta}\left[\epsilon_n u_n^*\right]' \otimes q_i^{s_n}(X_i)\right)'\left(\frac{1}{n}\sum_{i=1}^n g_i(\bar{\theta}_n) g_i\left(\bar{\theta}_n\right)'\right)^{-1}\hat{g}\left(\hat{\theta}_n\right) + o_p(n^{-1}) \\
=\ & 2\left(\frac{1}{n}\sum_{i=1}^n \frac{d\rho(Z_i,\hat{\theta}_n)}{d\theta}\left[\epsilon_n u_n^*\right]' \otimes q_i^{s_n}(X_i)\right)'\left(\frac{1}{n}\sum_{i=1}^n \Sigma(X_i,\bar{\theta}_n)\otimes q_i q_i'\right)^{-1}\hat{g}(\hat{\theta}_n) + o_p(n^{-1}).
\end{aligned}$$

Then we can write

$$\begin{aligned}
&\left.\frac{d\hat{L}_n\left(\theta(\epsilon_n)\right)}{d\epsilon_n}\right|_{\epsilon_n=0} \\
=\ & \frac{2}{n}\left\{Q(X_i,\bar{\theta}_n)\left(Q\left(\bar{\theta}_n\right)' Q(\bar{\theta}_n)\right)^{-1}\sum_{j=1}^n Q(X_j,\bar{\theta}_n)\Sigma(X_j,\bar{\theta}_n)^{-1/2}\frac{d\rho(Z_j,\hat{\theta}_n)}{d\theta}\left[\epsilon_n u_n^*\right]\right\}' \\
& \times\left\{Q(X_i,\bar{\theta}_n)\left(Q\left(\bar{\theta}_n\right)' Q(\bar{\theta}_n)\right)^{-1}\sum_{j=1}^n Q(X_j,\bar{\theta}_n)\Sigma(X_j,\bar{\theta}_n)^{-1/2}\rho(Z_j,\hat{\theta}_n)\right\} \\
=\ & \frac{2\epsilon_n}{n}\sum_{i=1}^n\left\{\frac{dm(X_i,\theta_0)}{d\theta}\left[u_n^*\right]\right\}\Sigma(X_i,\theta_0)^{-1}\rho(Z_i,\theta_0) + \epsilon_n\left\langle u_n^*, \hat{\theta}_n - \theta_0\right\rangle + o_p\left(\epsilon_n n^{-1/2}\right).
\end{aligned}$$

For the second-order term,

$$\begin{aligned}
& \left.\frac{d^2\hat{L}_n(\theta(\epsilon_n))}{d\epsilon_n^2}\right|_{\epsilon_n=s} \\
= & \left(\frac{1}{n}\sum_{i=1}^n \frac{d\rho(Z_i,\theta(s))}{d\theta}\left[\epsilon_n u_n^*\right]\otimes q_i\right)' \left(\frac{1}{n}\sum_{i=1}^n g_i\left(\bar{\theta}_n\right) g_i\left(\bar{\theta}_n\right)'\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^n \frac{d\rho(Z_i,\theta(s))}{d\theta}\left[\epsilon_n u_n^*\right]\otimes q_i\right) \\
& +\hat{g}\left(\theta(s)\right)' \left(\frac{1}{n}\sum_{i=1}^n g_i\left(\bar{\theta}_n\right) g_i\left(\bar{\theta}_n\right)\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^n \frac{d\rho^2(Z_i,\theta(s))}{d\theta d\theta}\left[\epsilon_n u_n^*\right]\otimes q_i\right) \\
= & O_p\left(\epsilon_n^2\right) = o_p(n^{-1}).
\end{aligned}$$

It follows that

$$\sqrt{n}\left\langle u_n^*, \hat{\theta}_n - \theta_0\right\rangle = -\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{dm(X_i,\theta_0)}{d\theta}\left[u_n^*\right]'\Sigma(X_i,\theta_0)^{-1}\rho(Z_i,\theta_0) + o_p\left(1\right). \tag{B.1}$$

By (B.5) in Lemma B.2,

$$\delta_n^* = \eta_n^{*\prime}\bar{p}^{k_n}(\cdot)$$

and

$$\eta_n^* = \Omega_n \frac{d\phi(\theta_0)}{d\theta}\left[\bar{p}^{k_n}(\cdot)\right].$$

It implies that

$$\begin{aligned}
\left\langle u_n^*, \hat{\theta}_n - \theta_0\right\rangle &= \left\langle \delta_n^*/||\delta_n^*||, \hat{\theta}_n - \theta_0\right\rangle \\
&= -\frac{1}{n||\delta_n^*||}\sum_{i=1}^n \frac{dm(X_i,\theta_0)}{d\theta}\left[\eta_n^{*\prime}\bar{p}^{k_n}(\cdot)'\right]'\Sigma_0(X_i)^{-1}\rho(Z_i,\theta_0) + o_p(n^{-1/2}) \\
&= -\frac{1}{n||\delta_n^*||}\frac{d\phi(\theta_0)}{d\theta}\left[\bar{p}^{k_n}(\cdot)\right]' B_n^{-1}\sum_{i=1}^n \frac{dm(X_i,\theta_0)}{d\theta}\left[\bar{p}^{k_n}(\cdot)\right]'\Sigma_0(X_i)^{-1}\rho(Z_i,\theta_0) + o_p(n^{-1/2})
\end{aligned}$$

with

$$\begin{aligned}
||\delta_n^*|| &= E\left[\left(\frac{dm(X,\theta_0)}{d\theta}\left[\delta_n^*\right]\right)'\Sigma_0(X)\left(\frac{dm(X,\theta_0)}{d\theta}\left[\delta_n^*\right]\right)\right] \\
&= \frac{d\phi(\theta_0)}{d\theta}\left[\bar{p}^{k_n}(\cdot)\right]'\Omega_n E\left[\frac{dm(X,\theta_0)}{d\theta}\left[\bar{p}^{k_n}(\cdot)'\right]'\Sigma_0(X)^{-1}\frac{dm(X,\theta_0)}{d\theta}\left[\bar{p}^{k_n}(\cdot)'\right]\right] \\
&\quad \times\Omega_n\frac{d\phi(\theta_0)}{d\theta}\left[\bar{p}^{k_n}(\cdot)\right] \\
&= \frac{d\phi(\theta_0)}{d\theta}\left[\bar{p}^{k_n}(\cdot)\right]'\left\{E\left[\frac{dm(X,\theta_0)}{d\theta}\left[\bar{p}^{k_n}(\cdot)'\right]'\Sigma_0(X)^{-1}\frac{dm(X,\theta_0)}{d\theta}\left[\bar{p}^{k_n}(\cdot)'\right]\right]\right\}^{-1}\frac{d\phi(\theta_0)}{d\theta}\left[\bar{p}^{k_n}(\cdot)\right].
\end{aligned}$$

The conclusion follows. □

**Lemma B.1.** *Suppose that Assumptions of Theorem 3.1 are satisfied. Then uniformly over $\theta \in \mathcal{N}_{osn}$, we have*

*(i)*

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{d\hat{\Psi}(X_i,\theta)}{d\theta}\left[u_n^*\right]\right\}'\hat{\Psi}(X_i,\theta) = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{dm(X_i,\theta)}{d\theta}\left[u_n^*\right]\right\}'\Sigma_0(X_i)^{-1}\rho(Z_i,\theta_0) + \langle u_n^*, \theta-\theta_0\rangle + o_p(n^{-1/2});$$

*(ii)* $\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{d^2\hat{\Psi}(X_i,\theta)}{d\theta d\theta}\left[u_n^*,u_n^*\right]\right\}'\hat{\Psi}(X_i,\theta) = o_p(n^{-1/4})$;
*and (iii)* $\frac{1}{n}\sum_{i=1}^{n}\left\|\frac{d\hat{\Psi}(X_i,\theta)}{d\theta}\left[u_n^*\right]\right\|_E^2 = O_p(1)$, $\frac{1}{n}\sum_{i=1}^{n}\left\|\frac{d^2\hat{\Psi}(X_i,\theta)}{d\theta d\theta}\left[u_n^*\right]\right\|_E^2 = O_p(1)$.

*Proof.* The proof is analogous to the ones of Lemma B.3-Lemma B.6 and is omitted for brevity. □

To prove Theorem 3.1, we first present and prove Lemma B.2 and Theorem B.1. Then we show the proof of Theorem 3.1.

**Lemma B.2.** *(Empirical Riesz) Suppose Assumptions of 3.1 hold. Let $\delta_n^*$ be the empirical Riesz representer defined in (3.8), and*

$$\begin{aligned} & B_n \\ \equiv & \begin{pmatrix} I_{11} & I_{n,12} \\ I_{n,21} & I_{n,22} \end{pmatrix} \\ \equiv & \begin{pmatrix} E\left[\left\|\Sigma_0(X)^{-1/2}\frac{dm(X,\theta_0)}{d\beta'}\right\|_E^2\right] & E\left[\left(\frac{dm(X,\theta_0)}{d\beta'}\right)'\Sigma_0(X)^{-1}\left(\frac{dm(X,\theta_0)}{dh}\left[p^{k_n}(\cdot)'\right]\right)\right] \\ I_{n,12}' & E\left[\left\|\Sigma_0(X)^{-1/2}\frac{dm(X,\theta_0)}{dh}\left[p^{k_n}(\cdot)'\right]\right\|_E^2\right] \end{pmatrix}. \end{aligned} \tag{B.2}$$

*Let $\mathbf{v}_n^* = I_{n,22}^{-1}I_{n,21}$. For $\delta_n = (\delta_{\beta,n}', p^{k_n}(\cdot)'\gamma_n)'$, we have*
*(i)*

$$\delta_{\beta,n}^* = I_n^{11}\left(\frac{\partial\phi(\theta_0)}{\partial\beta} - \mathbf{v}_n^{*\prime}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right]\right) \tag{B.3}$$

*and*

$$\delta_{h,n}^* = p^{k_n}(\cdot)'\gamma_n^*,\ \gamma_n^* = I_{n,22}^{-1}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right] - \mathbf{v}_n^*\delta_{\beta,n}^*. \tag{B.4}$$

*(ii) For regular functional $\phi(\theta) = \lambda'\beta$, we have*

$$\begin{aligned} \delta_{\beta,n}^* &= I_n^{11}\frac{\partial\phi(\theta_0)}{\partial\beta} \\ \delta_{h,n}^* &= -\psi^{k_n}(\cdot)'\mathbf{v}_n^*I_n^{11}\frac{\partial\phi(\theta_0)}{\partial\beta} \end{aligned}$$

*and (iii) for irregular functional $\phi(\theta) = \phi(h)$, we have*

$$\begin{aligned} \delta_{\beta,n}^* &= -I_n^{11}\mathbf{v}_n^{*\prime}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right] \\ \delta_{h,n}^* &= p^{k_n}(\cdot)'I_n^{22}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right]. \end{aligned}$$

*Proof.* By the definition of $\delta_n^*$ and Riesz representation theorem,

$$\begin{aligned}\frac{d\phi(\theta_0)}{d\theta}[\delta_n^*] &= \|\delta_n^*\|^2 = \sup_{\delta_n\in\bar{\Delta}_n:\langle\delta_n,\delta_n\rangle\neq 0} \frac{\left|\frac{\partial\phi(\theta_0)}{\partial\beta'}(\delta_\beta)+\frac{\partial\phi(\theta_0)}{\partial h}[\delta_{h,n}]\right|}{E\left[\left(\frac{dm(X,\theta_0)}{d\theta}[\delta_n]\right)'\Sigma_0(X)^{-1}\left(\frac{dm(X,\theta_0)}{d\theta}[\delta_n]\right)\right]} \\ &= \sup_{\eta_n=\left(\delta_\beta',\gamma_n'\right)\in\mathbf{R}^{d_\beta+k_n},\eta_n\neq 0}\frac{\eta_n' A_n A_n'\eta_n}{\eta_n' B_n\eta_n}, \end{aligned} \tag{B.5}$$

where $A_n = \left(\frac{\partial\phi(\theta_0)}{\partial\beta'}, \frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)'\right]\right)'$ and $B_n$ is defined in (B.2) with the inverse of $B_n$ to be

$$B_n^{-1} = \begin{pmatrix} I_n^{11} & I_n^{12} \\ I_n^{21} & I_n^{22}\end{pmatrix} = \begin{pmatrix} \left(I_{11}-I_{n,12}I_{n,22}^{-1}I_{n,21}\right)^{-1} & -I_{11}^{-1}I_{n,21}I_n^{22} \\ -I_{n,22}^{-1}I_{n,21}I_n^{11} & \left(I_{n,22}-I_{n,21}I_{11}^{-1}I_{n,12}\right)^{-1}\end{pmatrix}.$$

By solving (B.5), for $\delta_n = (\delta_{\beta,n}', p^{k_n}(\cdot)'\gamma_n)'$, we have

$$\begin{pmatrix}\delta_{\beta,n}^* \\ \gamma_n^*\end{pmatrix} = B_n^{-1}A_n = \begin{pmatrix} I_n^{11} & I_n^{12} \\ I_n^{21} & I_n^{22}\end{pmatrix}\begin{pmatrix}\frac{\partial\phi(\theta_0)}{\partial\beta} \\ \frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right]\end{pmatrix} = \begin{pmatrix} I_n^{11}\lambda + I_n^{12}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right] \\ I_n^{21}\lambda + I_n^{22}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right]\end{pmatrix}. \tag{B.6}$$

It implies that

$$\begin{aligned}\delta_{\beta,n}^* &= I_n^{11}\frac{\partial\phi(\theta_0)}{\partial\beta} + I_n^{12}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right] = I_n^{11}\frac{\partial\phi(\theta_0)}{\partial\beta} + I_n^{21'}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right] \\ &= I_n^{11}\left(\frac{\partial\phi(\theta_0)}{\partial\beta} - \mathbf{v}_n^{*'}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right]\right), \end{aligned} \tag{B.7}$$

where the last equality is by the definition of $I_n^{21}$ and $\mathbf{v}_n^*$. Moreover, we have

$$I_n^{22} = I_{n,22}^{-1} + I_{n,22}^{-1}I_{n,21}I_n^{11}I_{n,12}I_{n,22}^{-1} \tag{B.8}$$

by the Woodbury matrix identity. Combining (B.6) and (B.8), it follows that

$$\begin{aligned}\gamma_n^* &= I_n^{21}\frac{\partial\phi(\theta_0)}{\partial\beta} + I_n^{22}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right] \\ &= -I_{n,22}^{-1}I_{n,21}I_n^{11}\frac{\partial\phi(\theta_0)}{\partial\beta} + I_{n,22}^{-1}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right] + I_{n,22}^{-1}I_{n,21}I_n^{11}I_{n,12}I_{n,22}^{-1}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right] \\ &= -\mathbf{v}_n^*I_n^{11}\frac{\partial\phi(\theta_0)}{\partial\beta} + I_{n,22}^{-1}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right] + \mathbf{v}_n^*I_n^{11}\mathbf{v}_n^{*'}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right] \\ &= I_{n,22}^{-1}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right] - \mathbf{v}_n^*I_n^{11}\left(\frac{\partial\phi(\theta_0)}{\partial\beta} - \mathbf{v}_n^{*'}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right]\right). \end{aligned} \tag{B.9}$$

Combining (B.7) and (B.9), part (i) follows.

For part (ii), because $A_n = \left(\frac{\partial\phi(\theta_0)}{\partial\beta'}, 0\right)'$, we have

$$\begin{pmatrix}\delta_{\beta,n}^* \\ \gamma_n^*\end{pmatrix} = B_n^{-1}A_n = \begin{pmatrix} I_n^{11} & I_n^{12} \\ I_n^{21} & I_n^{22}\end{pmatrix}\begin{pmatrix}\frac{\partial\phi(\theta_0)}{\partial\beta} \\ 0\end{pmatrix} = \begin{pmatrix} I_n^{11}\frac{\partial\phi(\theta_0)}{\partial\beta} \\ I_n^{21}\frac{\partial\phi(\theta_0)}{\partial\beta}\end{pmatrix},$$

it follows that $\delta_{\beta,n}^* = I_n^{11}\frac{\partial\phi(\theta_0)}{\partial\beta}$ and $\delta_{h,n}^* = p^{k_n}(\cdot)'\gamma_n^* = p^{k_n}(\cdot)'I_n^{21}\frac{\partial\phi(\theta_0)}{\partial\beta} = -p^{k_n}(\cdot)'\mathbf{v}_n^*I_n^{11}\frac{\partial\phi(\theta_0)}{\partial\beta}$.

For part (iii), because $A_n = \left(0, \frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)'\right]\right)'$, we have

$$\begin{pmatrix} \delta^*_{\beta,n} \\ \gamma^*_n \end{pmatrix} = B_n^{-1}A_n = \begin{pmatrix} I_n^{11} & I_n^{12} \\ I_n^{21} & I_n^{22} \end{pmatrix}\begin{pmatrix} 0 \\ \frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right] \end{pmatrix} = \begin{pmatrix} I_n^{12}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right] \\ I_n^{22}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right] \end{pmatrix},$$

we have $\delta^*_{\beta,n} = -I_n^{11}\mathbf{v}_n^{*\prime}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right]$ and $\delta^*_{h,n} = p^{k_n}(\cdot)'\gamma^*_n = p^{k_n}(\cdot)'I_n^{22}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right]$.

□

Theorem B.1 presents a preliminary joint asymptotic normality result for SGMM estimator.

**Theorem B.1.** *Suppose that Assumptions 2.1, 3.1-A.4 and A.1-A.14 hold. Suppose $\phi(\theta) = \lambda'\beta + \phi_h(h)$. Suppose $\phi_h(\cdot)$ satisfies (3.7). Then*

$$\sqrt{n}V_n^{-1/2}\begin{pmatrix} \hat{\beta}_n - \beta_0 \\ \hat{\phi}_h\left(\hat{h}_n\right) - \phi_h\left(h_0\right) \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} I_{d_\beta} & 0 \\ 0 & 1 \end{pmatrix}\right),$$

*where*

$$\begin{aligned}
V_n &= \begin{pmatrix} \Omega_\beta & -\Omega_\beta\upsilon_n' \\ -\upsilon_n\Omega_\beta & \bar{V}_{\phi_{h,n}} + \upsilon_n\Omega_\beta\upsilon_n' \end{pmatrix}_{(d_\beta+1)\times(d_\beta+1)}, \\
\Omega_\beta &= \left(E\left[D_{\varpi^*}(X)'\Sigma_0(X)^{-1}D_{\varpi^*}(X)\right]\right)^{-1}, \\
\upsilon_n &= -\frac{\partial\phi(\theta_0)}{\partial h}[p^{k_n}(\cdot)]'I_{n,22}^{-1} \times I_{n,21} \\
\Omega_{h,n} &= \left(E\left[\frac{dm(x,\theta_0)}{dh}\left[p^{k_n}(\cdot)'\right]'\Sigma_0(X)^{-1}\frac{dm(x,\theta_0)}{dh}\left[p^{k_n}(\cdot)'\right]\right]\right)^{-1}, \\
\bar{V}_{\phi_{h,n}} &= \frac{\partial\phi(\theta_0)}{\partial h}[p^{k_n}(\cdot)']\Omega_{h,n}\frac{\partial\phi(\theta_0)}{\partial h}[p^{k_n}(\cdot)]
\end{aligned}$$

*Proof.* As is shown in the proof of Theorem 3.1,

$$\sqrt{n}\left\langle u_n^*, \hat{\theta}_n - \theta_0\right\rangle = -\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{dm(X_i,\theta_0)}{d\theta}\left[u_n^*\right]'\Sigma_0(X_i)^{-1}\rho(Z_i,\theta_0) + o_p(1),$$

where $u_n^* = \delta_n^*/\left\|\delta_n^*\right\|_{sd}$. By the definition of $\delta_n^*$,

$$\begin{aligned}
\frac{dm(X,\theta_0)}{d\theta}\left[\delta_n^*\right] &= \frac{dm(X,\theta_0)}{d\beta}\left[\delta^*_{\beta,n}\right] + \frac{dm(X,\theta_0)}{dh}\left[\delta^*_{h,n}\right] = \frac{dm(X,\theta_0)}{d\beta}\left(\delta^*_{\beta,n}\right) + \frac{dm(X,\theta_0)}{dh}\left[p^{k_n}(\cdot)'\gamma^*_{h,n}\right] \\
&= \frac{dm(X,\theta_0)}{d\beta}I_n^{11}\left(\frac{\partial\phi(\theta_0)}{\partial\beta} - \mathbf{v}_n^{*\prime}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right]\right) \\
&\quad + \frac{dm(X,\theta_0)}{dh}\left[p^{k_n}(\cdot)'\left\{I_{n,22}^{-1}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right] - \mathbf{v}_n^*I_n^{11}\left(\frac{\partial\phi(\theta_0)}{\partial\beta} - \mathbf{v}_n^{*\prime}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right]\right)\right\}\right]
\end{aligned}$$

where the third equality follows from Lemma B.2 and the fourth equality follows from a direct calculation. Then the variance-covariance matrix

$$\left\|\delta_n^*\right\|^2 = E\left[\frac{dm(X,\theta_0)}{d\theta}\left[\delta_n^*\right]'\Sigma_0(X)^{-1}\frac{dm(X,\theta_0)}{d\theta}\left[\delta_n^*\right]\right]$$

can be decomposed into three terms such that

$$||\delta_n^*||^2 = T_1 + T_2 + 2T_3,$$

where

$$\begin{aligned} T_1 &= E\left[\left\{\left(\frac{dm(X,\theta_0)}{d\beta} - \frac{dm(X,\theta_0)}{dh}\left[p^{k_n}(\cdot)'\mathbf{v}_n^*\right]\right) I_n^{11}\left(\frac{\partial\phi(\theta_0)}{\partial\beta} - \mathbf{v}_n^{*\prime}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right]\right)\right\}' \Sigma_0(X)^{-1} \right. \\ &\quad \left. \times \left\{\left(\frac{dm(X,\theta_0)}{d\beta} - \frac{dm(X,\theta_0)}{dh}\left[p^{k_n}(\cdot)'\mathbf{v}_n^*\right]\right) I_n^{11}\left(\frac{\partial\phi(\theta_0)}{\partial\beta} - \mathbf{v}_n^{*\prime}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right]\right)\right\}\right], \end{aligned}$$

$$T_2 = E\left[\left(\frac{dm(X,\theta_0)}{dh}\left[p^{k_n}(\cdot)'I_{n,22}^{-1}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right]\right]\right)' \times \Sigma_0(X)^{-1}\left(\frac{dm(X,\theta_0)}{dh}\left[p^{k_n}(\cdot)'I_{n,22}^{-1}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right]\right]\right)\right]$$

and

$$\begin{aligned} T_3 &= E\left[\left\{\left(\frac{dm(X,\theta_0)}{d\beta} - \frac{dm(X,\theta_0)}{dh}\left[p^{k_n}(\cdot)'\mathbf{v}_n^*\right]\right) I_n^{11}\left(\frac{\partial\phi(\theta_0)}{\partial\beta} - \mathbf{v}_n^{*\prime}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right]\right)\right\}' \Sigma_0(X)^{-1} \right. \\ &\quad \left. \times \frac{dm(X,\theta_0)}{dh}\left[p^{k_n}(\cdot)'I_{n,22}^{-1}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right]\right]\right]. \end{aligned}$$

For the first term $T_1$, we have

$$\begin{aligned} & T_1 \\ = & \left(\frac{\partial\phi(\theta_0)}{\partial\beta} - \mathbf{v}_n^{*\prime}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right]\right)' \times I_n^{11} \\ & \times E\left[\left(\frac{dm(X,\theta_0)}{d\beta} - \frac{dm(X,\theta_0)}{dh}\left[p^{k_n}(\cdot)'\mathbf{v}_n^*\right]\right)' \Sigma_0(X)^{-1}\right. \\ & \left. \times \left(\frac{dm(X,\theta_0)}{d\beta} - \frac{dm(X,\theta_0)}{dh}\left[p^{k_n}(\cdot)'\mathbf{v}_n^*\right]\right)\right] \\ & \times I_n^{11}\left(\frac{\partial\phi(\theta_0)}{\partial\beta} - \mathbf{v}_n^{*\prime}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right]\right) \\ = & \left(\frac{\partial\phi(\theta_0)}{\partial\beta} - \mathbf{v}_n^{*\prime}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right]\right)' \times I_n^{11} E\left[D_{\varpi_n^*}(X)'\Sigma_0(X)^{-1}D_{\varpi_n^*}(X)\right] \\ & \times I_n^{11} \times \left(\frac{\partial\phi(\theta_0)}{\partial\beta} - \mathbf{v}_n^{*\prime}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right]\right), \end{aligned}$$

where the first equality follows from direct calculation and the second equality follows from the definition of $D_{\varpi_n^*}(X)$. For the second term $T_2$, by the definition of $\Omega_{h,n}$ and linearity of pathwise derivatives, we have

$$\begin{aligned} T_2 &= \frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)'\right] I_{n,22}^{-1} E\left[\frac{dm(X,\theta_0)}{dh}\left[p^{k_n}(\cdot)'\right]' \Sigma_0(X)^{-1}\frac{dm(X,\theta_0)}{dh}\left[p^{k_n}(\cdot)'\right]\right] \times I_{n,22}^{-1}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right] \\ &= \frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)'\right] I_{n,22}^{-1}\Omega_h I_{n,22}^{-1}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right] \\ &= \frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)'\right]\Omega_h^{-1}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right], \end{aligned}$$

where the last equality follows from the fact that $I_{n,22}^{-1} = \Omega_{h,n}$.

For the third term $T_3$, we have

$$\begin{aligned}
T_3 &= \left(\frac{\partial\phi(\theta_0)}{\partial\beta} - \mathbf{v}_n^{*\prime}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right]\right)' I_n^{11} \\
&\quad \times E\left[\left(\frac{dm(X,\theta_0)}{d\beta'} - \frac{dm(X,\theta_0)}{dh}\left[p^{k_n}(\cdot)'\mathbf{v}_n^*\right]\right)'\Sigma_0(X)^{-1}\frac{dm(X,\theta_0)}{dh}\left[p^{k_n}(\cdot)'\right]\right] \times I_{n,22}^{-1}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(y_2)\right] \\
&= \left(\frac{\partial\phi(\theta_0)}{\partial\beta} - \mathbf{v}_n^{*\prime}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right]\right)' I_n^{11}\left[I_{n,12} - I_{n,12}\times I_{n,22}\times I_{n,22}^{-1}\right]\times I_{n,22}^{-1}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right] \\
&= 0,
\end{aligned}$$

where the last equality follows by the definition of $B_n$ and $\mathbf{v}_n^*$. Let $\upsilon_n$ be a $1\times d_\beta$ vector such that

$$\upsilon_n = -\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)'\right] I_{n,22}^{-1} I_{n,21} = -\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)'\right]\mathbf{v}_n^*$$

Thus,

$$\begin{aligned}
\|\delta_n^*\|^2 &= \left(\frac{\partial\phi(\theta_0)}{\partial\beta} - \mathbf{v}_n^{*\prime}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right]\right)' I_n^{11}\times E\left[D_{\varpi_n^*}(X)'\Sigma_0(X)^{-1}D_{\varpi_n^*}(X)\right]\times I_n^{11} \\
&\quad \times\left(\frac{\partial\phi(\theta_0)}{\partial\beta} - \mathbf{v}_n^{*\prime}\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right]\right) \\
&\quad +\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)'\right]\Omega_h\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right] \\
&\to \left(\frac{\partial\phi(\theta_0)}{\partial\beta} + \upsilon_n'\right)'\Omega_\beta\left(\frac{\partial\phi(\theta_0)}{\partial\beta} + \upsilon_n'\right) + \frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)'\right]\Omega_h\frac{\partial\phi(\theta_0)}{\partial h}\left[p^{k_n}(\cdot)\right] \\
&= \left(\lambda', \frac{\partial\phi(\theta_0)}{\partial h}[p^{k_n}(\cdot)']\right)\begin{pmatrix}\Omega_\beta & -\Omega_\beta\mathbf{v}_n^{*\prime} \\ -\mathbf{v}_n^*\Omega_\beta & \Omega_h + \mathbf{v}_n^*\Omega_\beta\mathbf{v}_n^{*\prime}\end{pmatrix}\left(\lambda', \frac{\partial\phi(\theta_0)}{\partial h}[p^{k_n}(\cdot)']\right)'.
\end{aligned}$$

This complete the proof.

□

**Proof of Theorem 3.1.**

Following the proof in Theorem B.1, let $\frac{\partial\phi(\theta_0)}{\partial\beta} = \lambda$, where $\lambda$ is a $d_\beta-$vector of ones. Then

$$\begin{aligned}
&\mathrm{Var}\left(\frac{dm(X_i,\theta_0)}{d\theta}\left[\delta_{\beta,n}^* + V_{\phi_{h,n}}^{-1/2}\delta_{h,n}^*\right]'\Sigma_o(X_i)^{-1}\rho(Z_i,\theta_0)\right) \\
\to &\left(\lambda + V_{\phi_{h,n}}^{-1/2}\upsilon_n'\right)'\Omega_\beta^{-1}\left(\lambda + V_{\phi_{h,n}}^{-1/2}\upsilon_n'\right) + V_{\phi_{h,n}}^{-1/2}V_{\phi_{h,n}}V_{\phi_{h,n}}^{-1/2} \\
\to &\left(\lambda + V_{\phi_{h,n}}^{-1/2}\upsilon_n'\right)'\Omega_\beta^{-1}\left(\lambda + V_{\phi_{h,n}}^{-1/2}\upsilon_n'\right) + 1 \to \lambda'\Omega_\beta^{-1}\lambda + 1.
\end{aligned}$$

The conclusion follows by employing Wald's device. □

**Proof of Theorem 3.2.**

See the Supplemental Appendix.

### B.1.2 Proofs of Section 3.2

We first present several lemmas that are useful for the proof of main theorems in this section. The proofs of these lemmas are in the Supplemental Appendix.

**Lemma B.3.** *Suppose that Assumptions in Theorem 3.3 hold. Then uniformly over $w \in \mathcal{W}$*

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \frac{dm(X_i, \theta_0)}{d\theta} \left[ u_n^*(w) \right] \right\}' \Sigma_0(X_i)^{-1/2} \hat{\Psi}(X_i, \theta_0) \\ = & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \frac{dm(X_i, \theta_0)}{d\theta} \left[ u_n^*(w) \right] \right\}' \Sigma_0(X_i)^{-1/2} \rho(Z_i, \theta_0) + o_p(1). \end{aligned}$$

**Lemma B.4.** *Suppose that Assumptions of Theorem 3.3 hold. Then uniformly over $w \in \mathcal{W}$*

$$\begin{aligned} & \mathbb{G}_n \left[ \frac{dm(X_i, \theta_0)}{d\theta} \left[ u_n^*(w) \right] \Sigma_o(X_i)^{-1/2} \left\{ \hat{\Psi}(X_i, \theta_n) - \hat{\Psi}(X_i, \theta_0) \right\} \right] \\ = & \mathbb{G}_n \left[ \frac{dm(X_i, \theta_0)}{d\theta} \left[ u_n^*(w) \right] \Sigma_o(X_i)^{-1} \left\{ m(X_i, \theta_n) - m(X_i, \theta_0) \right\} \right] + o_p(1) \end{aligned}$$

*uniformly over $(w, \theta_n) \in \mathcal{W} \times \mathcal{N}_{osn}$*

**Lemma B.5.** *Suppose that Assumptions of Theorem 3.3 hold. Then*

$$\mathbb{G}_n \left[ \frac{dm(X_i, \theta_0)}{d\theta} \left[ u_n^*(w) \right]' \Sigma_0(X_i)^{-1} \left\{ m(X_i, \theta_n) - m(X_i, \theta_0) \right\} \right] = o_p(1)$$

*uniformly over $(w, \theta_n) \in \mathcal{W} \times \mathcal{N}_{osn}$.*

**Lemma B.6.** *Suppose Assumptions in Theorem 3.3 hold. Then*

$$\frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{d\hat{\Psi}(X_i, \theta_n)}{d\theta} \left[ u_n^*(w) \right] \right\}' \hat{\Psi}(X_i, \theta_n) = \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{d\Psi(X_i, \theta_0)}{d\theta} \left[ u_n^*(w) \right] \right\}' \hat{\Psi}(X_i, \theta_n) + o_p\left( n^{-1/2} \right)$$

*uniformly on $(w, \theta_n) \in \mathcal{W} \times \mathcal{N}_{osn}$.*

**Proof of Theorem 3.3.**

The proof consists of two main steps. Step 1 establishes uniform linearization properties. Step 2 provides the strong approximation results.

**Step 1.** The goal is to establish that

$$\left\langle u_n^*(w), \hat{\theta}_n - \theta_0 \right\rangle = \frac{1}{n} \sum_{i=1}^{n} \frac{dm(X_i, \theta_0)}{d\theta} \left[ u_n^*(w) \right]' \Sigma_0(X_i)^{-1} \rho(Z_i, \theta_0) + o_p\left( n^{-1/2} \right).$$

for all $w \in \mathcal{W}$.

Similar to the proof of Theorem 3.1, by Assumption A.14 on the uniform derivatives and second-order Taylor expansion, for $t_n = o_p(n^{-1/2})$, we have

$$\left. \frac{d\hat{L}_n\left(\theta(t_n)\right)}{dt_n} \right|_{t_n = 0} = \left( \frac{1}{n} \sum_{i=1} \frac{d\rho(Z_i, \hat{\theta}_n)}{d\theta} \left[ t_n u_n^*(w) \right] \otimes q_i \right)' \left( \frac{1}{n} \sum_{i=1}^{n} g_i\left(\bar{\theta}_n\right) g_i\left(\bar{\theta}_n\right)' \right)^{-1} \hat{g}\left(\hat{\theta}_n\right), \quad \text{(B.10)}$$

and

$$\begin{aligned}
&\left.\frac{d^2 L_n\left(\theta(t_n)\right)}{dt_n^2}\right|_{t=s} \\
= &\left(\frac{1}{n}\sum_{i=1}\frac{d\rho(z_i,\theta(s))}{dh}\left[t_n u_n^*\right]\otimes q_i\right)'\left(\frac{1}{n}\sum_{i=1}^n g_i\left(\bar{\theta}_n\right) g_i\left(\bar{\theta}_n\right)'\right)^{-1}\times\left(\frac{1}{n}\sum_{i=1}\frac{d\rho(z_i,\theta(s))}{dh}\left[t_n u_n^*\right]\otimes q_i\right) \\
&+\left(\frac{1}{n}\sum_{i=1}^n\frac{d^2\rho(z_i,\theta(s))}{d\theta d\theta}\left[t_n u_n^*, t_n u_n^*\right]\otimes q_i\right)'\left(\frac{1}{n}\sum_{i=1}^n g_i\left(\bar{\theta}_n\right) g_i\left(\bar{\theta}_n\right)\right)^{-1}\hat{g}\left(\theta(s)\right)
\end{aligned}$$

for some $s\in[0,1]$. By Assumption A.7,

$$\sup_{(\theta_n,w)\in\mathcal{N}_{0sn}\times\mathcal{W}}\left\|\frac{1}{n}\sum_{i=1}^n\frac{d\rho(Z_i,\theta_n)}{d\theta}\left[u_n^*(w)\right]\otimes q^{s_n}(X_i)\right\|_E=O_p(1)$$

and

$$\sup_{(\theta_n,w)\in\mathcal{N}_{0sn}\times\mathcal{W}}\left\|\frac{1}{n}\sum_{i=1}^n\frac{d^2\rho(Z_i,\theta_n)}{d\theta d\theta}\left[u_n^*(w),u_n^*(w)\right]\otimes q^{s_n}(X_i)\right\|_E=O_p(1).$$

It implies that $\left.\frac{d^2 L_n(\theta(t))}{dt^2}\right|_{t=s}=O_p\left(t_n^2\right)$. Furthermore,

$$\begin{aligned}
&\left.\frac{d\hat{L}_n\left(\theta(t_n)\right)}{dt_n}\right|_{t_n=0} \\
= &\left(\frac{1}{n}\sum_{i=1}\frac{d\rho(Z_i,\hat{\theta}_n)}{d\theta}\left[t_n u_n^*(w)\right]\otimes q_i\right)'\left(\frac{1}{n}\sum_{i=1}^n\Sigma(X_i,\bar{\theta}_n)\otimes q_i q_i'\right)^{-1}\hat{g}\left(\hat{\theta}_n\right)+o_p\left(t_n n^{-1/2}\right) \\
= &\frac{1}{n}\sum_{i=1}^n\left\{Q(X_i,\bar{\theta}_n)\left(Q(\bar{\theta}_n)'Q(\bar{\theta}_n)\right)^{-1}\sum_{j=1}^n Q(X_j,\bar{\theta}_n)\Sigma(X_j,\bar{\theta}_n)^{-1/2}\frac{d\rho(Z_i,\hat{\theta}_n)}{d\theta}\left[t_n u_n^*(w)\right]\right\}' \\
&\times\left\{Q(X_i,\bar{\theta}_n)\left(Q(\bar{\theta}_n)'Q(\bar{\theta}_n)\right)^{-1}\sum_{j=1}^n Q(X_j,\bar{\theta}_n)\Sigma(X_j,\bar{\theta}_n)^{-1/2}\rho(Z_j,\hat{\theta}_n)\right\}+o_p(n^{-1}), \qquad \text{(B.11)}
\end{aligned}$$

where the first equality follows from Lemma B.7 and the second equality follows from direct calculations. By Lemma C.1 and Lemma B.6, (B.11) can be simplified to

$$\left.\frac{d\hat{L}_n\left(\theta(t_n)\right)}{dt_n}\right|_{t_n=0}=\frac{1}{n}\sum_{i=1}^n\frac{dm(X_i,\theta_0)}{d\theta}\left[u_n^*(w)\right]'\Sigma_0(X_i)^{-1/2}\hat{\Psi}(X_i,\hat{\theta}_n)+o_p(n^{-1}).$$

Note that for all $(w, \theta_n) \in \mathcal{W} \times \mathcal{N}_{osn}$, by direct calculation, we have

$$\begin{aligned}
& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{dm(X_i, \theta_0)}{d\theta} \left[u_n^*(w)\right]' \Sigma_0(X_i)^{-1/2} \left\{ \hat{\Psi}(X_i, \theta_n) - \hat{\Psi}(X_i, \theta_0) \right\} \\
= \; & \mathbb{G}_n \left[ \frac{dm(X_i, \theta_0)}{d\theta} \left[u_n^*(w)\right]' \Sigma_0(X_i)^{-1/2} \left\{ \hat{\Psi}(X_i, \theta_n) - \hat{\Psi}(X_i, \theta_0) \right\} \right] \\
& + \sqrt{n} E \left[ \frac{dm(X_i, \theta_0)}{d\theta} \left[u_n^*(w)\right]' \Sigma_0(X_i)^{-1/2} \left\{ \hat{\Psi}(X_i, \theta_n) - \hat{\Psi}(X_i, \theta_0) \right\} \right]. \qquad \text{(B.12)}
\end{aligned}$$

By Lemma B.4, for all $w \in \mathcal{W}$,

$$\begin{aligned}
& \mathbb{G}_n \left[ \frac{dm(X_i, \theta_0)}{d\theta} \left[u_n^*(w)\right]' \Sigma_0(X_i)^{-1/2} \left\{ \hat{\Psi}(X_i, \theta_n) - \hat{\Psi}(X_i, \theta_0) \right\} \right] \\
= \; & \mathbb{G}_n \left[ \frac{dm(X_i, \theta_0)}{d\theta} \left[u_n^*(w)\right]' \Sigma_0(X_i)^{-1} \left\{ m(X_i, \theta_n) - m(X_i, \theta_0) \right\} \right] + o_p(1) \\
= \; & o_p(1),
\end{aligned}$$

where the second equality follows from Lemma B.5.

For second term on the right hand side of (B.12), note that uniformly over $\theta_n \in \mathcal{N}_{osn}$ and $w \in \mathcal{W}$,

$$\begin{aligned}
& E \left[ \left| \frac{dm(X_i, \theta_0)}{d\theta} \left[u_n^*(w)\right]' \Sigma_0(X_i)^{-1/2} \left\{ \hat{\Psi}(X_i, \theta_n) - \Psi(X_i, \theta_n) \right\} \right|^2 \right] \\
\leq \; & E \left[ \left\| \frac{dm(X_i, \theta_0)}{d\theta} \left[u_n^*(w)\right]' \Sigma(X_i)^{-1/2} \right\|_E^2 \right] \times \sup_{x \in \mathcal{X}, \theta \in \mathcal{N}_{osn} \cup \{\theta_0\}} \left\| \hat{\Psi}(X_i, \theta_n) - \Psi(X_i, \theta_n) \right\|_E^2 \\
= \; & O_p\left(\varrho_n^2\right) = o_p(n^{-1/2}) \qquad \text{(B.13)}
\end{aligned}$$

where the last equality follows from Lemma C.1. Similarly,

$$E \left[ \left| \frac{dm(X_i, \theta_0)}{d\theta} \left[u_n^*(w)\right]' \Sigma_0(X_i)^{-1/2} \left\{ \hat{\Psi}(X_i, \theta_0) - \Psi(X_i, \theta_0) \right\} \right|^2 \right] = o_p(n^{-1/2}). \qquad \text{(B.14)}$$

Combining (B.13) and (B.14), we have

$$\begin{aligned}
& E \left[ \frac{dm(X_i, \theta_0)}{d\theta} \left[u_n^*(w)\right]' \Sigma_0(X_i)^{-1/2} \left\{ \hat{\Psi}(X_i, \theta_n) - \hat{\Psi}(X_i, \theta_0) \right\} \right] \\
= \; & E \left[ \frac{dm(X_i, \theta_0)}{d\theta} \left[u_n^*(w)\right]' \Sigma_0(X_i)^{-1} \left\{ m(X_i, \theta_n) - m(X_i, \theta_0) \right\} \right] \\
= \; & E \left[ \frac{dm(x_i, \theta_0)}{d\theta} \left[u_n^*(w)\right]' \Sigma_0(X_i)^{-1} \left\{ \frac{dm\left(X_i, \theta_n\right)}{d\theta} \left[\theta_n - \theta_0\right] \right\} \right] \\
= \; & E \left[ \frac{dm(x_i, \theta_0)}{d\theta} \left[u_n^*(w)\right]' \Sigma_0(X_i)^{-1} \left\{ \frac{dm\left(X_i, \theta_n\right)}{d\theta} \left[\theta_n - \theta_0\right] - \frac{dm\left(X_i, \theta_0\right)}{d\theta} \left[\theta_n - \theta_0\right] \right\} \right] \\
& + \left\langle u_n^*(w), \theta_n - \theta_0 \right\rangle + o_p\left(n^{-1/2}\right) = \left\langle u_n^*(w), \theta_n - \theta_0 \right\rangle + o_p\left(n^{-1/2}\right),
\end{aligned}$$

where the last equality follows from Assumption A.14 that

$$\begin{aligned} & E\left[\left\{\frac{dm(X_i,\theta_0)}{d\theta}\left[u_n^*(w)\right]\right\}'\Sigma_0(X_i)^{-1}\left\{\frac{dm(X_i,\theta_n)}{d\theta}\left[\theta_n-\theta_0\right]-\frac{dm(X_i,\theta_0)}{d\theta}\left[\theta_n-\theta_0\right]\right\}\right] \\ = \quad & o_p(n^{-1/2}) \end{aligned}$$

uniformly on $\theta\in\mathcal{N}_0$ and $\theta_n\in\mathcal{N}_{0sn}$ and $w\in\mathcal{W}$. Therefore, uniformly over $\theta_n\in\mathcal{N}_{osn}$ and $w\in\mathcal{W}$, we have

$$\begin{aligned} & \left.\frac{d\hat{L}_n\left(\theta(t_n)\right)}{dt_n}\right|_{t_n=0} \\ = \quad & \frac{t_n}{n}\sum\frac{dm(Z_i,\theta_0)}{d\theta}\left[u_n^*(w)\right]'\Sigma(X_i,\theta_0)^{-1}\rho(Z_i,\theta_0)+t_n\left\langle u_n^*(w),\hat{\theta}_n-\theta_0\right\rangle+o_p(t_n n^{-1/2}), \end{aligned}$$

and it implies that

$$\begin{aligned} \sqrt{n}\left\langle u_n^*(w),\hat{\theta}_n-\theta_0\right\rangle &= -\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{dm(X_i,\theta_0)}{d\theta}\left[u_n^*(w)\right]'\Sigma_o(X_i)^{-1}\rho(Z_i,\theta_0)+o_p\left(1\right) \\ &= -\alpha_n(w)'\Omega_n^{1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{dm(X_i,\theta_0)}{d\theta}\left[\bar{p}^{k_n}(\cdot)'\right]'\Sigma_o(X_i)^{-1}\rho(Z_i,\theta_0)+o_p(1) \end{aligned}$$

where the second equality follows from the definition that $\alpha_n(w)=\frac{\Omega_n^{1/2}A_n(w)}{\left\|\Omega_n^{1/2}A_n(w)\right\|_E}$ and $u_n^*(w)$.

By Assumption A.15,

$$\begin{aligned} \left\|\alpha_n\left(w_1\right)-\alpha_n\left(w_2\right)\right\|_E &\lesssim \left\|\frac{A_n\left(w_1\right)}{\left\|\Omega_n^{1/2}A_n(w_1)\right\|_E}-\frac{A_n(w_2)}{\left\|\Omega_n^{1/2}A_n(w_2)\right\|_E}\right\|_E \\ &\lesssim \frac{\left\|A_n(w_1)-A_n(w_2)\right\|_E}{\left\|\Omega_n^{1/2}A_n(w_1)\right\|_E}+\left\|A_n(w_2)\right\|_E\left|\frac{1}{\left\|\Omega_n^{1/2}A_n(w_1)\right\|_E}-\frac{1}{\left\|\Omega_n^{1/2}A_n(w_2)\right\|_E}\right| \\ &= \frac{\left\|A_n(w_1)-A_n(w_2)\right\|_E}{\left\|\Omega_n^{1/2}A_n(w_1)\right\|_E}+\left\|A_n(w_2)\right\|_E\frac{\left|\left\|\Omega_n^{1/2}A_n(w_2)\right\|_E-\left\|\Omega_n^{1/2}A_n(w_1)\right\|_E\right|}{\left\|\Omega_n^{1/2}A_n(w_1)\right\|_E\left\|\Omega_n^{1/2}A_n(w_2)\right\|_E} \\ &\lesssim \frac{\left\|A_n(w_1)-A_n(w_2)\right\|_E}{\left\|A_n(w_1)\right\|_E}\lesssim\xi_{k_n}^L\left\|w_1-w_2\right\|_E \end{aligned}$$

where the last inequality follows from Assumption A.15.

For $l=1\ldots,d_\rho$, since $\max_{1\le t,l\le d_\rho}\left|\Sigma_{tl}(X,\theta_0)^{-1}\right|=O_p(1)$, consider

$$\mathcal{F}=\left\{f=\alpha_n(w)\frac{dm_l(X,\theta_0)}{d\theta}\left[\bar{p}^{k_n}(\cdot)'\right]\Sigma_0(X)^{-1}\rho_l(Z,\theta_0)\right\}. \tag{B.15}$$

Then $\mathrm{Var}\left[\alpha_{n,}(w)'\frac{dm_l(X,\theta_0)}{d\theta}\left[\bar{p}^{k_n}(\cdot)'\right]\Sigma_0(X)^{-1}\rho_l(Z,\theta_0)\right]\lesssim 1$ and

$$\sup_{w\in\mathcal{W}}\left|\alpha_{n,l}(w)\frac{dm_l(X,\theta_0)}{d\theta}\left[\bar{p}^{k_n}(\cdot)'\right]\Sigma_0(X)^{-1}\rho_l(Z,\theta_0)\right|\lesssim\xi_{\rho,k_n}n^{1/m}=o_p(b_n^{-1}),$$

where the inequality follows from Assumption 3.5 (ii), A.15 A.17 and A.19. Furthermore, for $w^1, w^2 \in \mathcal{W}$,

$$\begin{aligned} & f(w^1) - f(w^2) \\ \leq\ & \xi_{k_n}^L \left\|w^1 - w^2\right\| \max_{1\leq i\leq n} |\rho_t(Z_i,\theta_0)| \left\|\Sigma_0(X_i)^{-1}\right\| \max_{1\leq i\leq n} \left\| \frac{dm_l(X_i,\theta_0)}{d\theta} \left[\bar{p}^{k_n}(\cdot)\right] \right\| \\ \lesssim\ & \xi_{k_n}^L \left\|w^1 - w^2\right\| n^{1/m} \xi_{\rho,k_n}. \end{aligned}$$

It implies that for some $C > 0$,

$$\sup_Q N\left(\mathcal{F}, L^2(Q), \varepsilon n^{1/m}\xi_{\rho,k_n}\right) \leq \left(C\xi_k^L/\varepsilon\right)^{d_w}.$$

Thus, by Theorem 6.1 in BCCK (see also Giné and Koltchinskii, 2006),

$$E\left[\sup_{w\in\mathcal{W}} \left| \alpha_n(w)' \mathbb{G}_n \left[ \frac{dm(X_i,\theta_0)}{d\theta} \left[\bar{p}^{k_n}(\cdot)'\right] \Sigma_0(X_i)^{-1} \rho(Z_i,\theta_0) \right] \right| \right] \lesssim \sqrt{\log n}. \tag{B.16}$$

**Step 2.** (Strong Approximation)
Our proof applies Yurinskii's coupling (Theorem 10 in Pollard, 2002). Let

$$S_{1i} = \Omega_n^{1/2} \frac{dm(X_i,\theta_0)}{d\theta} \left[\bar{p}^{k_n}(\cdot)'\right]' \Sigma_0(X_i)^{-1} \rho(Z_i,\theta_0)$$

be a copy of the first order approximation to our estimator. As all eigenvalues of $\Omega_n$ and $\Sigma_0$ are bounded away from zero,

$$\begin{aligned} E\left\|S_{1i}\right\|^3 &\lesssim \left\|\Sigma_0(X_i)^{-1}\right\|^3 E\left[ \max_{1\leq l\leq d_\rho} \left\| \frac{dm_l(X_i,\theta_0)}{d\theta} \left[\bar{p}^{k_n}(\cdot)'\right] \right\|^3 \right] \left( \max_{1\leq l\leq d_\rho} \max_{1\leq i\leq n} E\left[ \left\|\rho_l(Z_i,\theta_0)\right\|^3 |X_i \right] \right) \\ &\lesssim \sup_{1\leq i\leq n} \max_{1\leq l\leq d_\rho} E\left[ \left\| \frac{dm_l(X_i,\theta_0)}{d\theta} \left[\bar{p}^{k_n}(\cdot)'\right] \right\|^2 \right] \sup_{1\leq i\leq n} \max_{1\leq l\leq d_\rho} \left\| \frac{d\rho_l(Z_i,\theta_0)}{d\theta} \left[\bar{p}^{k_n}(\cdot)'\right] \right\| \\ &\lesssim (k_n + d_\beta)\, \xi_{\rho,k_n} < \infty, \end{aligned}$$

where we used the contraction property of conditional expectation, Assumption A.7 and A.17-A.19. By applying Yurinskii's coupling, $\forall \varepsilon > 0$, as $b_n^6 \left(k_n + d_\beta\right)^2 \xi_{\rho,k_n}^2 \log^2 n/n \to 0$, we have

$$\begin{aligned} \Pr\left\{ \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n S_{1i} - \mathcal{N}_{d_\beta+k_n} \right\| > 3\delta b_n^{-1} \right\} &\lesssim \frac{n\left(k_n + d_\beta\right)^2 \xi_{\rho,k_n}}{\left(\delta b_n^{-1}\sqrt{n}\right)^3} \left(1 + \left(k_n + d_\beta\right)^{-1} \log\left(k_n + d_\beta\right)^2 \xi_{\rho,k_n}\right) \\ &\lesssim \frac{b_n^3 (k_n + d_\beta)^2 \xi_{\rho,k_n}}{\delta^3\sqrt{n}} \left(1 + \frac{\log n}{(k_n + d_\beta)}\right) \to 0. \end{aligned}$$

It implies that

$$\sqrt{n}\left\langle u_n^*(w), \hat{\theta}_n - \theta_0 \right\rangle = \frac{A_n(w)}{\left\|\Omega_n^{1/2} A_n(w)\right\|_E}' \Omega_n \mathbb{G}_n \left[ \frac{dm(X_i,\theta_0)}{d\theta} \left[\bar{p}^{k_n}(\cdot)'\right]' \Sigma_o(X_i)^{-1} \rho(Z_i,\theta_0) \right] + o_p\left(b_n^{-1}\right).$$

Thus, by the definition of $S_{1i}$, we have

$$\begin{aligned}\sqrt{n}\left\langle u_n^*(w), \hat{\theta}_n - \theta_0 \right\rangle &= \frac{A_n(w)'\Omega_n^{1/2}}{\left\|\Omega_n^{1/2}A_n(w)\right\|_E}\Omega_n^{1/2}\mathbb{G}_n S_{1i} + o_p\left(b_n^{-1}\right) \\ &= \frac{A_n(w)'\Omega_n^{1/2}}{\left\|\Omega_n^{1/2}A_n(w)\right\|_E}\mathcal{N}_{d_\beta+k_n} + o_p\left(b_n^{-1}\right)\end{aligned}$$

in $\ell^\infty(\mathcal{W})$. Therefore, under the assumption that $\sup_w \sqrt{n}\left|r_n(w)\right| / \left\|\delta_n^*(w)\right\| = o_p(b_n^{-1})$, we have

$$\frac{\sqrt{n}\left(\hat{\phi}_n(w) - \phi_0(w)\right)}{\|\delta_n^*\|} = \frac{\sqrt{n}\left\langle \delta_n^*(w), \hat{\theta}_n - \theta_0 \right\rangle}{\left\|A_n(w)'\Omega_n^{1/2}\right\|_E} + o_p\left(b_n^{-1}\right) = \frac{A_n(w)'\Omega_n^{1/2}}{\left\|A_n(w)'\Omega_n^{1/2}\right\|_E}\mathcal{N}_{d_\beta+k_n} + o_p\left(b_n^{-1}\right)$$

in $\ell^\infty(\mathcal{W})$. □

### B.1.3 Proofs of Section 3.3

The following lemma implies results in Lemma 4.1.

**Lemma B.7.** *(Consistency and Convergence Rates of Variance Estimators) Suppose Assumptions of 3.3 hold. Let* $b_n^{-1} = o_p\left(\frac{1}{\sqrt{\log n}}\right)$. *Then uniformly over* $w \in \mathcal{W}$, *we have (i)* $\frac{\left\|\hat{\delta}_n^*(w) - \delta_n^*(w)\right\|}{\|\delta_n^*(w)\|} = o_p\left(b_n^{-1}\right)$, *(ii)* $\left|\frac{\left\|\hat{\delta}_n^*(w)\right\|}{\|\delta_n^*(w)\|} - 1\right| = o_p\left(b_n^{-1}\right)$, *(iii)* $\|\hat{u}_n^*(w) - u_n^*(w)\| = o_p\left(b_n^{-1}\right)$, *(iv)* $\frac{\left\|\tilde{\delta}_n^*(w) - \delta_n^*(w)\right\|}{\|\delta_n^*(w)\|} = o_p\left(b_n^{-1}\right)$, *(v)* $\left|\frac{\left\|\tilde{\delta}_n^*(w)\right\|}{\|\delta_n^*(w)\|} - 1\right| = o_p\left(b_n^{-1}\right)$ *and (vi)* $||\tilde{u}^*(w) - u_n^*(w)|| = o_p(b_n^{-1})$.

The proof of Lemma B.7 is presented in the Supplemental Appendix.

**Proof of Theorem 3.5.**

We show the proof in two main steps. In the first step, we prove that

$$t_n(w) = t^*(w) + o_p(b_n^{-1})$$

uniformly over $w \in \mathcal{W}$. And in the second step, we prove that

$$\Pr\left\{\sup_{w\in\mathcal{W}}|t_n(w)| \le c_n(1-\tau)\right\} = 1 - \tau + o(1).$$

**Step 1.** Assumption A.16 implies that for for all $w \in \mathcal{W}$

$$\begin{aligned}&\sqrt{n}\frac{\phi(\hat{\theta}_n)[w] - \phi(\theta_0)[w]}{\|\delta_n^*(w)\|} = \sqrt{n}\frac{\frac{\partial\phi(\theta_0)}{\partial\theta}[\hat{\theta}_n - \theta_0][w]}{\|\delta_n^*(w)\|} + o_p(1) = \sqrt{n}\frac{\frac{\partial\phi(\theta_0)}{\partial\theta}[\hat{\theta}_n - \theta_{0,n} + \theta_{0,n} - \theta_0][w]}{\|\delta_n^*(w)\|} + o_p(1) \\ &= \sqrt{n}\left\langle \delta_n^*(w), \hat{\theta}_n - \theta_{0n} \right\rangle / \|\delta_n^*(w)\| + o_p(1) = \sqrt{n}\left\langle \delta_n^*(w), \hat{\theta}_n - \theta_0 \right\rangle / \|\delta_n^*(w)\| + o_p(1)\end{aligned}$$

where the last equality is by the orthogonality property of $\theta_{0,n}$.

Moreover, by the triangle inequality,

$$\begin{aligned}
&\left|\frac{\left\langle \hat{\delta}_n^*(w), \hat{\theta}_n - \theta_0 \right\rangle_n}{\left\|\hat{\delta}_n^*(w)\right\|_n} - \frac{\left\langle \delta_n^*(w), \hat{\theta}_n - \theta_0 \right\rangle}{\|\delta_n^*(w)\|}\right| \\
\lesssim\ &\left|\frac{\left\langle \hat{\delta}_n^*(w) - \delta_n^*(w), \hat{\theta}_n - \theta_0 \right\rangle_n}{\left\|\hat{\delta}_n^*(w)\right\|_n}\right| + \frac{\left|\left\langle \hat{\delta}_n^*(w), \hat{\theta}_n - \theta_0 \right\rangle\right|}{\left\|\hat{\delta}_n^*(w)\right\|}\left|1 - \frac{\left\|\hat{\delta}_n^*(w)\right\|_n}{\|\delta_n^*(w)\|}\right| \qquad \text{(B.17)}
\end{aligned}$$

uniformly over $w \in \mathcal{W}$.

By triangle inequality and Lemma B.7, $\left|\frac{\left\langle \hat{\delta}_n^*(w) - \delta_n^*(w), \hat{\theta}_n - \theta_0 \right\rangle_n}{\left\|\hat{\delta}_n^*(w)\right\|_n}\right| = o_p(n^{-1/2})$, $\left|1 - \frac{\left\|\hat{\delta}_n^*(w)\right\|}{\|\delta_n^*(w)\|}\right| = o_p\left(b_n^{-1}\right)$ and $\frac{\left|\left\langle \hat{\delta}_n^*(w), \hat{\theta}_n - \theta_0 \right\rangle\right|}{\left\|\hat{\delta}_n^*(w)\right\|} = O_p(n^{-1/2})$, it implies that (B.17) is $o_p(n^{-1/2})$.

Furthermore, Theorem 3.3 implies that

$$\sup_{w \in \mathcal{W}}\left|\frac{\sqrt{n}\left\langle \delta_n^*(w), \hat{\theta}_n - \theta_0 \right\rangle}{\|\delta_n^*(w)\|} - \frac{A_n(w)'\Omega_n^{1/2}\mathcal{N}_{d_\beta + k_n}}{\left\|A_n(w)'\Omega_n^{1/2}\right\|}\right| = o_p(1).$$

Therefore,

$$\frac{\phi\left(\hat{\theta}_n\right)[w] - \phi(\theta_0)[w]}{\hat{\sigma}_{\phi,n}(w)} = \frac{A_n(w)'\Omega_n^{1/2}\mathcal{N}_{d_\beta + k_n}}{\left\|A_n(w)'\Omega_n^{1/2}\right\|} + o_p(1)$$

**in** $\ell^\infty(\mathcal{W})$

**Step 2.** Note that by triangle inequality,

$$\left|\sup_{w \in \mathcal{W}}\left|\hat{t}_n^*(w)\right| - \sup_{w \in \mathcal{W}}\left|t_n^*(w)\right|\right| \leq \sup_{w \in \mathcal{W}}\left|\left(\frac{A(w)'\hat{\Omega}^{1/2}}{\sqrt{n}\hat{\sigma}_{\hat{\phi}}(w)} - \frac{A(w)'\Omega^{1/2}}{\sqrt{n}\sigma_\phi(w)}\right)\mathcal{N}_{d_\beta + k_n}\right|.$$

Let $E_{\mathcal{N}_{d_\beta+k_n}}[\cdot]$ be the expectation with respect to the distribution of $\mathcal{N}_{d_\beta+k_n}$. Conditionally on the data,

$$\begin{aligned}
E_{\mathcal{N}_{d_\beta+k_n}}\left[\left(\hat{t}_n^*(w) - t_n^*(w)\right)^2\right]^{1/2} &= \left\|\frac{\hat{A}_n(w)'\hat{\Omega}_n^{1/2}}{\sqrt{n}\hat{\sigma}_{\phi,n}(w)} - \frac{A_n(w)'\Omega_n^{1/2}}{\sqrt{n}\hat{\sigma}_{\phi,n}(w)}\right\|_E \\
&\leq \left\|\frac{A_n(w)'\Omega_n^{1/2}}{\sqrt{n}\sigma_{\phi,n}(w)}\right\|_E\left|\frac{\sigma_{\phi,n}(w)}{\hat{\sigma}_{\phi,n}(w)} - 1\right| + \frac{\|A_n(w)\|_E}{\sqrt{n}\hat{\sigma}_{\phi,n}(w)}\left\|\hat{\Omega}_n^{1/2} - \Omega_n^{1/2}\right\|_E \\
&\quad + \left\|\frac{\hat{A}_n(w) - A_n(w)}{\sqrt{n}\sigma_{\phi,n}(w)}\right\|.
\end{aligned}$$

Since $\left\|\hat{\Omega}_n^{1/2} - \Omega_n^{1/2}\right\|_E \leq \left\|\hat{\Omega}_n - \Omega_n\right\|_E \left\|\Omega_n^{-1}\right\|_E^{1/2}$, we have

$$\frac{\|A_n(w)\|_E}{\sqrt{n}\hat{\sigma}_\phi(w)}\left\|\hat{\Omega}_n^{1/2} - \Omega_n^{1/2}\right\|_E \leq \frac{\|A_n(w)\|_E}{\sqrt{n}\hat{\sigma}_\phi(w)}\left\|\hat{\Omega}_n - \Omega_n\right\|_E\left\|\Omega_n^{-1}\right\|_E^{1/2} \lesssim_p b_n^{-1},$$

where the second inequality is from Theorem B.7. Similarly, as all eigenvalues of $\Omega_n$ are bounded away from

zero, by the triangle inequality,

$$\left|\frac{\hat{\sigma}_{\hat{\phi},n}(w)}{\sigma_{\phi,n}(w)}-1\right| \leq \frac{\left\|A_n(w)'\left(\hat{\Omega}_n^{1/2}-\Omega_n^{1/2}\right)\right\|_E}{\left\|A_n(w)'\Omega_n^{1/2}\right\|_E} \lesssim_p \left\|\hat{\Omega}_n^{1/2}-\Omega_n^{1/2}\right\|_E = o_p(b_n^{-1}).$$

Moreover, by Assumption A.19,

$$\left\|\frac{\hat{A}_n(w)-A_n(w)}{\sqrt{n}\sigma_{\phi,n}(w)}\right\| \lesssim \frac{\left\|\hat{A}_n(w)-A_n(w)\right\|}{\|A_n(w)\|} \lesssim \xi_{\phi,n}K_n\varrho_{s,n} = o_p(b_n^{-1})$$

Thus,

$$E_{\mathcal{N}_{k_n}}\left[\left(\hat{t}_n^*(w)-t_n^*(w)\right)^2\right]^{1/2} = o_p(b_n^{-1}).$$

Furthermore, uniformly over $w_1, w_2 \in \mathcal{W}$,

$$\begin{aligned} & E_{\mathcal{N}_{k_n}}\left[\left(\left(\hat{t}_n^*(w_1)-t_n^*(w_1)\right)-\left(\hat{t}_n^*(w_2)-t_n^*(w_2)\right)\right)^2\right]^{1/2} \\ \leq\ & \left\|\frac{A_n(w_1)'\Omega^{1/2}}{\sqrt{n}\sigma_\phi(w_1)}-\frac{A_n(w_2)'\Omega^{1/2}}{\sqrt{n}\sigma_\phi(w_2)}\right\|_E + \left\|\frac{\hat{A}_n(w_1)'\hat{\Omega}^{1/2}}{\sqrt{n}\hat{\sigma}_\phi(w_1)}-\frac{\hat{A}_n(w_2)'\Omega^{1/2}}{\sqrt{n}\hat{\sigma}_\phi(w_2)}\right\|_E. \end{aligned}$$

By triangle inequality, uniformly over $w_1, w_2 \in \mathcal{W}$,

$$\begin{aligned} \left\|\frac{A_n(w_1)'\Omega_n^{1/2}}{\sqrt{n}\sigma_\phi(w_1)}-\frac{A_n(w_2)'\Omega_n^{1/2}}{\sqrt{n}\sigma_\phi(w_2)}\right\|_E &\leq \frac{\|A_n(w_1)-A_n(w_2)\|_E}{\sqrt{n}\sigma_\phi(w_1)} + \frac{\|A_n(w)\|_E}{\sqrt{n}}\frac{|\sigma_\phi(w_2)-\sigma_\phi(w_1)|}{\sigma_\phi(w_1)\sigma_\phi(w_2)} \\ &\lesssim \frac{\|A_n(w_1)-A_n(w_2)\|_E}{\|A_n(w_1)\|_E} \lesssim_p \xi_{k_n}^L \|w_1-w_2\|_E. \end{aligned}$$

Similarly, uniformly over $w_1, w_2 \in \mathcal{W}$,

$$\left\|\frac{\hat{A}(w_1)'\hat{\Omega}^{1/2}}{\sqrt{n}\hat{\sigma}_{\hat{f}}(w_1)}-\frac{\hat{A}(w_2)'\Omega^{12}}{\sqrt{n}\hat{\sigma}_{\hat{f}}(w_2)}\right\|_E \lesssim_p \xi_{k_n}^L \|w_1-w_2\|_E.$$

It follows that

$$E_{\mathcal{N}_{k_n}}\left[\left(\left(\hat{t}_n^*(w_1)-t_n^*(w_1)\right)-\left(\hat{t}_n^*(w_2)-t_n^*(w_2)\right)\right)^2\right]^{1/2} \lesssim_p \xi_{k_n}^L \|w_1-w_2\|_E.$$

Thus, there exists a sequence $\{\ell_n\}$ s.t. $\ell_n \to 0$ and

$$P\left\{\left|\sup_{w\in\mathcal{W}}\left|\hat{t}_n^*(w)\right| - \sup_{w\in\mathcal{W}}\left|t_n^*(w)\right|\right| > \ell_n (\log n)^{-1/2}\right\} \to 0.$$

The following proof follows from the arguments in the proof of Theorem 5.6 in BCCK. We present here for completeness of the proof.

Denote $\tilde{c}_n(1-\tau)$ as the $(1-\tau)-$quantile of $\sup_{w\in\mathcal{W}}|t_n^\star(w)|$. By

$$\left|\sup_{w\in\mathcal{W}}\left|\hat{t}_n^*(w)\right| - \sup_{w\in\mathcal{W}}\left|t_n^*(w)\right|\right| = o_p\left(\ell_n (\log n)^{-1/2}\right)$$

and the fact that closeness in probability implies closeness of conditional quantiles (Lemma A.3 in BCCK, 2013), for a sequence $\{\nu_n\}$ such that $\nu_n = o(1)$, we have

$$\begin{aligned}\Pr\left\{c_n(1-\tau) < \tilde{c}_n(1-\tau-\nu_n) - \ell_n (\log n)^{-1/2}\right\} &= o(1),\\ \Pr\left\{c_n(1-\tau) > \tilde{c}_n(1-\tau+\nu_n) + \ell_n (\log n)^{-1/2}\right\} &= o(1).\end{aligned}$$

Furthermore, by the strong approximation result in Theorem 3.3, there exists a sequence of $\{\beta_n\}$ of constants and a sequence $\{Z_n\}$ of random variables such that $\beta_n = o(1)$, $Z_n$ equals in distribution to $||t_n^*||_{\mathcal{W}}$, and

$$\Pr\left\{\left|\sup_{w\in\mathcal{W}}|t_n(w)| - Z_n\right| > \beta_n/\sqrt{\log n}\right\} = o(1).$$

It implies that for universal constant $C$, we have

$$\begin{aligned}\Pr\left\{\sup_{w\in\mathcal{W}}|t_n(w)| \le c_n(1-\tau)\right\} &\le \Pr\left\{Z_n \le c_n(1-\tau) + \beta_n/\sqrt{\log n}\right\} + o(1)\\ &\le \Pr\left\{Z_n \le \tilde{c}_n(1-\tau+\nu_n + C(\ell_n+\beta_n))\right\} + o(1)\\ &\le \Pr\left\{Z_n \le \tilde{c}_n\left(1-\tau+\nu_n + C(\ell_n+\beta_n)\right)\right\} + o(1)\\ &= 1-\tau+o(1).\end{aligned}$$

where the third inequality follows from the "anti-concentration for separable Gaussian processes" (Lemma 5.3 in BCCK).

Moreover, since $c_n(1-\tau) \lesssim_p \sqrt{\log n}$, we have $2c_n(1-\tau)\hat{\sigma}_n \lesssim_p \sqrt{\log n}\sigma(w)$, uniformly over $w \in \mathcal{W}$. □

**Proof of Theorem 3.6.**

Part (i): The assumption on the linear independence of functionals implies that

$$\begin{aligned}&n\left(\boldsymbol{\phi}(\hat{\theta}_n)[w] - \boldsymbol{\phi}(\theta_0)[w]\right)' \hat{V}_{\boldsymbol{\phi},n}^{-1}\left(\boldsymbol{\phi}(\hat{\theta}_n)[w] - \boldsymbol{\phi}(\theta_0)[w]\right)\\ =\ &\sum_{j=1}^{J}\left\{\sqrt{n}\left(\phi_j(\hat{\theta}_n)[w] - \phi_j(\theta_0)[w]\right) / \left\|\hat{\delta}_{n,j}^*(w)\right\|\right\}^2\\ =\ &\sum_{j=1}^{J}\left\{\sqrt{n}\left\langle\delta_{n,j}^*(w), \hat{\theta}_n - \theta_0\right\rangle / \left\|\hat{\delta}_{n,j}^*(w)\right\|\right\}^2 + o_p(1). \qquad \text{(B.18)}\end{aligned}$$

For each $j = 1, \ldots, J$, by Theorem 3.3, for all $w \in \mathcal{W}$, we have

$$\begin{aligned}&\sqrt{n}\left\langle\delta_{n,j}^*(w), \hat{\theta}_n - \theta_0\right\rangle / \left\|\delta_{n,j}^*(w)\right\|\\ =\ &-\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{dm(X_i,\theta_0)}{d\theta}[u_{n,j}^*(w)]'\Sigma_0(X_i)^{-1}\rho(Z_i,\theta_0) + o_p(b_n^{-1})\\ =\ &V_{\phi,n}^{-1/2}A_n(w)'\Omega_n^{1/2}\mathcal{N}_{d_\beta+k_n} + o_p(b_n^{-1}). \qquad \text{(B.19)}\end{aligned}$$

Note that by Lemma B.7, for $j = 1, \ldots, J$,

$$\begin{aligned}
&\left| \frac{\left\langle \hat{\delta}^*_{n,j}(w), \hat{\theta}_n - \theta_0 \right\rangle_n}{\left\| \hat{\delta}^*_{n,j}(w) \right\|_n} - \frac{\left\langle \delta^*_{n,j}(w), \hat{\theta}_n - \theta_0 \right\rangle}{\left\| \delta^*_{n,j}(w) \right\|} \right| \\
\lesssim & \left| \frac{\left\langle \hat{\delta}^*_{n,j}(w) - \delta^*_{n,j}(w), \hat{\theta}_n - \theta_0 \right\rangle_n}{\left\| \hat{\delta}^*_{n,j}(w) \right\|_n} \right| + \frac{\left| \left\langle \hat{\delta}^*_{n,j}(w), \hat{\theta}_n - \theta_0 \right\rangle \right|}{\left\| \hat{\delta}^*_{n,j}(w) \right\|} \left| 1 - \frac{\left\| \hat{\delta}^*_{n,j}(w) \right\|_n}{\left\| \delta^*_{n,j}(w) \right\|} \right| = o_p(b_n^{-1}). \quad \text{(B.20)}
\end{aligned}$$

Combining (B.19) and (B.20) yields that

$$\left\| \frac{\left\langle \hat{\boldsymbol{\delta}}^*_n(w), \hat{\theta}_n - \theta_0 \right\rangle_n}{\left\| \hat{\boldsymbol{\delta}}^*_n(w) \right\|_n} - \sqrt{n} V_{\phi,n}^{-1/2} A_n(w)' \Omega_n^{1/2} \mathcal{N}_{d_\beta + k_n} \right\| = o_p\left(b_n^{-1}\right).$$

Therefore, for all $w \in \mathcal{W}$,

$$\begin{aligned}
\text{Wald}_n(w) &= \sum_{j=1}^{J} \left( \frac{A_{n,j}(w)' \Omega_n^{1/2}}{\left\| A_{n,j}(w)' \Omega_n^{1/2} \right\|_E} \mathcal{N}_{d_\beta + k_n} \right)^2 + o_p(1) \\
&= \mathcal{N}'_{d_\beta + k_n} \Omega_n^{1/2} A_n(w) V_{\phi,n}^{-1} A_n(w)' \Omega_n^{1/2} \mathcal{N}_{d_\beta + k_n} + o_p(1).
\end{aligned}$$

Next, we show that for a sequence $\{\ell_n\}$ such that $\ell_n \to 0$,

$$P\left\{ \left| \sup_{w \in \mathcal{W}} \left| \hat{T}^*_n(w) \right| - \sup_{w \in \mathcal{W}} |T^*_n(w)| \right| > \ell_n / \sqrt{\log n} \right\} = o(1).$$

Note that by triangle inequality and Assumption 3.8,

$$\begin{aligned}
&\left| \sup_{w \in \mathcal{W}} \left| \hat{T}^*_n(w) \right| - \sup_{w \in \mathcal{W}} |T^*_n(w)| \right| \\
\leq & \sup_{w \in \mathcal{W}} \left| \hat{T}^*_n(w) - T^*_n(w) \right| \\
= & \sup_{w \in \mathcal{W}} \left| \left\| \hat{V}_{\phi,n}^{-1/2} \hat{A}_n(w)' \hat{\Omega}_n^{1/2} \mathcal{N}_{d_\beta + k_n} \right\|_E^2 - \left\| V_{\phi,n}^{-1/2} A_n(w)' \Omega_n^{1/2} \mathcal{N}_{d_\beta + k_n} \right\|_E^2 \right| \\
\leq & \sup_{w \in \mathcal{W}} \sum_{j=1}^{J} \left| \left( \frac{\hat{A}_{n,j}(w)' \hat{\Omega}_n^{1/2}}{\left\| \hat{A}_{n,j}(w)' \hat{\Omega}_n^{1/2} \right\|_E} - \frac{A_{n,j}(w)' \Omega_n^{1/2}}{\left\| A_{n,j}(w)' \Omega_n^{1/2} \right\|_E} \right) \mathcal{N}_{d_\beta + k_n} \right|^2 + o_p(b_n^{-1}).
\end{aligned}$$

Then the results follows by similar calculations in the proof of Theorem 3.5 and extended continuous mapping theorem.

Part (ii): By the definition of $\sup - \text{QLR}_n$, we first consider the difference of $L_n(\tilde{\theta}_n) - L_n(\hat{\theta}_n)$. Since $\boldsymbol{\phi}(\tilde{\theta}_n)[w] = \boldsymbol{\phi}(\theta_0)[w]$ for all $w \in \mathcal{W}$, under the null, $\left\langle \mathbf{u}^*_n(w), \tilde{\theta}_n - \theta_0 \right\rangle = o_p(n^{-1/2})$.

Let $\tilde{\theta}_n(t_n, w) = \tilde{\theta}_n - t_n \mathbf{u}^*_n(w)$. By the definition of $\hat{\theta}_n$,

$$\hat{L}_n(\tilde{\theta}_n) - \hat{L}_n(\hat{\theta}_n) \geq \hat{L}_n\left(\tilde{\theta}_n\right) - \hat{L}_n\left(\tilde{\theta}_n(t_n, w)\right) - o_p(n^{-1}).$$

Furthermore,

$$\begin{aligned}
& \hat{L}_n\left(\tilde{\theta}_n\right) - \hat{L}_n\left(\tilde{\theta}_n(t_n, w)\right) - o_p(n^{-1}) \\
= \; & \hat{L}_n\left(\tilde{\theta}_n\right) - \hat{L}_n\left(\tilde{\theta}_n - t_n \mathbf{u}_n^*(w)\right) - o_p(n^{-1}) \\
= \; & \frac{2t_n}{n} \sum_{i=1}^{n} \frac{dm(X_i, \theta_0)}{d\theta} \left[\mathbf{u}_n^*(w)\right]' \Sigma_0(X_i)^{-1} \rho(Z_i, \theta_0) + 2t_n \left\langle \mathbf{u}_n^*(w), \tilde{\theta}_n - \theta_0 \right\rangle - t_n^2 + o_p(n^{-1}). \\
= \; & \frac{2t_n}{n} \sum_{i=1}^{n} \frac{dm(X_i, \theta_0)}{d\theta} \left[\mathbf{u}_n^*(w)\right]' \Sigma_0(X_i)^{-1} \rho(Z_i, \theta_0) - t_n^2 + O_p(n^{-1/2} t_n) + o_p(n^{-1}) \\
= \; & \frac{2t_n}{n} \sum_{i=1}^{n} \frac{dm(X_i, \theta_0)}{d\theta} \left[\mathbf{u}_n^*(w)\right]' \Sigma_0(X_i)^{-1} \rho(Z_i, \theta_0) - t_n^2 + o_p(n^{-1}).
\end{aligned} \tag{B.21}$$

Minimizing the above distance yields that

$$t_n = \frac{1}{n} \sum_{i=1}^{n} \frac{dm(X_i, \theta_0)}{d\theta} \left[\mathbf{u}_n^*(w)\right]' \Sigma_0(X_i)^{-1} \rho(Z_i, \theta_0) = O_p(n^{-1/2}) \tag{B.22}$$

and hence for all $w$, we have

$$\hat{L}_n\left(\tilde{\theta}_n\right) - \hat{L}_n\left(\hat{\theta}_n\right) \geq \left( \frac{1}{n} \sum_{i=1}^{n} \frac{dm(X_i, \theta_0)}{d\theta} \left[\mathbf{u}_n^*(w)\right]' \Sigma_0(X_i)^{-1} \rho(Z_i, \theta_0) \right)^2 + o_p(n^{-1}). \tag{B.23}$$

On the other hand, let $\theta_n^* = \hat{\theta}_n(t_n, w) = \hat{\theta}_n + t_n \mathbf{u}_n^*(w)$, by similar arguments in the proof of Theorem 3.3, for all $w \in \mathcal{W}$,

$$\begin{aligned}
& \hat{L}_n\left(\hat{\theta}_n(t_n, w)\right) - \hat{L}_n\left(\hat{\theta}_n\right) \\
= \; & \frac{2t_n}{n} \sum_{i=1}^{n} \frac{dm(X_i, \theta_0)}{d\theta} \left[\mathbf{u}_n^*(w)\right]' \Sigma_0(X_i)^{-1} \rho(Z_i, \theta_0) + 2t_n \left\langle \mathbf{u}_n^*(w), \hat{\theta}_n - \theta_0 \right\rangle + t_n^2 + o_p(n^{-1}) \\
= \; & \left( \frac{1}{n} \sum_{i=1}^{n} \frac{dm(X_i, \theta_0)}{d\theta} \left[\mathbf{u}_n^*(w)\right]' \Sigma_0(X_i)^{-1} \rho(Z_i, \theta_0) \right)^2 + o_p(n^{-1}),
\end{aligned}$$

where the last equality follows from (B.22), $t_n \lesssim n^{-1/2}$ and the result in Theorem 3.3 such that for all $w \in \mathcal{W}$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{dm(X_i, \theta_0)}{d\theta} \left[\mathbf{u}_n^*(w)\right]' \Sigma_0(X_i)^{-1} \rho(Z_i, \theta_0) + \left\langle \mathbf{u}_n^*(w), \hat{\theta}_n - \theta_0 \right\rangle = o_p(1).$$

If $\boldsymbol{\phi}\left(\hat{\theta}_n(t_n, w)\right)[w] = \boldsymbol{\phi}(\theta_0)[w]$ for all $w \in \mathcal{W}$, then by the definition of $\tilde{\theta}_n$,

$$\begin{aligned}
n\left(\hat{L}_n(\tilde{\theta}_n) - \hat{L}_n\left(\hat{\theta}_n\right)\right) \; & \leq \; n\left(\hat{L}_n\left(\hat{\theta}_n(t_n, w)\right) - \hat{L}_n(\hat{\theta}_n)\right) + o_p(1) \\
& = \; \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{dm(X_i, \theta_0)}{d\theta} \left[\mathbf{u}_n^*(w)\right]' \Sigma_0(X_i)^{-1} \rho(Z_i, \theta_0) \right)^2 + o_p(1).
\end{aligned} \tag{B.24}$$

Combining (B.23) and (B.24), we conclude that uniformly over $w \in \mathcal{W}$,

$$n\left(\hat{L}_n(\tilde{\theta}_n) - \hat{L}_n\left(\hat{\theta}_n\right)\right) = \left(\frac{1}{n}\sum_{i=1}^{n} \frac{dm(X_i,\theta_0)}{d\theta}\left[\mathbf{u}_n^*(w)\right]' \Sigma_0(X_i)^{-1}\rho(Z_i,\theta_0)\right)^2 + o_p(1).$$

However, if $\boldsymbol{\phi}\left(\hat{\theta}_n(t_n,w)\right)[w] \neq \boldsymbol{\phi}(\theta_0)[w]$ for all $w \in \mathcal{W}$, under the null, we show that there exists $t_n^* \in \mathcal{T}_n$ w.p.a.1. such that (i) $t_n^* = t_n + o_p(n^{-1/2}) = O_p(n^{-1/2})$ and (ii) $\boldsymbol{\phi}\left(\hat{\theta}_n(t_n^*,w)\right)[w] = \boldsymbol{\phi}(\theta_0)[w]$ for all $w \in \mathcal{W}$. The proof strategy is based on the one in Shen and Shi (2005) or Chen and Pouzo (2014).

Let $\bar{M}_n = o\left(||\boldsymbol{\delta}_n^*||n^{-1/2}\right)$. By the Riesz representation theorem, for all $w \in \mathcal{W}$,

$$-\bar{M}_n \leq \boldsymbol{\phi}(\theta + t_n u_n^*(w))[w] - \boldsymbol{\phi}(\theta_0)[w] - \langle \theta - \theta_0, \delta_n^*(w)\rangle - t_n||\boldsymbol{\delta}_n^*(w)|| \leq \bar{M}_n,$$

Thus, for any $r \in \left\{|r| \leq 2K_n \max_{1\leq j\leq J}||\delta_{n,j}^*(w)||\varrho_n\right\}$, let

$$\underline{t} = -\langle \mathbf{u}_n^*(w), \theta - \theta_0\rangle - 2\bar{M}_n \left\|\boldsymbol{\delta}_n^*(w)\right\|^{-1} + r\left\|\boldsymbol{\delta}_n^*(w)\right\|^{-1}$$

and

$$\bar{t} = -\langle \mathbf{u}_n^*(w), \theta - \theta_0\rangle + 2\bar{M}_n \left\|\boldsymbol{\delta}_n^*(w)\right\|^{-1} + r\left\|\boldsymbol{\delta}_n^*(w)\right\|^{-1}.$$

To show that $\underline{t} \in \mathcal{T}_n$, note that since

$$|\underline{t}| \times ||\boldsymbol{\delta}_n^*(w)|| \lesssim ||\theta - \theta_0|| + 2\left|\bar{M}_n\right| + |r(w)|,$$

we have $|\underline{t}| \leq 4K_n^2\varrho_n$. Then $\underline{t} \in \mathcal{T}_n$. By the same argument, $\bar{t} \in \mathcal{T}_n$. Therefore, by the definitions of $\underline{t}$ and $\bar{t}$, for all $w \in \mathcal{W}$,

$$\boldsymbol{\phi}(\theta + \bar{t}u_n^*(w))[w] - \boldsymbol{\phi}(\theta_0)[w] \quad \geq \quad r(w) + \bar{M}_n > r(w)$$

and

$$\boldsymbol{\phi}(\theta + \underline{t}u_n^*(w))[w] - \boldsymbol{\phi}(\theta_0)[w] \quad \leq \quad r(w) - \bar{M}_n < r(w).$$

By continuity of $\phi(\theta + tu_n^*)$ and the mean value theorem, there exists some $t_n^* \in [\underline{t}, \bar{t}]$ such that $t_n^* \in \mathcal{T}_n$ and $\boldsymbol{\phi}(\theta + t_n^*\mathbf{u}_n^*(w))[w] = \mathbf{r}(w)$. It implies that $\boldsymbol{\phi}\left(\hat{\theta}_n(t_n^*,w)\right)[w] - \boldsymbol{\phi}(\theta_0)[w] = 0$. Furthermore, by similar arguments as above, we show that

$$\begin{aligned}
& L_n\left(\hat{\theta}_n(t_n^*,w)\right) - L_n\left(\hat{\theta}_n(t,w)\right) - o_p(n^{-1}) \\
= \quad & \left(L_n\left(\hat{\theta}_n(t_n^*,w)\right) - L_n(\hat{\theta}_n)\right) - \left(L_n\left(\hat{\theta}_n(t,w)\right) - L_n\left(\hat{\theta}_n\right)\right) - o_p(n^{-1}) \\
= \quad & t_n^* \sum_{j=1}^{J}\left[\left\langle \hat{\theta}_n - \theta_0, u_{n,j}^*(w)\right\rangle - \frac{1}{n}\sum_{i=1}^{n}\frac{dm(X_i,\theta_0)}{d\theta}\left[u_{n,j}^*(w)\right]'\Sigma(X_i)^{-1}\rho(Z_i,\theta_0)\right] + o_p(n^{-1}) \\
= \quad & O_p(n^{-1/2}) \times o_p(n^{-1/2}) + o_p(n^{-1}) = o_p(n^{-1}).
\end{aligned}$$

Therefore, by orthogonality of $\{\delta^*_{n,j}(w)\}$ for all $w$,

$$\begin{aligned} n\left(\hat{L}_n(\tilde{\theta}_n) - \hat{L}_n\left(\hat{\theta}_n\right)\right) &\leq n\left(\hat{L}_n\left(\hat{\theta}_n(t^*_n, w)\right) - L_n\left(\hat{\theta}_n\right)\right) \\ &= \sum_{j=1}^{J}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{dm(X_i,\theta_0)}{d\theta}\left[\mathbf{u}^*_n(w)\right]'\Sigma_0(X_i)^{-1}\rho(Z_i,\theta_0)\right)^2 + o_p(1) \\ &= \sum_{j=1}^{J}\left(\sqrt{n}\left\langle\delta^*_{j,n}(w), \hat{\theta}_n - \theta_0\right\rangle / \left\|\delta^*_{j,n}(w)\right\|\right)^2 + o_p(1) \end{aligned}$$

and

$$\begin{aligned} n\left(\hat{L}_n(\tilde{\theta}_n) - \hat{L}_n\left(\hat{\theta}_n\right)\right) &\geq n\hat{L}_n\left(\tilde{\theta}_n\right) - \hat{L}_n\left(\tilde{\theta}_n(t^*_n, w)\right) \\ &= \sum_{j=1}^{J}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{dm(X_i,\theta_0)}{d\theta}\left[\mathbf{u}^*_n(w)\right]'\Sigma_0(X_i)^{-1}\rho(Z_i,\theta_0)\right)^2 + o_p(1) \\ &= \sum_{j=1}^{J}\left(\sqrt{n}\left\langle\delta^*_{j,n}(w), \hat{\theta}_n - \theta_0\right\rangle // \left\|\delta^*_{j,n}(w)\right\|\right)^2 + o_p(1) \end{aligned}$$

for all $w \in \mathcal{W}$. Since for each $j = 1, \ldots, J$, for all $w \in \mathcal{W}$, by Theorem 3.3,

$$\frac{\sqrt{n}\left\langle\delta^*_{j,n}(w), \hat{\theta}_n - \theta_0\right\rangle}{\left\|\delta^*_{j,n}(w)\right\|} = \frac{A_{n,j}(w)'\Omega_n^{1/2}}{\left\|A_{n,j}(w)'\Omega_n^{1/2}\right\|}\mathcal{N}_{d_\beta+k_n} + o_p(1).$$

It yields that

$$n\left(\hat{L}_n\left(\tilde{\theta}_n\right) - \hat{L}_n\left(\hat{\theta}_n\right)\right) = \mathcal{N}'_{d_\beta+k_n}\Omega_n^{1/2}A_n(w)V^{-1}_{\phi,n}A_n(w)'\Omega_n^{1/2}\mathcal{N}_{d_\beta+k_n} + o_p(1).$$

(iii) By similar argument as in the proof of Theorem 3.3, let $\tilde{\mathbf{u}}^*_n(w) = \tilde{\boldsymbol{\delta}}^*_n(w)/||\tilde{\boldsymbol{\delta}}^*_n(w)||$ for $w \in \mathcal{W}$. Then for all $w \in \mathcal{W}$,

$$\begin{aligned} &\frac{1}{2}\frac{d\tilde{L}_n(\tilde{\theta}_n)}{d\theta}\left[\tilde{\mathbf{u}}^*_n(w)'\right] \\ = \ &\frac{1}{n}\sum_{i=1}^{n}\frac{dm(X_i,\theta_0)}{d\theta}\left[\tilde{\boldsymbol{\delta}}^*_n(w)/\left\|\tilde{\boldsymbol{\delta}}^*_n(w)\right\|\right]'\Sigma_0(X_i)^{-1}\rho(Z_i,\theta_0) + \left\langle\tilde{\theta}_n - \theta_0, \tilde{\mathbf{u}}^*_n(w)\right\rangle + o_p(n^{-1/2}). \end{aligned}$$

By the definition of $\tilde{\theta}_n$, for all $w$, we have

$$0 = \sqrt{n}\left(\boldsymbol{\phi}\left(\tilde{\theta}_n\right)[w] - \boldsymbol{\phi}(\theta_0)[w]\right) / \left\|\tilde{\boldsymbol{\delta}}^*_n(w)\right\| = \sqrt{n}\left\langle\tilde{\theta}_n - \theta_0, \tilde{\boldsymbol{u}}^*_n(w)\right\rangle + o_p(1).$$

It implies that $\sqrt{n}\left\langle \tilde{\theta}_n - \theta_0, \tilde{\mathbf{u}}_n^*(w) \right\rangle = o_p(1)$. Thus,

$$\begin{aligned}\sqrt{n}\frac{1}{2}\frac{d\tilde{L}_n(\tilde{\theta}_n)}{d\theta}\left[\tilde{\mathbf{u}}_n^*(w)'\right] &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{dm(X_i,\theta_0)}{d\theta}\left[\tilde{\boldsymbol{\delta}}_n^*(w)/\left\|\tilde{\boldsymbol{\delta}}_n^*(w)\right\|\right]'\Sigma_0(X_i)^{-1}\rho(Z_i,\theta_0) + o_p(1) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{dm(X_i,\theta_0)}{d\theta}\left[\boldsymbol{\delta}_n^*(w)/\left\|\boldsymbol{\delta}_n^*(w)\right\|\right]'\Sigma_0(X_i)^{-1}\rho(Z_i,\theta_0) + o_p(1),\end{aligned}$$

where the second equality follows from Lemma B.7. Then the results follow by the same steps above as the ones for sup-Wald$_n$ and sup $-$QLR$_n$. □

### B.1.4 Multiplier Bootstrap for Uniform Inference Under Point Identification

we show that the following multiplier bootstrap procedures are valid for uniform inference under point identification.

Consider a sequence $\{\zeta_i\}_{i=1}^n$ that are i.i.d. draws from the standard exponential distribution and independent of the data. For each draw of such weights, we define the multiplier bootstrap draw of the SGMM estimator $\hat{\theta}_n^\star$ as a solution to the following criterion weighted by $\{\zeta_i\}_{i=1}^n$ such that

$$\hat{L}_n^\star(\theta) = \left(\frac{1}{n}\sum_{i=1}^n \zeta_i\rho(Z_i,\theta)\otimes q_i^{s_n}\right)'\left(\frac{1}{n}\sum_{i=1}^n \Sigma(X_i,\tilde{\theta}_n)\otimes q_i^{s_n}q_i^{s_n'}\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n \zeta_i\rho(Z_i,\theta)\otimes q_i^{s_n}\right) + \lambda_n \text{Pen}(h). \tag{B.25}$$

Let $g_i^\star(\hat{\theta}_n) = (\zeta_i - 1)\rho(Z_i,\hat{\theta}_n^\star)$ and the bootstrap variance estimator be

$$\begin{aligned}\hat{V}_{\phi,n}^\star &= \frac{d\phi(\hat{\theta}_n)}{d\theta}\left[\bar{\psi}^{k_n}(\cdot)'\right]'\times\left(\frac{1}{n}\sum_{i=1}^n \frac{d\rho(Z_i,\hat{\theta}_n)}{d\theta}\left[\bar{p}^{k_n}(\cdot)'\right]\otimes q_i\right)'\left(\frac{1}{n}\sum_{i=1} g_i^\star\left(\hat{\theta}_n\right)g_i^\star\left(\hat{\theta}_n\right)'\right)^{-1} \\ &\quad\times\left(\frac{1}{n}\sum_{i=1}^n \frac{d\rho(Z_i,\hat{\theta}_n)}{d\theta}\left[\bar{p}^{k_n}(\cdot)'\right]\otimes q_i\right)\times\frac{d\phi(\hat{\theta}_n)}{d\theta}\left[\bar{\psi}^{k_n}(\cdot)'\right].\end{aligned}$$

We have $E\left[\hat{V}_{\phi,n}^\star|Z^n\right] = \hat{V}_{\phi,n}$.

Let

$$\sup-\text{Wald}_n^\star \equiv \sup_{w\in\mathcal{W}}\left\{n\left(\boldsymbol{\phi}\left(\hat{\theta}_n^\star\right)[w] - \boldsymbol{\phi}\left(\hat{\theta}_n\right)[w]\right)'\hat{V}_{\hat{\boldsymbol{\phi}}_n}^{\star-1}(w)\left(\boldsymbol{\phi}\left(\hat{\theta}_n^\star\right)[w] - \boldsymbol{\phi}\left(\hat{\theta}_n\right)[w]\right)\right\}, \tag{B.26}$$

$$\sup-\text{QLR}_n^\star \equiv n\sup_{w\in\mathcal{W}}\left\{\min_{\theta\in\Theta_n:\phi(\theta)=\hat{r}(w)} L_n^*(\theta) - \min_{\theta\in\Theta_n} L_n^\star(\theta)\right\},$$

where $\hat{r}(w) = \phi(\hat{\theta}_n)[w]$ and

$$\begin{aligned}\sup-\text{LM}_n^\star &\equiv \frac{n}{4}\sup_{w\in\mathcal{W}}\left\{\left(\frac{d\hat{L}_n\left(\tilde{\theta}_n\right)}{d\theta}\left[\bar{p}^{k_n}(\cdot)'\right]\right)[w]'\left(\frac{\partial\boldsymbol{\phi}(\tilde{\theta}_n)}{\partial\theta}\left[\bar{p}^{k_n}(\cdot)\right]\right)[w]\right. \\ &\quad\left.\times\tilde{V}_{\boldsymbol{\phi},n}^{-1}(w)\left(\frac{\partial\boldsymbol{\phi}(\tilde{\theta}_n)}{\partial\theta}\left[\bar{p}^{k_n}(\cdot)\right]\right)'[w]\left(\frac{d\hat{L}_n\left(\tilde{\theta}_n\right)}{d\theta}\left[\bar{p}^{k_n}(\cdot)'\right]\right)[w]\right\}.\end{aligned}$$

**Theorem B.2.** *Suppose that Assumptions of Theorem 3.6 hold. Then for*

$$T_n^*(w) = \mathcal{N}'_{d_\beta+k_n}\Omega_n^{1/2}\mathbf{A}_n(w)V_{\boldsymbol{\phi}}^{-1}\mathbf{A}_n(w)'\Omega_n^{1/2}\mathcal{N}_{d_\beta+k_n}, \tag{B.27}$$

*then we have*

*(i)* $\sup-Wald_n^\star \overset{d}{=} \sup_w\{T_n^*(w)\} + o_p(1)$; $\sup-QLR_n^\star \overset{d}{=} \sup_w\{T_n^*(w)\} + o_p(1)$ *and* $\sup-LM_n^\star \overset{d}{=} \sup_w\{T_n^*(w)\} + o_p(1)$.

*(ii) The results continue to hold in $P-$probability if we replace the unconditional $P$ by the conditional probability $P^\star(\cdot|\mathcal{D})$ given $\mathcal{D} = \{(x_i, y_i) : i = 1, \ldots, n\}$.*

*Proof.* The proof of Theorem is similar to the one of Theorem 3.6. Thus, we only present the main steps for the $\sup-\text{Wald}_n^\star$. For the $\sup-\text{Wald}_n^\star$ statistic, note that the weight $\zeta_i$ is independent of data with $E[\zeta_i] = 1$, $E[\zeta_i^2] = 1$ and $\max_{1\le i\le n}\zeta_i \lesssim_p \log n$. By replacing the envelop of $\xi_{\rho,k_n}$ we used in the proof of Theorem 3.3 by $\xi_{\rho,k_n}\log n$, we have

$$\begin{aligned}
& n\left(\boldsymbol{\phi}(\hat\theta_n^\star)[w] - \boldsymbol{\phi}(\hat\theta_n)[w]\right)' \hat V_{\boldsymbol{\phi},n}^{-1}\left(\boldsymbol{\phi}(\hat\theta_n^\star)[w] - \boldsymbol{\phi}(\hat\theta_n)[w]\right) \\
= & \sum_{j=1}^J \left\{\frac{\sqrt{n}\left(\phi_j(\hat\theta_n^\star)[w] - \phi_j(\hat\theta_n)[w]\right)}{\left\|\hat\delta_{n,j}^*(w)\right\|}\right\}^2 \\
= & \sum_{j=1}^J \left\{\sqrt{n}\left\langle \delta_{n,j}^*(w), \hat\theta_n^\star - \hat\theta_n\right\rangle / \left\|\delta_{n,j}^*(w)\right\|\right\}^2 + o_p(b_n^{-1}),
\end{aligned} \tag{B.28}$$

where the second equality follows from Lemma B.7.

For $j = 1, \ldots, J$, it can be shown that uniformly over $w \in \mathcal{W}$.

$$\begin{aligned}
\sqrt{n}\left\langle u_{n,j}^*, \hat\theta_n^\star - \hat\theta_n\right\rangle &= \sqrt{n}\left\langle u_{n,j}^*(w), \hat\theta_n^\star - \theta_0\right\rangle - \sqrt{n}\left\langle u_{n,j}^*(w), \hat\theta_n - \theta_0\right\rangle \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n (\zeta_i - 1)\frac{dm(X_i,\theta_0)}{d\theta}\left[u_{n,j}^*(w)\right]'\Sigma_0(X_i)^{-1}\rho(Z_i,\theta_0) + o_p(1) \\
&= \frac{A_{n,j}(w)'\Omega_n^{1/2}}{\sqrt{n}||\Omega_n A_{n,j}(w)||}\Omega_n^{1/2}\sum_{i=1}^n (\zeta_i - 1)\frac{dm(X_i,\theta_0)}{d\theta}\left[\bar p^{k_n}(\cdot)'\right]\Sigma_0(X_i)^{-1}\rho(Z_i,\theta_0) + o_p(1)
\end{aligned}$$

where the last equality follows from $\delta_{n,j}^*(w) = \bar p^{k_n}(\cdot)'\Omega_n A_{n,j}(w)$ and

$$u_{n,j}^*(w) = \bar p^{k_n}(\cdot)'\Omega_n A_{n,j}(w)/\left\|\Omega_n A_{n,j}(w)\right\|.$$

Let $\zeta_i^o = \zeta_i - 1$.Note that for $\mathcal{F}$ defined in (B.15), $\left\|\zeta(f^1 - f^2)\right\|_{Q,2} \le \|\zeta\|_{Q,2}\left\|f^1 - f^2\right\|_{Q,2}$, then

$$N\left(\varepsilon||\zeta||_{Q,2}||F||_{Q,2}, \zeta\mathcal{F}, L_2(Q)\right) \le N\left(\varepsilon||\zeta||_{Q,2}, \mathcal{F}, L_2(Q)\right).$$

Let $\{\tau_i\}_{i=1}^n$ be a sequence of i.i.d. Rademacher random variables defined by $\Pr(\tau = 1) = \Pr(\tau = -1) = \frac{1}{2}$. Let

$$\mathbb{G}_n^o \equiv \frac{\alpha_n(w)'}{\sqrt{n}}\Omega_n^{1/2}\sum_{i=1}^n \tau_i\frac{dm(X_i,\theta_0)}{d\theta}[\bar p^{k_n}(\cdot)']\rho(Z_i,\theta_0).$$

By Theorem 2.2.4 of van der Vaart and Wellner (1996), for $\psi_2(y_2) = \exp\{y^2\} - 1$. For the Orlicz norm

$||\cdot||_{\psi_2}$,

$$\begin{aligned}\left\|\left\|\mathbb{G}_n^0\right\|_{\zeta F}\right\|_{\psi_2|D,\zeta} &\lesssim \int_0^{\left\|\|\zeta f\|_n\right\|_{\mathcal{F}}} \sqrt{1+\log N\left(\varepsilon,\zeta\mathcal{F},L_2(\mathbb{P}_n)\right)}d\varepsilon \\ &\simeq \int_0^{\frac{\left\|\|\zeta f\|_n\right\|_{\mathcal{F}}}{\|\zeta\|_n\times||F||_n}} \sqrt{1+\log N\left(u||\zeta||_n\times||F||_n,\zeta\mathcal{F},L_2\left(\mathbb{P}_n\right)\right)}||\zeta||_n\times||F||_n du \\ &\lesssim \int_0^1 \sqrt{1+\log N\left(u||F||_n,\mathcal{F},L_2\left(\mathbb{P}_n\right)\right)}||\zeta||_n\times||F||_n du.\end{aligned}$$

Next, we take expectations on both sides of the above equation. By the definition of $\zeta$ ($E[\zeta]=1$) and the symmetrization inequality (Lemma 2.3.1 of van der Vaart and Wellner, 1996), we have

$$E\left[\sup_{w\in\mathcal{W}}\alpha_n(w)'\Omega_n\mathbb{G}_n\left[\zeta\frac{dm(X_i,\theta_0)}{d\theta}[\bar{p}^{k_n}(\cdot)']\rho(Z_i,\theta_0)\right]\right] \lesssim \sqrt{\log n}$$

by combining the results from (B.16).

Furthermore, since $E\left[\left(\zeta^o\right)^2\right]=1$ and $E\left[|\zeta^o|^3\right]\lesssim 1$, we can apply Yurinskii's coupling for the weighted process and obtain that

$$\Pr\left\{\left\|\frac{1}{\sqrt{n}}\Omega_n^{1/2}\sum_{i=1}^n\zeta_i^o\frac{dm(X_i,\theta_0)}{d\theta}\left[\bar{p}^{k_n}(\cdot)'\right]'\Sigma_o(X_i)^{-1}\rho(Z_i,\theta_0)-\mathcal{N}_{d_\beta+k_n}\right\|>3\delta b_n^{-1}\right\} \to 0.$$

It implies that

$$\begin{aligned}\sqrt{n}\left\langle u_n^*(w),\hat{\theta}_n^\star-\hat{\theta}_n\right\rangle &= \frac{A_n(w)}{\left\|\Omega_n^{1/2}A_n(w)\right\|_E}'\Omega_n\mathbb{G}_n\left[\zeta_i^o\frac{dm(X_i,\theta_0)}{d\theta}\left[\bar{p}^{k_n}(\cdot)'\right]'\Sigma_o(X_i)^{-1}\rho(Z_i,\theta_0)\right]+o_p\left(b_n^{-1}\right) \\ &= \frac{A_n(w)'\Omega_n^{1/2}}{\left\|\Omega_n^{1/2}A_n(w)\right\|_E}\mathcal{N}_{d_\beta+k_n}+o_p\left(b_n^{-1}\right).\end{aligned}$$

Part (ii) follows from Theorem 4.5 (3) in BCCK.

□

## B.2 Proofs of Sections 4

Throughout this subsection, our results are based on a general SGMM criterion function such that

$$\bar{L}_n(\theta_n)=\hat{g}(\theta_n)'\hat{W}\hat{g}(\theta_n)+\lambda_n\text{Pen}(h_n).$$

We assume Assumption A.20 holds.

**Proof of Lemma 4.1.**

Lemma 4.1 is a direct application of Lemma C.4 in the Supplemental Appendix.□

**Proof of Lemma 4.2.**

Without loss of generality, we assume $d_\rho=1$ and $J=1$. Under Assumption 4.3 and Assumption 4.4, for

any $\theta_0 \in \Theta_0$, consider the mapping $\mathbf{F} : \mathbf{F}(\rho(\cdot,\theta))[w] = \phi(\theta)[w]$. Notice that $\frac{d\mathbf{F}(\rho_0)}{d\rho}[\rho - \rho_0] : \bar{\mathbf{V}} \to R$ is a linear functional of $\rho$, then if

$$\sup_{\nu(Z,\theta)=\rho(Z,\theta)-\rho(Z,\theta_0)\neq 0:\nu\in\bar{\mathbf{V}}} \frac{\left|\left(\frac{d\mathbf{F}(\rho(\cdot,\theta_0))}{d\rho}[\nu]\right)[w]\right|^2}{E\left[E\left[\nu(Z,\theta)|X\right]'\Sigma(X)^{-1}E\left[\nu(Z,\theta)|X\right]\right]} < \infty, \tag{B.29}$$

then functional $\boldsymbol{\phi}(\theta)$ are functional on $\bar{\mathbf{V}}$, by the Riesz representation theorem, for any $\theta_0 \in \Theta_0$, there is a $\nu^*(\cdot,\theta_0) \in \bar{\mathbf{V}}$ such that

$$\|\nu^*(\cdot,\theta_0,w)\|_{wp}^2 = \sup_{\nu\in\bar{\mathbf{V}}:\langle\nu,\nu\rangle\neq 0} \frac{\left|\left(\frac{d\mathbf{F}(\rho(\cdot,\theta_0))}{d\rho}[\nu]\right)[w]\right|^2}{E\left[E\left[\nu(Z,\theta)|X\right]'\Sigma(X)^{-1}E\left[\nu(Z,\theta)|X\right]\right]}, \tag{B.30}$$

and

$$\begin{aligned}\left(\frac{d\mathbf{F}(\rho_0)}{d\rho}[\rho-\rho_0]\right)[w] &= \langle\nu^*(\cdot,\theta_0,w),\rho-\rho_0\rangle_{wp} \\ &= \left[E\left[\nu^*(Z,\theta_0,w)|X\right]'\Sigma(X)^{-1}E\left[\rho(Z,\theta)-\rho(Z,\theta_0)\right]\right].\end{aligned}$$

Then

$$\begin{aligned}\left(\frac{d\phi(\theta_0)}{d\theta}[\theta-\theta_0]\right)[w] &= \left(\frac{d\mathbf{F}(\rho_0)}{d\rho}\cdot\frac{d\rho(Z,\theta_0)}{d\theta}[\theta-\theta_0]\right)[w] = \left\langle\nu^*(\cdot,\theta_0,w),\frac{d\rho(Z,\theta_0)}{d\theta}[\theta-\theta_0]\right\rangle_{wp} \\ &= E\left[E\left[\nu^*(Z,\theta_0,w)|X\right]'\Sigma(X)^{-1}E\left[\frac{d\rho(Z,\theta_0)}{d\theta}[\theta-\theta_0]\right]\right].\end{aligned}$$

If

$$\sup_{\nu(Z,\theta)=\rho(Z,\theta)-\rho(Z,\theta_0)\neq 0:\nu\in\bar{\mathbf{V}}} \frac{\left|\left(\frac{dF(\rho(\cdot,\theta_0))}{\partial\theta}[\nu]\right)[w]\right|^2}{E\left[E\left[\nu(Z,\theta)|X\right]'\Sigma(X)^{-1}E\left[\nu(Z,\theta)|X\right]\right]} = \infty,$$

we still have

$$\sup_{\nu_n(Z,\theta)=\rho(Z,\theta)-\rho(Z,\theta_{0n})\neq 0:\nu\in\bar{\mathbf{V}}_n} \frac{\left|\left(\frac{d\mathbf{F}(\rho(\cdot,\theta_{0n}))}{d\rho}[\nu]\right)[w]\right|^2}{E\left[E\left[\nu(Z,\theta)|X\right]'\Sigma(X)^{-1}E\left[\nu(Z,\theta)|X\right]\right]} < \infty,$$

then there exists a $\nu_n^*(Z,\theta_0)$ such that

$$\left(\frac{d\mathbf{F}(\rho_0)}{d\rho}[\rho-\rho_{0n}]\right)[w] = E\left[E\left[\nu_n^*(Z,\theta_0,w)|X\right]'\Sigma(X)^{-1}E\left[\rho(Z,\theta)-\rho(Z,\theta_{0n})|X\right]\right] \tag{B.31}$$

□

**Proof of Theorem 4.1.**

By Assumption 4.4, without loss of generality, we assume $J = 1$. Let $\nu_n^*(\bar{\theta}_0,w)$ be the empirical Riesz representer and $\mu_n^*\left(\bar{\theta}_0,w\right) = \frac{\nu_n^*\left(\bar{\theta}_0,w\right)}{\left\|\nu_n^*(\bar{\theta}_0,w)\right\|_{sd}}$. For all $\bar{\theta}_0 \in \Theta_0 \cap \mathbb{R}$, we consider $\mathcal{B}_n(\bar{\theta}_0)$. By the definition of $\hat{\theta}_n$ (the unconstrained estimator), we have

$$\bar{L}_n\left(\tilde{\theta}_n\right) - \bar{L}_n\left(\hat{\theta}_n\right) \geq \bar{L}_n\left(\tilde{\theta}_n\right) - \bar{L}_n\left(\tilde{\theta}_n\left(t_n, w\right)\right) \tag{B.32}$$

where $\tilde{\theta}_n(t_n, w) \in \mathcal{B}_n(\bar{\theta}_0)$ satisfies $\rho(Z, \tilde{\theta}_n(t_n, w)) = \rho(Z, \tilde{\theta}_n) - t_n \times \mu_n^*(\bar{\theta}_0, w)$ for $\left\{t_n \in \mathcal{T}_n \equiv \left\{t \in [-1, 1] : t \lesssim n^{-1/2}\right\}\right\}$. To simplify notation, we write $\rho(Z, \theta) = \rho(\theta)$ for $\theta \in \mathcal{B}_n(\Theta_0) \cup \Theta_0$. By Lemma C.6,

$$\begin{aligned}
&\bar{L}_n\left(\tilde{\theta}_n\right) - \bar{L}_n\left(\tilde{\theta}_n\left(t_n, w\right)\right) \\
= \;& 2t_n\left(\frac{1}{n}\sum_{i=1}^{n} E\left[\mu_n^*(Z_i\bar{\theta}_0, w)|X_i\right]' \Sigma(X_i)^{-1}\rho(Z_i, \bar{\theta}_0) + \left\langle \rho(\tilde{\theta}_n) - \rho(\bar{\theta}_0), \mu_n^*(\bar{\theta}_0, w)\right\rangle (1 + o(1))\right) \\
& -t_n^2 \left\|\mu_n^*(\bar{\theta}_0, w)\right\|^2 + o_p\left(n^{-1}\right).
\end{aligned}$$

By the definition of $\tilde{\theta}_n$ and Lemma 4.2, for all $\bar{\theta}_0 \in \Theta_0 \cap R$

$$0 = \frac{\phi(\tilde{\theta}_n)[w] - \phi(\bar{\theta}_0)[w]}{||\nu_n^*(\bar{\theta}_0, w)||_{sd}} = \left\langle \rho\left(\tilde{\theta}_n\right) - \rho\left(\bar{\theta}_0\right), \mu_n^*(\bar{\theta}_0, w)\right\rangle_{wp} + o_p(n^{-1/2}),$$

which implies that $\left\langle \rho\left(\tilde{\theta}_n\right) - \rho\left(\bar{\theta}_0\right), \mu_n^*(\bar{\theta}_0, w)\right\rangle_{wp} = o_p\left(n^{-1/2}\right)$.
Thus,

$$\begin{aligned}
&\bar{L}_n\left(\tilde{\theta}_n\right) - \bar{L}_n\left(\tilde{\theta}_n\left(t_n\right)\right) \\
= \;& 2t_n\left(\frac{1}{n}\sum_{i=1}^{n} E\left[\mu_n^*(Z_i, \bar{\theta}_0, w)|X_i\right]' \Sigma(X_i)^{-1}\rho(Z_i, \bar{\theta}_0)\right) - t_n^2 \left\|\mu_n^*(\bar{\theta}_0, w)\right\|_{wp}^2 + o_p(n^{-1}),
\end{aligned} \tag{B.33}$$

which is minimized at

$$t_n = \left\|\mu_n^*(\bar{\theta}_0, w)\right\|_{wp}^{-2}\left(\frac{1}{n}\sum_{i=1}^{n} E\left[\mu_n^*(Z_i, \bar{\theta}_0, w)|X_i\right]' \Sigma(X_i)^{-1}\rho(Z_i, \bar{\theta}_0)\right) = O_p\left(n^{-1/2}\right). \tag{B.34}$$

By plugging (B.34) into (B.33) and combining (B.32) we have for all $\bar{\theta}_0 \in \Theta_0 \cap R$,

$$\begin{aligned}
& n\left(\bar{L}_n\left(\tilde{\theta}_n\right) - \bar{L}_n\left(\hat{\theta}_n\right)\right) \\
\geq \;& \left(\frac{1}{\|\mu_n^*(\theta_0, w)\|_{wp}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} E\left[\mu_n^*(Z_i, \bar{\theta}_0, w)|X_i\right]' \Sigma(X_i)^{-1}\rho(Z_i, \bar{\theta}_0)\right)^2 + o_p(1).
\end{aligned} \tag{B.35}$$

On the other hand, let $\rho(Z, \theta_n^*, w) = \rho(Z, \hat{\theta}_n) + t_n \times \mu_n^*(\bar{\theta}_0, w)$ for $t_n$ be the one in (B.34). Then by Lemma C.6,

$$\begin{aligned}
&\bar{L}_n(\theta_n^*) - \bar{L}_n(\hat{\theta}_n) \\
= \;& 2t_n\left\{\frac{1}{n}\sum_{i=1}^{n} E\left[\mu_n^*(Z_i, \bar{\theta}_0, w)|X_i\right]' \Sigma(X_i)^{-1}\rho(Z_i, \bar{\theta}_0) + \left\langle \rho\left(\hat{\theta}_n\right) - \rho(\theta_0), \mu_n^*(\bar{\theta}_0, w)\right\rangle_{wp}\right\} \\
& +t_n^2 \times \left\|\mu_n^*(\bar{\theta}_0, w)\right\|_{wp}^2 + o_p(n^{-1}).
\end{aligned} \tag{B.36}$$

Let $\varepsilon_n = o(n^{-1/2})$ for any $\theta_n \in \mathcal{B}_n(\theta_0) \subset \Theta_n$, consider a local pertubation $\rho(Z, \theta(\varepsilon_n)) = \rho(Z, \theta_n) \pm \varepsilon_n \times$

$\mu_n^*(\bar{\theta}_0, w)$. By the definition of $\hat{\theta}_n$, we have

$$\begin{aligned} -o_p(n^{-1}) &\leq \bar{L}_n\left(\hat{\theta}_n(\varepsilon_n)\right) - \bar{L}_n\left(\hat{\theta}_n\right) \\ &= \pm 2\varepsilon_n \left\{ \frac{1}{n}\sum_{i=1}^n E\left[\mu_n^*(\bar{\theta}_0, w)|X_i\right]' \Sigma(X_i)^{-1}\rho(Z_i, \bar{\theta}_0) \right\} \mp 2\varepsilon_n \left\langle \rho(\hat{\theta}_n) - \rho(\bar{\theta}_0), \mu_n^*(\bar{\theta}_0, w) \right\rangle_{wp} \\ &\quad + \varepsilon_n^2 \times ||\mu_n^*(\bar{\theta}_0, w)||_{wp}^2 + o_p(n^{-1}). \end{aligned}$$

where the second equality follows from essentially the same caculations as those of Lemma C.6. Thus, by the fact that $\varepsilon_n = o(n^{-1/2})$ and $||\mu_n^*(\bar{\theta}_0, w)||_{wp}^2 = O(1)$, we have

$$\frac{1}{n}\sum_{i=1}^n E\left[\mu_n^*(Z_i, \bar{\theta}_0, w)|X_i\right]' \Sigma(X_i)^{-1}\rho(Z_i, \bar{\theta}_0) + \left\langle \rho\left(\hat{\theta}_n\right) - \rho(\bar{\theta}_0), \mu_n^*(\bar{\theta}_0, w) \right\rangle_{wp} = o_p(n^{-1/2}). \tag{B.37}$$

Combining (B.36) and (B.37) yields that

$$\bar{L}_n(\theta_n^*) - \bar{L}_n\left(\hat{\theta}_n\right) = t_n^2 \times \|\mu_n^*(\theta_0, w)\|_{wp}^2 + o_p(n^{-1}).$$

Furthermore, by the definition of $t_n$ in (B.34), we have

$$(t_n)^2 \times \left\|\mu_n^*(\bar{\theta}_0, w)\right\|_{wp}^2 = \left\|\mu_n^*(\bar{\theta}_0, w)\right\|_{wp}^{-2} \left( \frac{1}{n}\sum_{i=1}^n E\left[\mu_n^*(Z_i, \bar{\theta}_0, w)|X_i\right]' \Sigma(X_i)^{-1}\rho(Z_i, \bar{\theta}_0) \right)^2.$$

Therefore,

$$\bar{L}_n\left(\theta_n^*\right) - \bar{L}_n\left(\hat{\theta}_n\right) = \left\|\mu_n^*(\bar{\theta}_0, w)\right\|_{wp}^{-2} \left( \frac{1}{n}\sum_{i=1}^n E\left[\mu_n^*(Z_i, \bar{\theta}_0, w)|X_i\right]' \Sigma(X_i)^{-1}\rho(Z_i, \bar{\theta}_0) \right)^2 + o_p(n^{-1}).$$

If $\mathbf{F}\left(\rho\left(\cdot, \theta_n^*, w\right)\right) = \mathbf{F}\left(\rho\left(\cdot, \theta_0, w\right)\right)$, then under the null, by the definition of $\tilde{\theta}_n$,

$$\begin{aligned} & \bar{L}_n(\tilde{\theta}_n) - \bar{L}_n(\hat{\theta}_n) \\ \leq\ & \bar{L}_n(\theta_n^*) - \bar{L}_n(\hat{\theta}_n) + o_p(n^{-1}) \\ =\ & \left\|\mu_n^*(\bar{\theta}_0, w)\right\|_{wp}^{-2} \left( \frac{1}{n}\sum_{i=1}^n E\left[\mu_n^*(Z_i, \bar{\theta}_0, w)|X_i\right]' \Sigma(X_i)^{-1}\rho(Z_i, \bar{\theta}_0) \right)^2 + o_p(n^{-1}). \end{aligned} \tag{B.38}$$

The conclusion follows by Assumption A.25, (B.35) and (B.38).

However, it is possible that $\mathbf{F}\left(\rho\left(\cdot, \theta_n^*, w\right)\right) \neq \mathbf{F}\left(\rho\left(\cdot, \theta_0, w\right)\right)$. If the latter is the case, we show that there exists $t_n^* \in \mathcal{T}_n$ w.p.a.1 such that $\rho(\theta_n^*(t_n^*, w)) = \rho(\hat{\theta}_n) + t_n^* \mu_n^*(\bar{\theta}_0, w)$ satisfying the following two conditions: (i) $\mathbf{F}(\rho(\cdot, \theta_n^*(t_n^*, w)) = \mathbf{F}(\rho(\cdot, \theta_0, w))$ for all $w \in \mathcal{W}$ and (ii) $t_n^* = -t_n + o_p(n^{-1/2})$.

The proof strategy of the existence of such $t_n^*$ is similar to the ones in the proof of Theorem 3.6. First, by the Riesz representation theorem and Assumption A.23, for $\hat{\theta}_n$, there is a $\bar{K}_n =$

$o(n^{-1/2}||\mu_n^*(\bar{\theta}_0, w)||_{wp})$ such that

$$\begin{aligned} -\bar{K}_n \quad \leq \quad & \mathbf{F}(\rho(\cdot, \hat{\theta}_n) + t_n \mu_n^*(\bar{\theta}_0, w))[w] - \mathbf{F}(\rho(\theta_0, w))[w] - \left\langle \rho(\hat{\theta}_n) - \rho(\theta_0), v_n^*(\bar{\theta}_0, w) \right\rangle_{wp} \\ & -t||\mu_n^*(\bar{\theta}_0, w)||_{wp} \leq \bar{K}_n. \end{aligned}$$

Thus, for any $r \in \left\{|r| \leq 2\bar{K}_n ||\mu_n^*(\bar{\theta}_0, w)||_{wp} \sigma_n\right\}$, let $\mathcal{T}_n \equiv \{t \in \mathbf{R} : |t| \leq 4\bar{K}_n^2 \sigma_n\}$,

$$\underline{t}||\mu_n^*(\bar{\theta}_0, w)||_{wp}^2 = -\left\langle \mu_n^*(\bar{\theta}_0, w), \rho(\hat{\theta}_n) - \rho(\theta_0) \right\rangle_{wp} - 2\bar{K}_n \left\|\nu_n^*(\bar{\theta}_0, w)\right\|_{sd}^{-1} + r \left\|\nu_n^*(\bar{\theta}_0, w)\right\|_{sd}^{-1}$$

and

$$\bar{t}||\mu_n^*(\bar{\theta}_0, w)||_{wp}^2 = -\left\langle \mu_{n,}^*(\bar{\theta}_0, w), \rho(\hat{\theta}_n) - \rho(\theta_0) \right\rangle_{wp} + 2\bar{K}_n \left\|\nu_n^*(\bar{\theta}_0, w)\right\|_{sd}^{-1} + r \left\|\nu_n^*(\bar{\theta}_0, w)\right\|_{sd}^{-1}.$$

To show that $\underline{t} \in \mathcal{T}_n$, note that since

$$|\underline{t}| \lesssim ||\rho(\hat{\theta}_n) - \rho(\theta_0)||_{wp} \times ||\mu_n^*(\bar{\theta}_0, w)||_{wp} + 2\left|\bar{K}_n\right| \left\|\nu_n^*(\bar{\theta}_0, w)\right\|_{sd}^{-1} \times + |r| \left\|\nu_n^*(\bar{\theta}_0, w)\right\|_{sd}^{-1},$$

we have $|\underline{t}| \leq 4K_n^2 \varrho_n$. Then $\underline{t} \in \mathcal{T}_n$. By the same argument, $\bar{t} \in \mathcal{T}_n$. Therefore, by the definitions of $\underline{t}$ and $\bar{t}$ and $\hat{\theta}_n$, we have

$$\mathbf{F}\left(\rho\left(\hat{\theta}_n\right) + \underline{t}\mu_n^*(\bar{\theta}_0, w)\right)[w] - \mathbf{F}(\rho(\cdot, \theta_0))[w] \leq r(w) - \bar{K}_n < r$$

and

$$\mathbf{F}\left(\rho\left(\hat{\theta}_n\right) + \bar{t}\mu_n^*(\bar{\theta}_0, w)\right)[w] - \mathbf{F}(\rho(\cdot, \theta_0))[w] \geq r(w) + \bar{K}_n > r$$

By continuity of $\mathbf{F}(\rho(\hat{\theta}_n) + t_n \mu_n^*(\bar{\theta}_0, w))$ and the mean value theorem, there exists some $t_n^* \in [\underline{t}, \bar{t}]$ such that $t_n^* \in \mathcal{T}_n$ and $\mathbf{F}(\rho(\cdot, \hat{\theta}_n) + t_n^* \mu_n^*(\bar{\theta}_0, w))[w] = \mathbf{F}(\rho(\cdot, \theta_0))[w]$. Thus, $t_n^*$ satisfies Condition (i). Second, we show that $t_n^* = -t_n + o_p(n^{-1/2})$. By Assumption A.23,

$$\begin{aligned} \left|\mathbf{F}\left(\rho(\cdot, \hat{\theta}_n) + t_n^* \mu_n^*(\bar{\theta}_0, w)\right)[w] - \mathbf{F}(\rho(\cdot, \theta_0))[w] - \frac{d\mathbf{F}(\rho_0)}{d\rho}\left[\rho(\theta_n^*(t_n^*, w)) - \rho(\theta_0)\right]\right| / ||\nu_n^*(\bar{\theta}_0, w)||_{wp} \quad &= \quad o_p(n^{-1/2}) \\ \left|\frac{d\mathbf{F}(\rho_0)}{d\rho}\left[\rho(\theta_n^*(t_n^*, w)) - \rho(\theta_0)\right]\right| / ||\nu_n^*(\bar{\theta}_0, w)||_{wp} \quad &= \quad o_p(n^{-1/2}), \end{aligned}$$

where $\mathbf{F}\left(\rho(\cdot, \hat{\theta}_n) + t_n^* \mu_n^*(\bar{\theta}_0, w)\right)[w] - \mathbf{F}(\rho(\cdot, \theta_0))[w] = 0$. By the definition of $\|\cdot\|_{wp}$, it implies that

$$\left\langle \rho(\hat{\theta}_n) - \rho(\theta_0), \nu_n^*(\bar{\theta}_0, w)/||\nu_n^*(\bar{\theta}_0, w)||_{wp} \right\rangle - t_n^* ||\mu_n^*(\bar{\theta}_0, w)||^2 = o_p(n^{-1/2}).$$

Since $||\mu_n^*(\bar{\theta}_0, w)|| = O(1)$, by the definition of $t_n$ and (B.37), we have $t_n^* = -t_n + o_p(n^{-1/2})$. Thus, there exists a $t_n^*$ that satisfies Condition (i) and (ii). □

**Proof of Theorem 4.2**.

By Assumption 4.4, without loss of generality, we assume $J = 1$. Let $\nu_n^*(w, \bar{\theta}_0)$ be the empirical Riesz representer and $\mu_n^*\left(w, \bar{\theta}_0\right) = \frac{\nu_n^*\left(w, \bar{\theta}_0\right)}{\left\|\nu_n^*\left(w, \bar{\theta}_0\right)\right\|_{sd}}$. For all $\bar{\theta}_0 \in \Theta_0 \cap R$, let $\tilde{\theta}_n^\star(t_n) \in \mathcal{B}_n(\bar{\theta}_0)$ satisfy $\rho(Z, \tilde{\theta}_n^\star(t_n)) = \rho(Z, \tilde{\theta}_n^\star) - t_n \times \mu_n^*(\bar{\theta}_0, w)$ for $\left\{t_n \in \mathcal{T}_n \equiv \left\{t \in [-1, 1] : t \lesssim n^{-1/2}\right\}\right\}$. By the definition of $\hat{\theta}_n$, we have

$$\bar{L}_n^\star\left(\tilde{\theta}_n^\star\right) - \bar{L}_n^\star\left(\hat{\theta}_n^\star\right) \geq \bar{L}_n^\star\left(\tilde{\theta}_n^\star\right) - \bar{L}_n^\star\left(\tilde{\theta}_n^\star\left(t_n\right)\right). \tag{B.39}$$

By Lemma C.6, for all $w \in \mathcal{W}$ and $\bar{\theta}_0 \in \Theta_0 \cap R$,

$$\begin{aligned}
& \bar{L}_n^\star\left(\tilde{\theta}_n^\star\right) - \bar{L}_n^\star\left(\tilde{\theta}_n^\star\left(t_n\right)\right) \\
= \; & 2t_n\left(\frac{1}{n}\sum_{i=1}^n \zeta_i E\left[\mu_n^*(Z_i,\bar{\theta}_0,w)|X_i\right]'\Sigma(X_i)^{-1}\rho(Z_i,\bar{\theta}_0) - \left\langle \rho\left(\tilde{\theta}_n^\star\right) - \rho\left(\theta_0\right), \mu_n^*(\bar{\theta}_0,w)\right\rangle_{wp}\right) \\
& -t_n^2\left\|\mu_n^*(\bar{\theta}_0,w)\right\|_{wp}^2 + o_{p^\star}(n^{-1}).
\end{aligned}$$

Then by Assumption 4.4 and Assumption A.23, we have for all $w \in \mathcal{W}$,

$$\left|\left\langle \rho\left(\tilde{\theta}_n^\star\right) - \rho\left(\hat{\theta}_n\right), \mu_n^*(\bar{\theta}_0,w)\right\rangle_{wp}\right| = o_p\left(n^{-1/2}\right).$$

Thus,

$$\begin{aligned}
& L_n^\star\left(\tilde{\theta}_n^\star\right) - L_n^\star\left(\tilde{\theta}_n^\star\left(t_n\right)\right) \\
= \; & 2t_n\left(\frac{1}{n}\sum_{i=1}^n \zeta_i E\left[\mu_n^*(Z_i,\bar{\theta}_0,w)|X_i\right]'\Sigma(X_i)^{-1}\rho(Z_i,\bar{\theta}_0) - \left\langle \rho\left(\tilde{\theta}_n^\star\right) - \rho\left(\hat{\theta}_n\right) + \rho(\hat{\theta}_n) - \rho(\theta_0), \mu_n^*(\bar{\theta}_0,w)\right\rangle_{wp}\right) \\
& -t_n^2||\mu_n^*(\bar{\theta}_0,w)||_{wp}^2 + o_{p^\star}(n^{-1}) \\
= \; & 2t_n\left(\frac{1}{n}\sum_{i=1}^n \zeta_i E\left[\mu_n^*(Z_i,\bar{\theta}_0,w)|X_i\right]'\Sigma(X_i)^{-1}\rho(Z_i,\bar{\theta}_0) - \left\langle \rho(\hat{\theta}_n) - \rho(\theta_0), \mu_n^*(\bar{\theta}_0,w)\right\rangle_{wp}\right) + o_{p^\star}(n^{-1}) \\
& -t_n^2||\mu_n^*(\bar{\theta}_0,w)||_{wp}^2 + o_{p^\star}(n^{-1}) \\
= \; & 2t_n\left(\frac{1}{n}\sum_{i=1}^n (\zeta_i - 1) E\left[\mu_n^*(Z_i,\bar{\theta}_0,w)|X_i\right]'\Sigma(X_i)^{-1}\rho(Z_i,\bar{\theta}_0)\right) - t_n^2||\mu_n^*(\bar{\theta}_0,w)||_{wp}^2 + o_{p^\star}(n^{-1}),
\end{aligned} \tag{B.40}$$

where the last equality follows from the equation that $t_n \lesssim n^{-1/2}$ and the following equation

$$\frac{1}{n}\sum_{i=1}^n E\left[\mu_n^*(Z_i,\bar{\theta}_0,w)|X_i\right]'\Sigma(X_i)^{-1}\rho(Z_i,\bar{\theta}_0) = -\left\langle \rho(\hat{\theta}_n) - \rho(\theta_0), \mu_n^*(\bar{\theta}_0,w)\right\rangle_{wp} + o_p(n^{-1/2})$$

implied by Lemma C.6(ii). Note that (B.40) is minimized at

$$t_n^\star = \left\|\mu_n^*(\bar{\theta}_0,w)\right\|_{wp}^{-2}\left(\frac{1}{n}\sum_{i=1}^n (\zeta_i - 1) E\left[\mu_n^*(Z_i,\bar{\theta}_0,w)\right]'\Sigma(X_i)^{-1}\rho(Z_i,\bar{\theta}_0)\right). \tag{B.41}$$

Therefore,

$$\begin{aligned}
& n\left\{L_n^\star\left(\tilde{\theta}_n^\star\right) - L_n^\star\left(\hat{\theta}_n^\star\right)\right\} \\
\geq \; & \left\{\frac{1}{\|\mu_n^*(\theta_0,w)\|_{wp}}\frac{1}{\sqrt{n}}\sum_{i=1}^n (\zeta_i - 1) E\left[\mu_n^*(Z_i,\bar{\theta}_0,w)|X_i\right]'\Sigma(X_i)^{-1}\rho(Z_i,\theta_0)\right\}^2 + o_{p^\star}(1).
\end{aligned} \tag{B.42}$$

Suppose there is a $\theta_n^\star \in \Theta_n$ satisfies $\rho(Z,\theta_n^\star) = \rho(Z,\hat{\theta}_n^\star) + t_n^\star \times \mu_n^*(\bar{\theta}_0) + o_p(n^{-1/2})$ and $\phi(\theta_n^\star) = \phi(\hat{\theta}_n)$. Then

we obtain that under the null, by the definition of $\tilde{\theta}_n^\star$,

$$\bar{L}_n(\tilde{\theta}_n^\star) - \bar{L}_n(\hat{\theta}_n^\star) \quad \leq \quad \bar{L}_n(\theta_n^\star) - \bar{L}_n(\hat{\theta}_n^\star) + o_p(n^{-1}).$$

By similar arguments as above, we can show that

$$\begin{aligned} & \bar{L}_n(\theta_n^\star) - \bar{L}_n(\hat{\theta}_n^\star) \\ = \quad & -2t_n^\star \left\{ \frac{1}{n}\sum_{i=1}^n (\zeta_i - 1) E\left[\mu_n^*(Z_i, \bar{\theta}_0, w)|X_i\right]' \Sigma(X_i)^{-1} \rho(Z_i, \bar{\theta}_0) + \left\langle \rho\left(\hat{\theta}_n^\star\right) - \rho\left(\hat{\theta}_n\right), \mu_n^*(\bar{\theta}_0, w) \right\rangle_{wp} \right\} \\ & + \left(t_n^*\right)^2 \times \|\mu_n^*(\theta_0, w)\|_{wp}^2 + o_p(n^{-1}). \end{aligned}$$

Since $t_n^\star = O_p(n^{-1/2})$ and

$$\frac{1}{n}\sum_{i=1}^n (\zeta_i - 1) E\left[\mu_n^*(Z_i, \bar{\theta}_0, w)|X_i\right]' \Sigma(X_i)^{-1} \rho(Z_i, \bar{\theta}_0) + \left\langle \rho\left(\hat{\theta}_n^\star\right) - \rho\left(\hat{\theta}_n\right), \mu_n^*(\bar{\theta}_0, w) \right\rangle_{wp} = o_p(n^{-1/2}),$$

we have

$$\bar{L}_n(\theta_n^\star) - \bar{L}_n(\hat{\theta}_n^\star) = \left(t_n^*\right)^2 \times \|\mu_n^*(\theta_0, w)\|_{wp}^2 + o_p(n^{-1}).$$

By the definition of $t_n^*$ in (B.41), for all $\bar{\theta}_0 \in \Theta_0 \cap R$,

$$\begin{aligned} & n\left(\bar{L}_n(\tilde{\theta}_n^\star) - \bar{L}_n(\hat{\theta}_n^\star)\right) \leq n\left(\bar{L}_n(\theta_n^\star) - \bar{L}_n(\hat{\theta}_n^\star)\right) \\ = \quad & \left\|\mu_n^*(\bar{\theta}_0, w)\right\|_{wp}^{-2} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n (\zeta_i - 1) E\left[\mu_n^*(Z_i, \bar{\theta}_0, w)|X_i\right]' \Sigma(X_i)^{-1} \rho(Z_i, \bar{\theta}_0) \right)^2 + o_p(1). \end{aligned} \tag{B.43}$$

The existence of $\theta_n^\star$ can be shown essentially by the same way as the one in the proof of Theorem 4.1. The conclusion follows by combining (B.42) and (B.43). □

**Proof of Theorem 4.3.**

See the Supplemental Appendix.

Table 1: Size as a Function of $\tau$ (DGP1): Pointwise Tests

| Degree of endogeneity | $\lambda = 0.2$ | | $\lambda = 0.2$ | | $\lambda = 0.8$ | | $\lambda = 0.8$ | |
|---|---|---|---|---|---|---|---|---|
| Nominal size | $\tau = 0.10$ | | $\tau = 0.05$ | | $\tau = 0.10$ | | $\tau = 0.05$ | |
| | $\beta_0 = 0$, $\mathbb{H}_0: \ h(y_1) = h_0(y_1)$ | | | | | | | |
| | $t_n$ | $\text{QLR}_n$ | $t_n$ | $\text{QLR}_n$ | $t_n$ | $\text{QLR}_n$ | $t_n$ | $\text{QLR}_n$ |
| $h(y_1^{0.25})$ | 0.098 | 0.102 | 0.045 | 0.059 | 0.113 | 0.098 | 0.055 | 0.057 |
| $h(y_1^{0.5})$ | 0.096 | 0.102 | 0.042 | 0.052 | 0.113 | 0.101 | 0.055 | 0.059 |
| $h(y_1^{0.75})$ | 0.094 | 0.101 | 0.045 | 0.053 | 0.112 | 0.101 | 0.052 | 0.057 |
| | $\beta_0 = 1$, $\mathbb{H}_0: \beta = \beta_0, h(y_1) = h_0(y_1)$ | | | | | | | |
| | $\text{Wald}_n$ | $\text{QLR}_n$ | $\text{Wald}_n$ | $\text{QLR}_n$ | $\text{Wald}_n$ | $\text{QLR}_n$ | $\text{Wald}_n$ | $\text{QLR}_n$ |
| $h(y_1^{0.25})$ | 0.112 | 0.109 | 0.062 | 0.058 | 0.102 | 0.098 | 0.063 | 0.055 |
| $h(y_1^{0.5})$ | 0.113 | 0.109 | 0.063 | 0.059 | 0.103 | 0.094 | 0.064 | 0.054 |
| $h(y_1^{0.75})$ | 0.112 | 0.106 | 0.066 | 0.054 | 0.105 | 0.095 | 0.060 | 0.056 |
| | $\beta_0 = 1$, $\mathbb{H}_0: \beta = \beta_0, \bigtriangledown h(y_1) = \bigtriangledown h_0(y_1)$ | | | | | | | |
| | $\text{Wald}_n$ | $\text{QLR}_n$ | $\text{Wald}_n$ | $\text{QLR}_n$ | $\text{Wald}_n$ | $\text{QLR}_n$ | $\text{Wald}_n$ | $\text{QLR}_n$ |
| $h(y_1^{0.25})$ | 0.095 | 0.101 | 0.041 | 0.057 | 0.110 | 0.101 | 0.046 | 0.052 |
| $h(y_1^{0.5})$ | 0.095 | 0.101 | 0.040 | 0.054 | 0.118 | 0.101 | 0.043 | 0.053 |
| $h(y_1^{0.75})$ | 0.094 | 0.102 | 0.040 | 0.055 | 0.111 | 0.103 | 0.042 | 0.058 |

Table 2: Empirical Coverage as a Function of Normal Level (DGP1): Uniform Results

| | $n = 500$ | | | | $n = 1,000$ | | | |
|---|---|---|---|---|---|---|---|---|
| | $\beta_0 = 0$ | | $\beta_0 = 1$ | | $\beta_0 = 0$ | | $\beta_0 = 1$ | |
| | $\mathbb{H}_0: h(w) = h_0(w)$ | | $\mathbb{H}_0: \beta = \beta_0, h(w) = h_0(w)$ | | $\mathbb{H}_0: h(w) = h_0(w)$ | | $\mathbb{H}_0: \beta = \beta_0, h(w) = h_0(w)$ | |
| | sup-t | | sup-Wald | | sup-t | | sup-Wald | |
| | $\lambda = 0.2$ | $\lambda = 0.8$ | $\lambda = 0.2$ | $\lambda = 0.8$ | $\lambda = 0.2$ | $\lambda = 0.8$ | $\lambda = 0.2$ | $\lambda = 0.8$ |
| $\alpha = 2$ | | | | | | | | |
| $1 - \tau = 0.90$ | 0.88 | 0.85 | 0.88 | 0.92 | 0.89 | 0.90 | 0.89 | 0.91 |
| $1 - \tau = 0.95$ | 0.93 | 0.94 | 0.93 | 0.97 | 0.94 | 0.94 | 0.96 | 0.96 |
| $\alpha = 3$ | | | | | | | | |
| $1 - \tau = 0.90$ | 0.90 | 0.96 | 0.87 | 0.91 | 0.90 | 0.90 | 0.98 | 0.87 |
| $1 - \tau = 0.95$ | 0.94 | 0.95 | 0.93 | 0.96 | 0.95 | 0.96 | 0.94 | 0.94 |

Table 3: Size as a Function of $\tau$ and $B_n$ (DGP 2)

| | | $\mathbb{H}_0 : \sin(0) = 0$; n=500 | | | | | |
|---|---|---|---|---|---|---|---|
| | | $B = 50$ | | $B = 10^2$ | | $B = 10^3$ | |
| nominal size | $B_n$ | QLR-test | J-test | QLR-test | J-test | QLR-test | J-test |
| $\tau = 0.1$ | 50 | 0.097 | 0.108 | 0.101 | 0.102 | 0.103 | 0.104 |
| $\tau = 0.1$ | 100 | | 0.128 | | 0.106 | | 0.116 |
| $\tau = 0.1$ | 1000 | | 0.154 | | 0. 126 | | 0.126 |
| $\tau = 0.05$ | 50 | 0.054 | 0.060 | 0.053 | 0.054 | 0.050 | 0.052 |
| $\tau = 0.05$ | 100 | | 0.076 | | 0.068 | | 0.062 |
| $\tau = 0.05$ | 1000 | | 0.096 | | 0.076 | | 0.068 |
| $\tau = 0.01$ | 50 | 0.009 | 0.006 | 0.011 | 0.010 | 0.011 | 0.010 |
| $\tau = 0.01$ | 100 | | 0.008 | | 0.012 | | 0.016 |
| $\tau = 0.01$ | 1000 | | 0.020 | | 0.012 | | 0.016 |

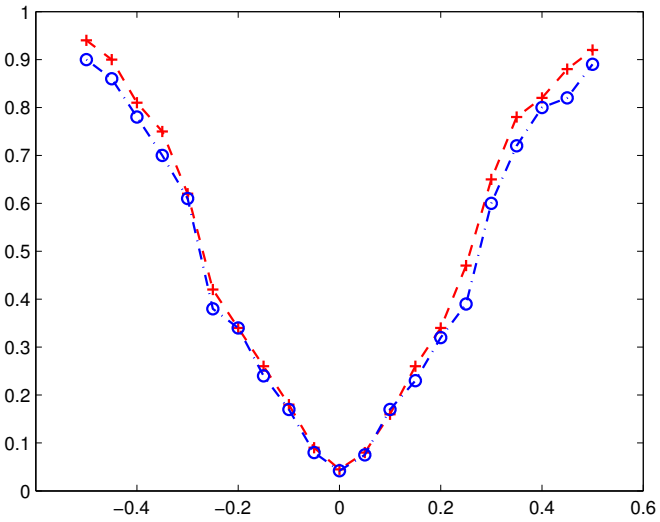


Figure 1. Power curves of the QLR test as a function of $B$ ($-*-$:$B = 10^2$, $-o-$: $B = 10^3$).

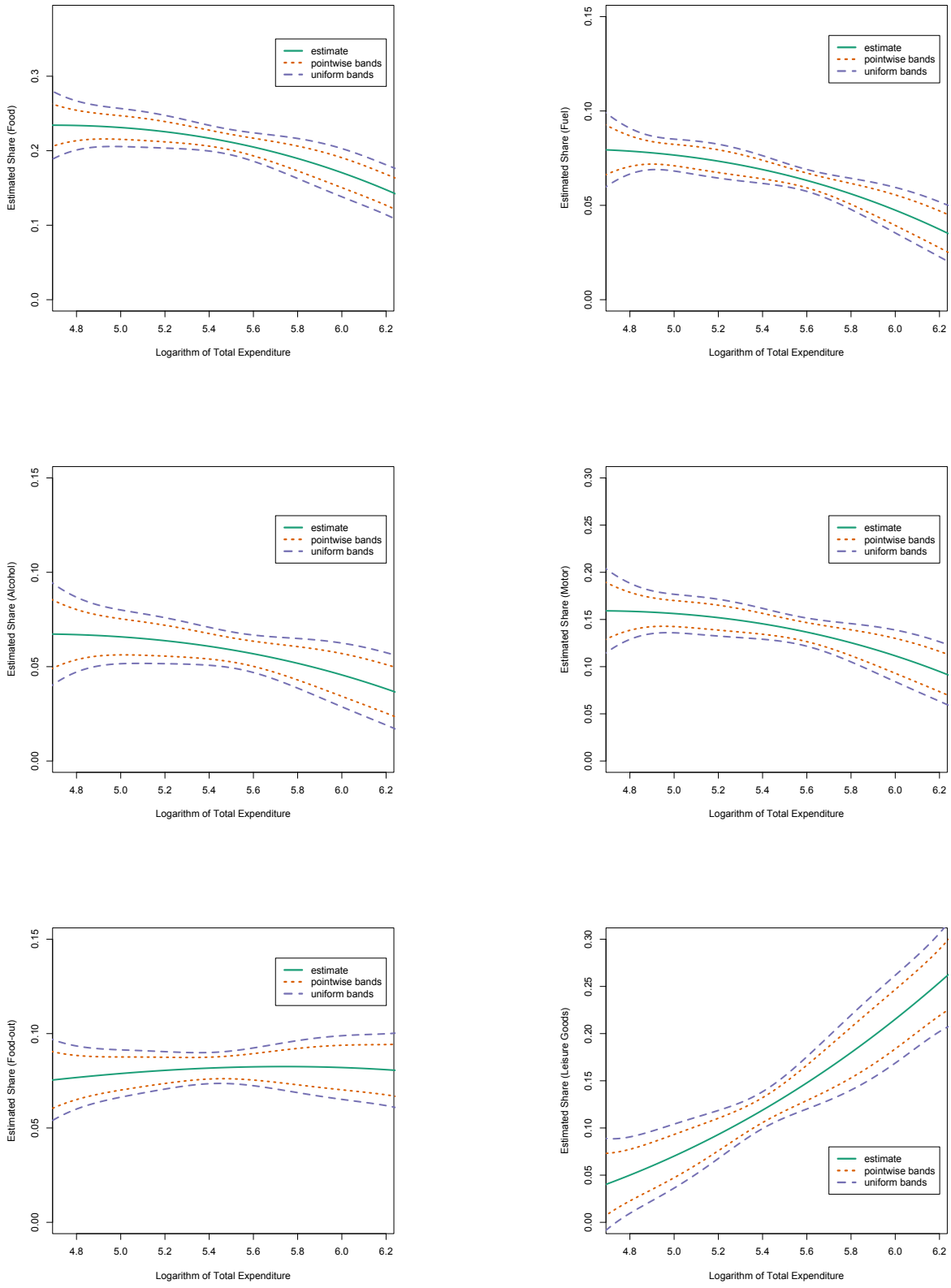


Figure 2: Estimates, Pointwise and Uniform Confidence Bands of Engel Curves

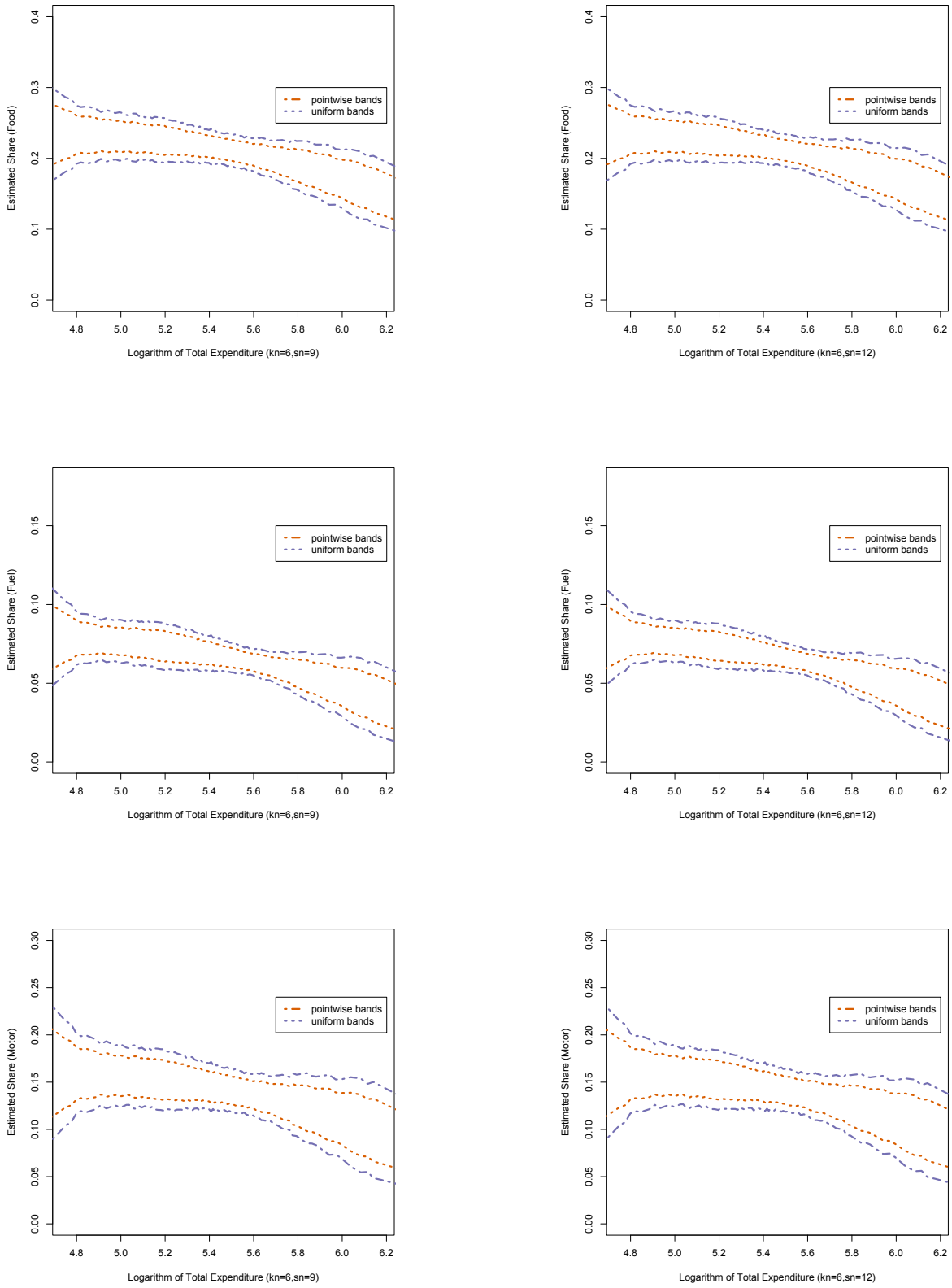


Figure 3: Pointwise and Uniform Confidence Bands of Engel Curves by QLR and sup-QLR