Article:
BRAVO, FRANCESCO (2026) Second order accurate inference for nonparametric estimating equations models. Journal of Nonparametric Statistics. ISSN: 1048-5252

https://doi.org/10.1080/10485252.2026.2658619

Second order accurate inference for nonparametric estimating equations models

Francesco Bravo*
University of York

March 2026

Abstract

This paper considers pointwise inference for nonparametric estimating equations models. The paper proposes two general test statistics that are based on a local version of the Generalized Empirical Likelihood (GEL) approach that can be used to test simple hypotheses about the unknown infinite dimensional parameters and to test for the correct specification of the chosen nonparametric estimating equations model. The paper shows that among the class of the proposed GEL test statistics, the empirical likelihood ratio is the only one admitting a Bartlett correction, however by appropriately modifying the other GEL based test statistics, it is still possible to obtain second order accurate inferences. The paper also proposes a new (local) version of the so-called efficient bootstrap that delivers the same level of second order accuracy as that of the (modified) GEL test statistics for the correct specification of the chosen nonparametric estimating equations model. Finally, the paper uses simulations and a real data example to illustrate the finite sample properties and applicability of the proposed inference methods.

Keywords: Bartlett correction, Coverage accuracy, Edgeworth expansion, Efficient bootstrap, Kernel estimation
Words count: 6969

*Address correspondence to: Department of Economics, University of York, York YO10 5DD, UK. E-mail: francesco.bravo@york.ac.uk. Web Page: https://sites.google.com/a/york.ac.uk/francescobravo/

1 Introduction

Parametric estimating equations (PEE) models, that is models where the unknown parameters of interest are finite dimensional, are widely used in Economics, Finance and Statistics. For example, many assets pricing models used in Finance naturally give rise to a set of (possibly conditional) estimating equations; likewise, quasi-likelihood methods for generalized linear models also give rise to a set of estimating equations. Estimation and inference for the unknown finite dimensional parameters of such models are typically carried out using Hansen's (1982) generalized method of moments (GMM) approach - see also Qu, Lindsay and Li (2000), or, alternatively, Newey and Smith's (2004) generalized empirical likelihood, with the asymptotic properties of the resulting estimators and test statistics being well-established. However, the chosen parametric specification is often too restrictive and can result in misspecified models. For example, the popular stochastic discount factor model widely used in the empirical assets pricing literature (see Cochrane (2001) for a comprehensive review) relies on a parametric specification of the utility function, which can result in a misspecified model, see for example Hansen and Jagannathan (1997) and more recently Gospodinov, Kan and Robotti (2013). Similarly, in quasi-likelihood estimation for generalized linear models the assumed linearity of the index might not be appropriate to model possible nonlinear relationships between the covariates, the response and the unknown parameters.

Nonparametric estimating equations (NPEE) models are useful extensions of PEE models, as they are robust to possible misspecification, while providing a very flexible way to model potentially complex nonlinear relationships often encountered with real data. Because of these desirable properties, several authors have considered NPEE models: Severini and Staniswalis (1994) considered nonparametric quasi-likelihood models, Cai (2003), Lewbel (2007) and Bravo (2022) considered generic NPEE models that can be used in the context of nonparametric binary choice, treatment effects and certain stochastic optimization models, Cai, Das, Xiong and Wu (2006) considered instrumental variables estimation of smooth coefficients model, Cai and Li (2008) considered nonparametric dynamic panel data models, Fang, Ren and Yuan (2011) and Cai, Ren and Sun (2015) considered nonparametric stochastic discount factor models.

All of the above papers focus mainly on estimation, whereas this paper considers pointwise inference. Pointwise inference is important for empirical applications: for example, it can be used in Economics in the context of demand elasticities and efficient frontiers, where it is often important to evaluate these functions at a specific point. Similarly, in the nonparametric treatment effect literature, it is important to evaluate the effectiveness of a given treatment at a specific point.

This paper focuses on the second order properties of two general test statistics that can be used to test simple hypotheses about the unknown infinite dimensional parameters as well as to test for the correct specification of the chosen NPEE model. The proposed test statistics are based on a kernel version of the Generalized Empirical Likelihood (GEL) approach. GEL was introduced by Smith (1997) and Newey and Smith (2004) as a quasi-likelihood alternative to GMM which includes as special cases Owen's (1988) empirical likelihood (EL) and continuous updating (CU) GMM (Hansen, Heaton and Yaron 1996). One important property of GEL is that, as opposed to GMM, it does not require the estimation of a weight matrix, which, as shown by Newey and Smith (2004) and Bravo (2022) is an important source of the finite sample bias for GMM estimators, and it might also explain the relatively poor finite sample performance of GMM based test statistics, such as Hansen's (1982) overidentifying restrictions statistic. To the best of our knowledge this is the first paper that develops and justifies rigorously a second order asymptotic theory for inference in NPEE models. It is important to note that in order to obtain second order accurate test statistics we need to deal with the bias resulting from the kernel estimation. One possibility is to directly estimate the bias. This is the approach followed by Calonico, Cattaneo and Farrell (2018) (although not in the context of the overidentified NPEE models considered in this paper). Alternatively, one can undersmooth, which is what we do in this paper, since, as noted by Hall (1992a), undersmoothing yields more accurate confidence regions and smaller errors in the rejection probabilities (see also Hall (1992b)). The new results of the paper are the following:

First, it considers inference for the unknown infinite dimensional parameter when the model is exactly identified and obtain Edgeworth expansions for local GEL test statistics that can be used to obtain confidence regions with improved coverage accuracy. This case is empirically relevant as it includes many nonparametric regression models, where there is some possible correlation between the covariates and the unobservable errors, see Section 4.2 for an example. The expansions show that in the class of local GEL test statistics, the one based on EL is the only one Bartlett correctable, that is, it is the only statistic that admits a scale correction that improves the order of magnitude of its coverage error. The expansions can also be used to obtain Bartlett-type corrections to the other members of the family of local GEL test statistics, as well as to obtain Cornish-Fisher type expansions for the upper quantile of the asymptotic χ^2 distribution of the test statistics.

Second, it considers a test for the correct local specification of the chosen NPEE model that is in the same spirit of Hansen's (1982) overidentifying restrictions test. This paper shows that by using the same linear reparametrization of Chen and Cui (2006) and Matsushita and Otsu (2013), the resulting local EL test statistic is still Bartlett correctable. Bartlett-type corrections for the other members of the local GEL family of test statistics for overidentifying restrictions are also considered, as well as Cornish-Fisher type expansions for the upper quantile of the relevant asymptotic χ^2 distribution.

Although theoretically interesting, both the Bartlett and Bartlett-type corrections for the local GEL test statistic for overidentifying restrictions are very complicated to compute in practice, so it is useful to consider alternative methods that deliver the same level of accuracy. The efficient bootstrap of Brown and Newey (2002) seems a natural way to incorporate the additional local information into the inference process, as it imposes the null hypothesis on the resampled observations. The third contribution of this paper is to show that a local version of the efficient bootstrap can be used to obtain second order improvements to the distribution of the local GEL test statistic for overidentifying restrictions. As an additional technical contribution, the paper establishes a new generic efficient bootstrap uniform law of large numbers for kernel based estimators that is of independent interest.

The fourth contribution of this paper is to illustrate its results both analytically and numerically with two examples: a nonparametric binary choice model and a smooth coefficient model with endogenous covariates, that is covariates that are possibly correlated with the unobservable errors.

The final contribution of this paper is to consider an empirical application, where the relationship between earnings, years of schooling and work experience is examined using a varying coefficients specification that allows for the endogeneity of the covariates, which shows the usefulness of the proposed inference methods.

These new results contribute to the literature on higher order asymptotics of EL and more generally GEL inference for models with infinite dimensional parameters, which includes Chen (1996), Chen and Qin (2000), Otsu, Xu and Matsushita (2015) and Bravo (2022) among others.

The rest of the paper is structured as follows: next section introduces the NPEE model, two motivating examples and the test statistics. Section 3 contains the main results. Section 4 contains the results of some Monte Carlo simulations that are used to illustrate the finite sample properties of the proposed test statistics. Section 5 contains the empirical application, whereas section 6 contains some concluding remarks. An Appendix contains all the proofs, whereas an online Appendix contains additional Monte Carlo results and robustness checks for the empirical application.

The following notation is used throughout the paper: "^{T}" indicates transpose, a prime and double prime indicate, respectively, first and second derivatives of a function with respect to its argument. Finally, as it is customary in the higher order asymptotics literature, the paper uses tensor (or index) notation - see for example McCullagh (1987), so for the components indices j, k, l etc., v_j,v_{jk} and v_{jkl} represent, respectively, a vector, a matrix and a three dimensional array. For such arrays the summation convention is also used, meaning that for any two (or more) repeated indices their sum is understood, so for example $v_j v_j = \sum_j v_j v_j$.

2 The NPEE model

Let $\{W_i\}_{i=1}^n$ denote a random sample from the distribution P_W of W taking values in $\mathcal{W} \subset \mathbb{R}^{d_W}$, let $\{Z_i\}_{i=1}^n$ denote a random sample of so-called instruments from the distribution P_Z taking values in $\mathcal{Z} \subset \mathbb{R}^{d_Z}$[1] and let $h \in \mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2 \times ... \times \mathcal{H}_k$ denote a k dimensional vector of unknown functions - the infinite dimensional unknown parameters, where $\mathcal{H}_j (j = 1, ..., k)$ are pseudo-metric spaces of functions.

The NPEE model we consider is defined through a vector of known functions $g : \mathcal{W} \times \mathcal{Z} \times \mathcal{H} \to \mathbb{R}^l$ $(l \geq k)$ such that

$$E\left[g\left(h\left(Z\right), W\right)|Z\right] = 0 \ a.s. \text{ for a unique } h_0 \in \mathcal{H}; \tag{1}$$

note that for a given point $z \in \mathcal{Z}$ (1) implies that

$$E\left[g\left(h_0\left(z\right), W\right)|Z = z\right] f(z) = 0, \tag{2}$$

where $f(\cdot)$ is the marginal density of Z.

The following two examples illustrate the type of NPEE models that are the focus of this paper:

Example 1 *Consider the following nonparametric binary choice model*

$$Y = I\left(h_0\left(Z\right) - \varepsilon \geq 0\right) \tag{3}$$

where $I(\cdot)$ is the indicator function, $h : \mathcal{Z} \to \mathbb{R}$ is an unknown function and ε is an unobservable error assumed to be independent of Z. Let F_ε denote the known distribution of ε; then $E\left(Y - F_\varepsilon\left(h_0\left(Z\right)\right)|Z\right) = 0$ a.s. and $g(h(Z), W) = Y - F_\varepsilon(h(Z))$.

Example 2 *Consider the following smooth coefficients model*

$$Y = X^T h_0\left(Z\right) + \varepsilon, \tag{4}$$

where X and Z are, respectively, a k and d_Z dimensional vectors of covariates and $h : \mathcal{Z} \to \mathbb{R}^k$ is a vector of unknown functions. We assume that the X covariates are endogenous and that $E\left(\varepsilon|Z\right) = 0$ a.s., so that the Z covariates can be used as instruments. Let $g\left(h\left(Z\right), W\right) = r\left(Z\right)\left(Y - X^T h\left(Z\right)\right)$, where $r : \mathcal{Z} \to \mathbb{R}^l$ $(l \geq k)$ is a vector of known functions. Then $E\left(r\left(Z\right)\varepsilon|Z\right) = 0$ a.s., where $\varepsilon = Y - X^T h_0\left(Z\right)$.

Let $b =: b(n)$ denote the bandwidth and $K : \mathcal{Z} \to \mathbb{R}$ denote a multivariate (product) kernel function; then (2) suggests that estimation and inference for h_0 can be based on

$$g\left(h\left(z\right), W\right) K\left(\frac{Z - z}{b}\right), \tag{5}$$

[1]Note that Z may contain some of the elements of W.

which forms the basis for the inference considered in this paper. To be specific, let

$$\Gamma_{\gamma}(\lambda, h, z) = \gamma\left(\lambda(z)^{T} g(h(z), W) K\left(\frac{Z-z}{b}\right)\right)$$

denote the local GEL criterion function, where γ is a known scalar valued function and λ is a vector of auxiliary unknown parameters. Examples of $\Gamma_{\gamma}(\lambda, h, z)$ include

$$\Gamma_{EL}(\lambda, h, z) = \log\left(1 - \lambda(z)^{T} g(h(z), W) K\left(\frac{Z-z}{b}\right)\right)$$

which corresponds to a local version of Owen's (1988) EL,

$$\Gamma_{ET}(\lambda, h, z) = -\exp\left(\lambda(z)^{T} g(h(z), W) K\left(\frac{Z-z}{b}\right)\right)$$

which corresponds to a local version of exponential tilting (see for example Kitamura and Stutzer (1997)), and finally

$$\Gamma_{CU}(\lambda, h, z) = -\frac{1}{2}\left(1 + \lambda(z)^{T} g(h(z), W) K\left(\frac{Z-z}{b}\right)\right)^{2}$$

which corresponds to a local version of the continuous updating (CU) estimator of Hansen et al. (1996).

Given a random sample $\{W_i^T, Z_i^T\}_{i=1}^{n\,T}$ and a suitable restriction, say $\mathcal{H}_C$, on the parameter space $\mathcal{H}$ to be defined precisely in Section 3 below, let

$$g_i^K(h(z)) := g(h(z), W_i) K\left(\frac{Z_i - z}{b}\right);$$

denote the local version of the NPEE $g(\cdot)$, and let

$$\widehat{h}(z) := \arg\min_{h \in \mathcal{H}_C} \widehat{\Gamma}_{\gamma}\left(\widehat{\lambda}, h, z\right) \tag{6}$$

where

$$\begin{aligned} \widehat{\Gamma}_{\gamma}(\lambda, h, z) &= \frac{1}{nb^{d_Z}} \sum_{i=1}^{n} \gamma\left(\lambda(z)^{T} g_i^K(h(z))\right) \text{ and} \\ \widehat{\lambda}(z) &= \arg\max_{\lambda \in \Lambda_n(\overline{h})} \widehat{\Gamma}_{\gamma}\left(\lambda, \overline{h}, z\right) \end{aligned} \tag{7}$$

for some $\overline{h} \in \mathcal{H}_C$, with $\Lambda_n(h) = \left\{\lambda(z)^{T} g_i^K(h(z)) \in \mathcal{V}_0, i = 1, ..., n\right\}$, where $\mathcal{V}_0$ is an open interval containing 0 for all $z \in \mathcal{Z}$, and $\widehat{\lambda}(z)$ can be interpreted as the local Lagrange multiplier estimator associated with (2).

The two test statistics we consider are

$$S_\gamma\left(\widehat{\lambda}, h_0\right) = 2nb^{d_Z}\left(\widehat{\Gamma}_\gamma\left(\widehat{\lambda}, h_0, z\right) - \widehat{\Gamma}_\gamma\left(0, h_0, z\right)\right) \tag{8}$$

and

$$S_\gamma\left(\widehat{\lambda}, \widehat{h}\right) = 2nb^{d_Z}\left(\widehat{\Gamma}_\gamma\left(\widehat{\lambda}, \widehat{h}, z\right) - \widehat{\Gamma}_\gamma\left(0, \widehat{h}, z\right)\right), \tag{9}$$

which can be used, respectively, to construct confidence regions for $h_0(\cdot)$ at $Z = z$ under the additional assumption that $\dim(g) = \dim(h)$[2] and to test for the local correct specification of $E\left[g\left(h_0(z), W\right)|Z = z\right] f(z) = 0$.

3 Asymptotic results

We first discuss the parameter space $\mathcal{H}_C$. Since the seminal work of Wong and Severini (1991, Theorem 1), $\mathcal{H}_C$ is often assumed to be a compact set (see for example Lewbel (2007) and Bravo (2022)). Note that compactness can also be deduced indirectly using various compact embedding results (see, for example, Nickl and Potscher (2007)) under some additional regularity conditions on $\mathcal{H}$. For example, suppose that $\mathcal{H} = \mathcal{W}^C_{p+p_0,\infty}$ for some finite positive constant C, that is $\mathcal{H}$ is the space of functions with Sobolev sup norm $\max_{|\alpha|\leq p+p_0} ||D^\alpha f||_\infty \leq C$, where D^α is a multi-index differential operator. Then the embedding $\mathcal{W}^C_{p+p_0,\infty} \hookrightarrow \mathcal{W}^C_{p,\infty}$ is compact and $\mathcal{H}_C = \mathcal{W}^C_{p,\infty}$. Let

$$G(z) = E\left[\partial g\left(h_0(Z), W\right)/\partial h^T|Z = z\right], \quad \Omega(z) = E\left[g\left(h_0(Z), W\right) g(h_0(Z), W)^T|Z = z\right],$$

and assume that:

A1 (i) There exists a unique $h_0 \in \mathcal{H}_C$ such that $E\left[g\left(h_0(z), W\right)|Z = z\right] = 0$, (ii) h_0 is twice continuously differentiable on $\mathcal{Z}$ with uniformly bounded derivatives, (iii) The d^Z-dimensional random vector Z has compact support $\mathcal{Z}$ and its density f is bounded away from 0 on $\mathcal{Z}$,

A2 (i) $\partial g(h)/\partial h^T$ exists and is continuous for each $h \in \mathcal{H}_C$ *a.s.*, (ii) the classes of functions $\mathcal{G}^K_1 = \left\{g^K(h(z)), b > 0, z \in \mathcal{Z}, h \in \mathcal{H}_C\right\}$ and $\mathcal{G}^K_{\partial h} = \left\{\partial g^K(h(z))/\partial h^T, b > 0, z \in \mathcal{Z}, h \in \mathcal{H}_C\right\}$ are Euclidean with integrable envelopes G_1 and G_{∂_1}, (iii) the envelope $G_2 = G_1^2$ is integrable,

[2] Note that second order accurate confidence regions for $h_0(z)$ in the general case $\dim(g) > \dim(h)$ can be constructed using the test statistic

$$S_\gamma(\widetilde{\lambda}, \widehat{\lambda}, h_0, \widehat{h}) = 2nb^{d_Z}(\widehat{\Gamma}_\gamma(\widetilde{\lambda}, h_0, z) - \widehat{\Gamma}_\gamma(\widehat{\lambda}, \widehat{h}, z)),$$

where $\widetilde{\lambda} = \arg\max_{\lambda\in\Lambda_n(h_0)} \widehat{\Gamma}_\gamma(\lambda, h_0, z)$. To do so, one needs to expand $\widehat{\Gamma}_\gamma(\widetilde{\lambda}, h_0, z)$ (along the lines of the case $\dim(g) = \dim(h)$), use the expansion for $\widehat{\Gamma}_\gamma(\widehat{\lambda}, \widehat{h}, z)$ and combine both expansions.

A3 for all $z \in \mathcal{Z}$ (i) $E\left\|g\left(h\left(z\right),W\right)\right\|^{\delta} < \infty$ for all $h \in \mathcal{H}_C$ and some $\delta > 2$, (ii) $\Omega\left(z\right)$ is positive definite, (iii) $rank\left(G\left(z\right)\right) = k$,

A4 $\gamma\left(v\left(z\right)\right)$ is twice continuously differentiable in v in a neighborhood of 0 for all $z \in \mathcal{Z}$ with $\gamma_j = \partial^j \gamma\left(v\right)/\partial v^j|_{v=0} = -1$ for $\left(j = 1, 2\right)$,

A5 The kernel function $K\left(\cdot\right)$ is symmetric and has compact support, say $[-1, 1]$.

Assumption **A1**(i) is a local identification condition, which is implied by (1) and can be often verified by imposing more primitive conditions on g and/or the covariates W and Z (see Section 4 below for two examples). **A1**(ii)-(iii) are standard in the nonparametric estimation literature; in particular A2(ii) ensures that the C used to define $\mathcal{H}_C$ exists. Assumption **A2**(i) requires smoothness of g in terms of its first derivative and is used to show the (pointwise) asymptotic normality of the estimators $\widehat{\lambda}$ and $\widehat{h}$. Assumption **A2**(ii) is a high level condition requiring that the classes of functions $\mathcal{G}_1^K$ and $\mathcal{G}_{\partial_1}^K$ are Euclidean with integrable envelopes - see Pakes and Pollard (1989) for a definition of an Euclidean class of functions. In Proposition 7.1 in the Supplemental appendix we provide a set of sufficient conditions that imply **A2**(ii). Assumptions **A3**-**A4** are similar to those used by Newey and Smith (2004); in particular **A4** is a normalization assumption that can be always be achieved by a simple rescaling of the local GEL criterion function -see Newey and Smith (2004)[p. 222]. Finally, **A5** is standard in the nonparametric estimation literature. Under the above regularity conditions, it is possible to obtain a uniform first order expansion for $\widehat{\lambda}(z)$, $\widehat{h}(z)$ and obtain their uniform convergence rate - see for example Bravo (2022). Note also that **A3**(i) for $h = h_0$ and a given $Z = z$, is sufficient for the Lyapunov central limit theorem to apply; combined with **A3**(iii) it yields the asymptotic normality of the estimators $\widehat{\lambda}\left(z\right)$ and $\widehat{h}\left(z\right)$. In particular, for $\left(nb^{d_Z}\right)^{1/2} \to \infty$

$$\left(nb^{d_Z}\right)^{1/2}\left(\widehat{h}\left(z\right) - h_0\left(z\right) - b^2 B\left(z\right)\right) \xrightarrow{d} N\left(0, \frac{\Sigma\left(z\right)}{f\left(z\right)}\right), \tag{10}$$

where

$$\begin{aligned}
\Sigma\left(z\right) &= \left(G\left(z\right)^T \Omega\left(z\right)^{-1} G\left(z\right)\right)^{-1} \int K^2\left(u\right) du, \\
B\left(z\right) &= \frac{1}{2f\left(z\right)} \sum_{j=1}^{d_Z} \left(h_{j0}''\left(z\right) f\left(z\right) + 2h_0'\left(z\right) f'\left(z\right)^T\right) \int u u_j K\left(u\right) du.
\end{aligned}$$

which shows that the asymptotic mean squared error $\left(AMSE\right)$ for $\widehat{h}\left(z\right)$ is

$$AMSE\left(\widehat{h}\left(z\right)\right) = \frac{b^4}{4}\left\|B\left(z\right)\right\|^2 + \frac{\int K^2\left(u\right) du}{nb^{d_Z}} tr\left(\Sigma\left(z\right)\right), \tag{11}$$

where $tr\left(\cdot\right)$ is the trace operator, which implies that the optimal bandwidth $b_*^{d_Z}$ minimizing (11) is

$$b_*^{d_Z} = \left(\frac{1}{n}\right)^{\frac{1}{d_Z+4}} \left(d_Z \int K^2\left(u\right) du tr\left(\Sigma\left(z\right)\right)\left\|B\left(z\right)\right\|^{-2}\right)^{\frac{1}{d_Z+4}}, \tag{12}$$

and the optimal convergence rate is of order $n^{-4/(d_Z+4)}$.

We consider the problems of testing simple hypotheses and constructing confidence regions for h_0 at a given point z - when $\dim(h) = \dim(g)$, and of testing for the correct local specification of (1). As mentioned in the Introduction, to obtain valid inferences for the unknown parameter h_0 one needs to correct for the asymptotic bias either by directly estimating it or with undersmoothing, which is the approach we follow. Note that to construct confidence regions for h_0 using a score or Wald statistic or to test for the correct specification of (1) requires the explicit estimation of $\Omega(z)$, $\Sigma(z)$ and of the density $f(z)$. On the other hand this additional estimation is not required by the local GEL statistics (8) and (9); indeed under the undersmoothing assumption $nb^{d_Z+4} \to 0$, a second order Taylor expansion and standard results on quadratic forms in asymptotically normal random vectors show that (8) and (9) enjoy the so-called Wilks' phenomenon (Wilks 1938), that is

$$S_\gamma\left(\widehat{\lambda}, h_0\right) = nb^{d_Z}\left(\frac{1}{nb^{d_Z}}\sum_{i=1}^{n} g_i^K(h_0)^T\right)\left(\frac{1}{nb^{d_Z}}\sum_{i=1}^{n} g_i^K(h_0)^{\otimes 2}\right)^{-1}\frac{1}{nb^{d_Z}}\sum_{i=1}^{n} g_i^K(h_0) + o_p(1) \xrightarrow{d} \chi_k^2,$$

$$S_\gamma\left(\widehat{\lambda}, \widehat{h}\right) = nb^{d_Z}\left(\frac{1}{nb^{d_Z}}\sum_{i=1}^{n} g_i^K\left(\widehat{h}\right)^T\right)\left(\frac{1}{nb^{d_Z}}\sum_{i=1}^{n} g_i^K\left(\widehat{h}\right)^{\otimes 2}\right)^{-1}\frac{1}{nb^{d_Z}}\sum_{i=1}^{n} g_i^K\left(\widehat{h}\right) + o_p(1) \xrightarrow{d} \chi_{l-k}^2,$$

where $g_i^K(h_0)^{\otimes 2} = g_i^K(h_0)g_i^K(h_0)^T$. Note that as with all specification tests, rejection of the null hypothesis could be due to the fact that either h could be misspecified, or because the instruments Z_i are not valid or because of a misspecified g. We first consider the second order properties of $S_\gamma\left(\widehat{\lambda}, h_0\right)$. Note that by inverting $S_\gamma\left(\widehat{\lambda}, h_0\right)$ one can construct confidence regions for h_0 with nominal coverage $1-\alpha$, that is for $P\left(\chi_k^2 \le c_\alpha\right) = 1-\alpha$ and

$$R_\alpha(h) = P\left(h | S_\gamma\left(\widehat{\lambda}, h\right) \le c_\alpha\right),$$

we have $P\left(h_0 \in R_\alpha(h)\right) = 1 - \alpha + o(1)$.

It is convenient to switch from matrix to the tensor notation typically used in higher order asymptotics, see for example McCullagh (1987). For the indices $j, l, ..$ running from 1 to k, let $\omega_{jl}(h) = E\left[g_j^K(h)\, g_l^K(h)/b^{d_Z}\right]$, $\left[\omega^{jl}(h)\right]^{1/2}$ denote the matrix square root of its inverse and, using the summation convention, define the standardized multivariate moments

$$\begin{aligned}\mu_{jlm\ldots}(h) &= E\frac{1}{b^{d_Z}}\left[\left[\omega^{jn}(h)\right]^{1/2} g_n^K(h)\left[\omega^{lo}(h)\right]^{1/2} g_o^K(h) \times\right. \\ &\quad \left.\left[\omega^{mp}(h)\right]^{1/2} g_p^K(h) \ldots\right],\end{aligned} \tag{13}$$

with $\mu_{jl}(h) = \delta_{jl}$ the Kronecker delta. Let $\overline{k} = k + k(k+1)/2 + k(k+1)(k+2)/3!$ denote the dimension of the vector

$$U_{\overline{k}}(h,K) = \Big[\left[\omega^{1m}(h)\right]^{1/2} g_m^K(h), ..., \left[\omega^{km}(h)\right]^{1/2} g_m^K(h), \\ \left[\omega^{1n}(h)\right]^{1/2} g_n^K(h) \left[\omega^{1o}(h)\right]^{1/2} g_o^K(h), ..., \\ \left[\omega^{kn}(h)\right]^{1/2} g_n^K(h) \left[\omega^{ko}(h)\right]^{1/2} g_o^K(h), \\ \left[\omega^{1n}(h)\right]^{1/2} g_n^K(h) \left[\omega^{1o}(h)\right]^{1/2} g_o^K(h) \left[\omega^{1p}(h)\right]^{1/2} g_p^K(h), ..., \\ \left[\omega^{kn}(h)\right]^{1/2} g_n^K(h) \left[\omega^{ko}(h)\right]^{1/2} g_o^K(h) \left[\omega^{kp}(h)\right]^{1/2} g_p^K(h)\Big]^T. \tag{14}$$

The following assumptions are sufficient for justifying the second order expansion for the coverage probability of the confidence region $R_\alpha(h)$ in the Edgeworth sense, see for example Hall (1991):

A4' for all $z \in \mathcal{Z}$ (i) $\gamma(v(z))$ is five times differentiable in v a neighborhood of 0, ii) there exists an integrable function φ such that $|\partial^5\gamma(v(z))/\partial v^5 - \gamma_5| \leq \varphi|v(z)|$,

A6 (i) $E\left(\|g(h_0,W)\|^{15}\right) < \infty$, (ii) for any $\theta \in \mathbb{R}^{\overline{k}}$ such that $\|\theta\| = 1$, there exists a partition of $[-1,1]$, $-1 = u_1 < ... < u_{m(\theta)} = 1$ such that $\theta^T U_{\overline{k}}(h_0,K)$ is either strictly increasing or decreasing on each interval (u_{m-1}, u_m) for $m = 1, ..., m(\theta)$.

Assumption **A6**(i) is sufficient to ensure the existence of the higher order moments required to prove the validity of the Edgeworth expansion of the signed squared root of the statistic $S_\gamma\left(\widehat{\lambda}, h_0\right)$, and is commonly assumed in the higher order asymptotics literature for empirical likelihood inference in both parametric estimating equations models (see for example Chen and Cui (2007) and Matsushita and Otsu (2013)) and nonparametric estimating equations models (see for example Otsu et al. (2015)); **A6**(ii) is similar to Assumption 5(c) of Horowitz (1998) and is used to verify Hall's (1991) analog of the standard Cramer condition (see also Hall (1992b)[Lemma 5.6]) which is used to justify the validity of the resulting Edgeworth expansion. Note that **A4'** and **A6** imply that the approximation error in the Edgeworth expansion of $S_\gamma\left(\widehat{\lambda}, h_0\right)$ is of order $O_p\left(\left(nb^{d_Z}\right)^{-2}\right)$. Before we state the main theoretical results of this section, it is useful to provide some intuitive explanations about them. The validity of the Edgeworth expansions of Theorems 1 and 2 follows by the (standard) argument based on the so-called signed squared root decomposition of the stochastic expansions (up to the order $O_p((nb^{d_Z})^{-3/2})$) of the test statistics (8) and (9). Both expansions are shown to be valid in the sense of Bhattacharya and Ghosh (1978); we then use the results of Chandra and Ghosh (1980) combined with the oddness/evenness property of the Hermite polynomials (Barndorff-Nielsen and Hall 1988) appearing in the formal Edgeworth expansions of $S_\gamma(\widehat{\lambda}, h_0)$ and $S_\gamma(\widehat{\lambda}, \widehat{h})$ to obtain the $O((nb^{d_Z})^{-2})$ error's order of magnitude in the expansions - see also DiCiccio and Romano (1989) for this point. For the Bartlett correction and generalized Bartlett correction given, respectively, in

Corollaries 1.1 and 1.2 the results follows by a standard application of the delta method and the general approximation results to a χ^2 distribution of Cox and Reid (1987a).

Theorem 1 *Under **A1**, **A4'**, **A6** and* $nb^{dz+4} \to 0$,

$$P\left(h_0 \in R_\alpha\left(h\right)\right) = 1 - \alpha + \frac{1}{nb^{dz}} \sum_{j=0}^{3} B_j\left(K\right) G_{k+2j}\left(c_\alpha\right) + O\left(\frac{1}{\left(nb^{dz}\right)^2}\right), \tag{15}$$

where

$$\begin{aligned} B_0\left(K\right) &= \frac{1}{72}\left(-b_3\left(K\right) + 3b_2\left(K\right) - 36b_1\left(K\right)\right), \\ B_1\left(K\right) &= \frac{1}{24}\left(b_3\left(K\right) - 2b_2\left(K\right) + 12b_1\left(K\right)\right), \\ B_2\left(K\right) &= \frac{1}{24}\left(-b_3\left(K\right) + b_2\left(K\right)\right), \quad B_3\left(K\right) = \frac{1}{72}b_3\left(K\right), \end{aligned}$$

and

$$\begin{aligned} b_1\left(K\right) &= \left(\gamma_3 - \frac{\gamma_4}{4}\right)\mu_{jjll}\left(h_0\right) + \left(1 + \frac{5\gamma_3}{3} + \frac{\gamma_3^2}{2}\right)\mu_{jlm}\left(h_0\right)\mu_{jlm}\left(h_0\right) + \\ &\quad \frac{\left(\gamma_3 + 2\right)^2}{36}\mu_{jjl}\left(h_0\right)\mu_{lmm}\left(h_0\right), \\ b_2\left(K\right) &= [3]\left[\left(\gamma_4 - 4\gamma_3 - 2\right)\mu_{jlmn}\left(h_0\right) + 2\left(4 + \frac{10\gamma_3}{3} + \frac{2\gamma_3^2}{3}\right)\mu_{jlm}\left(h_0\right)\mu_{noo}\left(h_0\right) + \right. \\ &\quad \left.\left(12 + 14\gamma_3 + 4\gamma_3^2\right)\mu_{jlo}\left(h_0\right)\mu_{mno}\left(h_0\right)\right]\delta_{jl}\delta_{mn}, \\ b_3\left(K\right) &= [15]\left(2 + \gamma_3\right)^2\mu_{jlm}\left(h_0\right)\mu_{nop}\left(h_0\right)\delta_{jl}\delta_{mn}\delta_{op}. \end{aligned}$$

Note that the terms $b_2\left(K\right)$ and $b_3\left(K\right)$ are 0 for EL (i.e. for $\gamma_3 = -2$ and $\gamma_4 = -6$), which implies that a simple scale correction to the original statistic - the Bartlett correction - improves the coverage accuracy of the local EL statistic to the order $\left(nb^{dz}\right)^{-2}$. Let

$$BC\left(K\right) = 1 + \frac{b_{1EL}\left(K\right)}{knb^{dz}} \tag{16}$$

denote the Bartlett correction.

Corollary 1.1 *Under the same conditions of Theorem 1,*

$$P\left(\frac{S_{EL}\left(\widehat{\lambda}, h_0\right)}{BC\left(K\right)} \leq c_\alpha\right) = 1 - \alpha + O\left(\frac{1}{\left(nb^{dz}\right)^2}\right).$$

It is important to note that expansion (15) can be used to improve the accuracy of the confidence regions for h_0 for all the other local GEL statistics. To be specific, we consider a modified local GEL statistic similar to that proposed by Cordeiro and Ferrari (1991) for parametric likelihood models, that is

$$S_{\gamma}^{m}\left(\widehat{\lambda},h_0\right)=S_{\gamma}\left(\widehat{\lambda},h_0\right)\left[1-2\sum_{j=1}^{3}\left(\sum_{l=j}^{3}B_l\left(K\right)\right)\frac{S_{\gamma}^{j-1}\left(\widehat{\lambda},h_0\right)}{nb^{d_Z}\Gamma_j}\right], \tag{17}$$

where $\Gamma_j=2^j\Gamma\left(k/2+j\right)/\Gamma\left(k/2\right)$, $\Gamma\left(\cdot\right)$ is the gamma function and the term in the square brackets can be interpreted as a generalized Bartlett correction. Let

$$R_{\alpha}^{m}\left(h\right)=P\left(h|S_{\gamma}^{m}\left(\widehat{\lambda},h\right)\leq c_{\alpha}\right)$$

denote the corresponding confidence region; the following corollary to Theorem 1 shows that $R_{\alpha}^{m}\left(h\right)$ is second order accurate.

Corollary 1.2 *Under the same assumptions of Theorem 1,*

$$P\left(h\in R_{\alpha}^{m}\left(h\right)\right)=1-\alpha+O\left(\frac{1}{\left(nb^{d_Z}\right)^2}\right). \tag{18}$$

It is also important to note that (17) can also be used to obtain the asymptotic expansion of the upper quantile of the distribution of $S_{\gamma}\left(\widehat{\lambda},h_0\right)$, that is for

$$d_{\gamma}=c\left(1+2\sum_{j=1}^{3}\left(\sum_{l=j}^{3}B_l\left(K\right)\right)\right)\frac{c^{j-1}}{nb^{d_Z}\Gamma_j}, \tag{19}$$

one has that

$$P\left(S_{\gamma}\left(\widehat{\lambda},h_0\right)\leq d_{\gamma}\right)=G_k\left(c\right)+O\left(\frac{1}{\left(nb^{d_Z}\right)^2}\right);$$

thus (19) can be used with c_{α} replacing c to directly improve the coverage accuracy of $R_{\alpha}\left(h\right)$.

In practice the $b_j\left(K\right)$'s are unknown and need to be estimated; let $\widehat{b}_j\left(K\right)$ denote the sample analog of the corresponding $b_j\left(K\right)$ with h_0 replaced by $\widehat{h}$ and let $S_{\gamma}^{\widehat{m}}\left(\widehat{\lambda},h_0\right)$ and $\widehat{d}_{\gamma}$ denote the modified local GEL statistics and upper quantile based on the $\widehat{b}_j\left(K\right)$'s.
Assume that

A2' (i) **A2** holds, (ii) the classes of functions $\mathcal{G}_{1\partial}=\{(\partial g^K(h)/\partial h^T)g_j^K(h),b>0,h\in\mathcal{H}_C\}$ and $\mathcal{G}_{2\partial}=\{(\partial g^K(h)/\partial h^T)g_j^K(h)g_l^K(h),b>0,h\in\mathcal{H}_C\}$ for $j,l=1,...,k$ are Euclidean with integrable envelopes $G_{1\partial}$ and $G_{2\partial}$, (iii) the envelopes $G_3=G_1^3$ and $G_4=G_1^4$ are integrable.

Corollary 1.3 *Under the same assumptions of Theorem 1 and* ***A2'***

$$\begin{aligned}P\left(S_{\gamma}^{\widehat{m}}\left(\widehat{\lambda},h_0\right)\leq c_{\alpha}\right)&=1-\alpha+O\left(\frac{1}{\left(nb^{d_Z}\right)^2}\right),\\ P\left(S_{\gamma}\left(\widehat{\lambda},h_0\right)\leq \widehat{d}_{\gamma}\right)&=G_k\left(c\right)+O\left(\frac{1}{\left(nb^{d_Z}\right)^2}\right).\end{aligned}$$

Corollary 1.3 is important because it shows that using the sample analogs $\widehat{b}_j(K)$ does not change the order of the approximation error of the modified test statistics, a fact first noted by Barndorff-Nielsen and Hall (1988) for Bartlett corrected likelihood ratio statistics.

Remark 1 *It is interesting to note that the standardized multivariate moments $\mu_{jkl}(h_0)$ and $\mu_{jklm}(h_0)$ appearing in the modified statistic $S_\gamma^m\left(\widehat{\lambda}, h_0\right)$ given in (17) (and in the asymptotic expansion of the upper quantile d_γ given in (19)) can be replaced by $\mu_{0jkl}(h_0)$ and $\mu_{0jklm}(h_0)$, where*

$$\begin{aligned}\mu_{0jkl}(h_0) &= E\left(\left[\Omega(z)^{jm}\right]^{1/2} g_m(h_0(Z))\left[\Omega(z)^{kn}\right]^{1/2} g_n(h_0(Z))\times\right.\\ &\qquad \left.\left[\Omega(z)^{lo}\right]^{1/2} g_o(h_0(Z))\,|Z=z\right)\frac{\int K^3(u)\,du}{\left(\int K^2(u)\,du\right)^{3/2}},\\ \mu_{0jklm}(h_0) &= E\left(\left[\Omega(z)^{jn}\right]^{1/2} g_n(h_0(Z))\left[\Omega(z)^{ko}\right]^{1/2} g_o(h_0(Z))\times\right.\\ &\qquad \left.\left[\Omega(z)^{lo}\right]^{1/2} g_o(h_0(Z))\left[\Omega(z)^{mp}\right]^{1/2} g_p(h_0(Z))\,|Z=z\right)\frac{\int K^4(u)\,du}{\int K^2(u)\,du}.\end{aligned} \tag{20}$$

Let b_j and B_j denote the $b_j(K)$ and $B_j(K)$ $(j=1,2,3)$ terms expressed in terms of $\mu_{0jkl}(h_0)$ and $\mu_{0jklm}(h_0)$. Since $\mu_{jkl}(h_0)\to\mu_{0jkl}(h_0)+O(b)$ and $\mu_{jklm}(h_0)\to\mu_{0jklm}(h_0)+O(b)$, the same arguments as those used in the proof of Theorem 1 imply the following corollary:

Corollary 1.4 *Under the same assumption of Theorem 1*

$$P(h_0\in R_\alpha(h)) = 1-\alpha+\frac{1}{nb^{d_Z}}\sum_{j=0}^{3} B_j G_{k+2j}(c_\alpha)+O\left(\frac{b}{(nb^{d_Z})}\right). \tag{21}$$

Corollary 1.4 shows that the coverage error $R_\alpha(h)$ based on the the $B_j's$ is at best of order $o\left(1/nb^{d_Z}\right)$ under the undersmoothing assumption of Theorem 1. Therefore, the approximation errors in Corollaries (1.1), (1.2) and (1.3) have the same order $o\left(1/nb^{d_Z}\right)$.

We now consider the overidentifying restrictions test statistic $S_\gamma\left(\widehat{\lambda},\widehat{h}\right)$ defined in (9). The following convention is used to denote the components of the estimating equations and of the parameters: the indices $r,s,t,...$ run from 1 to $l-k$, the indices $j,k,l,...$ run form 1 to l and the indices $a,b,...$ run from 1 to k. We use the same approach as that developed by Chen and Cui (2007) and Matsushita and Otsu (2013), which relies on an appropriate reparametrization of the local estimating equations $g^K(h)$. To be specific let T_{1jk} denote an $l\times l$ orthogonal matrix such that

$$T_{1jk}\left[\omega(h_0)^{kl}\right]^{1/2} E\frac{1}{b^{d_Z}}\left(\frac{\partial g_l^K(h_0)}{\partial h_a}\right)T_{2ab} = \left[\Lambda_{ab}, O_{k+rb}\right], \tag{22}$$

where T_{2ab} is a $k \times k$ orthogonal matrix, Λ_{ab} is a $k \times k$ nonsingular diagonal matrix, O_{k+rb} is an $(l-k) \times k$ matrix of zeroes and the index k in (22) is fixed and corresponds to the dimension of h, so that the matrix O_{k+rb} is an $(l-k) \times k$ matrix of zeroes. Let

$$U(h) = T_{1jk}\left[\omega(h)^{kl}\right]^{1/2} g_l^K(h) \tag{23}$$

denote the (standardized) reparametrized local estimating equation,

$$\begin{aligned}
\mu_{k+rk+sk+t...}(h) &= E\frac{1}{b^{d}z}\left[U_{k+r}(h)\,U_{k+s}(h)\,U_{k+t}(h)\,...\right],\\
\gamma_{k+rk+sk+t...}^{a_1...a_j,b_1...b_l,...}(h) &= E\frac{1}{b^{d}z}\left[\frac{\partial^j U_{k+r}(h)}{\partial h_{a_1}...\partial h_{a_j}}\frac{\partial^l U_{k+s}(h)}{\partial h_{b_1}...\partial h_{b_l}}U_{k+t}(h)\,...\right],\\
G_{k+rk+sk+t..}^{a_1...a_{m_1},b_1...b_{m_2}}(h) &= \frac{1}{(nb^d)^{1/2}}\sum\left(\frac{\partial^{m_1}U_{ik+r}(h)}{\partial h_{a_1}...\partial h_{a_{m_1}}}\frac{\partial^{m_2}U_{ik+s}(h)}{\partial h_{b_1}...\partial h_{b_{m_2}}}U_{ik+t}(h)\,...-\gamma(h)_{k+rk+sk+t...}^{a_1...a_{m_1},b_1...b_{m_2}}\right),
\end{aligned}$$

where, as in (22), the index k appearing in the multivariate moments $\mu_{k+rk+s...}$ and expected derivatives $\gamma_{k+rk+sk+t..}^{a_1...a_j,b_1...b_l,...}$ is fixed.

Let $\bar{l} = (l-k)+(l-k)((l-k)+1)/2+(l-k)((l-k)+1)((l-k)+2)/3!+k(k+1)/2+(l-k)k(k+1)/2+k(l-k)((l-k)+1)/2+k(l-k)((l-k)+1)((l-k)+2)/3!$ denote the dimension of the vector

$$\begin{aligned}
U_{\bar{l}}(h,K) = \;&[U_{k+1}(h),...,U_l(h),U_{k+1k+1}(h),...,U_{ll}(h),\\
&U_{k+1k+1k+1}(h),...,U_{lll}(h),\\
&G_{k+1}^1(h),...,G_{k+1}^k(h),...,G_l^k(h),G_{k+1}^{1,1}(h),...,\\
&G_{k+l}^{k,k}(h),G_{k+1k+1}^1(h),...,G_{k+lk+l}^k(h),G_{k+1k+1k+1}^1(h),...,\\
&G_{k+1k+1k+1}^k(h),...,G_{k+lk+lk+l}^k(h)]^T
\end{aligned}$$

Assume that:

A2" (i) **A2'**(i)-(ii) hold, (ii) the classes of functions

$$\begin{aligned}
\mathcal{G}_{3\partial}^K &= \left\{(\partial g^K(h)/\partial h^T g_j^K(h))g_k^K(h)g_m^K(h), b>0, h\in\mathcal{H}_C\right\},\\
\mathcal{G}_{1\partial^2}^K &= \left\{(\partial^2 g^K(h)/\partial h^T\partial h_a)g_j^K(h), b>0, h\in\mathcal{H}_C\right\},\\
\mathcal{G}_{2\partial^2}^K &= \left\{\partial^2 g^K(h)/\partial h^T\partial h_a g_j^K(h)g_k^K(h), b>0, h\in\mathcal{H}_C\right\},
\end{aligned}$$

for $(j,l,m=1,...,k; a=1,...,k)$ are Euclidean with integrable envelopes $G_{3\partial}$, $G_{1\partial^2}$ and $G_{2\partial^2}$,

A6' (i) $E\left(\|g(h_0,W)\|^{15}\right)<\infty$, $E\left(\left\|\partial g(h_0,W)/\partial h^T\right\|^5\right)<\infty$, $E\left(\left\|\partial^2 g(h_0,W)/\partial h^T\partial h_a\right\|^5\right)$ $<\infty$,
$E\left(\left\|\partial g(h_0,W)/\partial h^T g_j(h_0,W)\right\|^5\right)<\infty$ and $E\left(\left\|\partial g(h_0,W)/\partial h^T g_j(h_0,W)\,g_k(h_0)\right\|^5\right)<$

∞ for $a = 1, ..., k$ and $j, k = 1, ..., l$, (ii) for any $\theta \in \mathbb{R}^{\bar{l}}$ such that $\|\theta\| = 1$, there exists a partition of $[-1, 1]$, $-1 = u_1 < ... < u_{m(\theta)} = 1$ such that $\theta^T U_{\bar{l}}(h, K)$ is either strictly increasing or decreasing on each interval (u_{m-1}, u_m) for $m = 1, ..., m(\theta)$.

Let

$$\Delta^{ab} = T_{2ac} \Lambda^{bc}, \tag{24}$$

where Λ^{bc} is the matrix inverse of the diagonal matrix defined in (22) and the orthogonal matrix T_{2ac} is also defined in (22). Let $P\left(\chi^2_{l-k} \leq c_\alpha\right) = 1 - \alpha$; the following theorem establishes the second order expansion for the rejection probabilities of $S_\gamma\left(\widehat{\lambda}, \widehat{h}\right)$.

Theorem 2 *Under* ***A1***, ***A2"***, ***A6'*** *and* $nb^{dz+4} \to 0$,

$$P\left(S_\gamma\left(\widehat{\lambda}, \widehat{h}\right) \geq c_\alpha\right) = \alpha + \frac{1}{nb^{dz}} \sum_{j=0}^{3} C_j(K) G_{k+2j}(c_\alpha) + O\left(\frac{1}{(nb^{dz})^2}\right), \tag{25}$$

where

$$\begin{aligned}
C_0(K) &= \frac{1}{72}\left(-c_3(K) + 3c_2(K) - 36c_1(K)\right), \\
C_1(K) &= \frac{1}{24}\left(c_3(K) - 2c_2(K) + 12c_1(K)\right), \\
C_2(K) &= \frac{1}{24}\left(-c_3(K) + c_2(K)\right), \quad C_3(K) = \frac{1}{72} c_3(K),
\end{aligned}$$

$$
\begin{aligned}
c_{1\gamma}(\rho,K) &= \left(\frac{\gamma_4}{4}+2\right)\mu_{k+rk+rk+tk+t}(h_0)+\left(\frac{5\gamma_3}{6}+\frac{59}{36}\right)\mu_{k+rk+rk+t}(h_0)\,\mu_{k+tk+uk+u}(h_0)+\\
&\left(\frac{25}{9}+\frac{5\gamma_3}{3}+\frac{\gamma_3^2}{18}\right)\mu_{k+rk+tk+u}(h_0)\,\mu_{k+rk+tk+u}(h_0)+\\
&\frac{(3+\gamma_3)^2}{36}\mu_{k+rk+sk+s}(h_0)\,\mu_{k+rk+tk+t}(h_0)+[2]\,\Delta^{ab}\gamma^{a}_{k+rk+sb}(h_0)+\\
&[2]\,\frac{\gamma_3}{2}\Delta^{ab}\mu_{k+rk+sb}(h_0)\,\gamma^{a}_{k+uk+u}(h_0)-\frac{1}{2}\Delta^{ab}\mu_{k+rk+sb}(h_0)\,\gamma^{a}_{k+rk+t}(h_0)\\
&-\frac{1}{2}\Delta^{ab}\Delta^{bc}\Delta^{df}\Delta^{ef}\gamma^{ad}_{k+r}(h_0)\,\gamma^{be}_{k+r}(h_0)-\Delta^{ac}\Delta^{bc}\gamma^{ab}_{k+rk+r}(h_0)+\\
&\frac{1}{2}\Delta^{ac}\Delta^{bc}\Delta^{df}\Delta^{ef}\gamma^{ab}_{k+r}(h_0)\,\gamma^{cd}_{k+r}(h_0)+\Delta^{ab}\Delta^{cd}\gamma^{ab}_{k+r}(h_0)\,\gamma^{cd}_{k+r}(h_0)\\
&-\left(\frac{\gamma_3+3}{3}\right)\Delta^{ac}\Delta^{bc}\mu_{k+rk+tk+t}(h_0)\,\gamma^{ab}_{k+r}(h_0)-\\
&[2]\,\Delta^{ab}\Delta^{ce}\Delta^{de}\gamma^{a}_{k+rb}(h_0)\,\gamma^{cd}_{k+s}(h_0)\,,\\
c_{2\gamma}(\rho,K) &= [3]\Bigg[(\gamma_4+6)\,\tau_4+\frac{(\gamma_3^2+5\gamma_3+6)}{6}\tau_3+\frac{2}{9}\left(2\gamma_3^2+21\gamma_3+34\right)\tau_4-\\
&[4]\,(\gamma_3+2)\,\mu_{k+tk+sk+t}(h_0)\left(-\Delta^{ab}\gamma^{a}_{k+ub}(h_0)+\Delta^{ac}\Delta^{bc}\gamma^{ab}_{k+u}(h_0)\right)\\
&\frac{(\gamma_3+2)}{2}\left[\Delta^{ac}\Delta^{bc}\gamma^{a}_{k+rk+s}(h_0)\,\gamma^{b}_{k+tk+u}(h_0)+\Delta^{ab}\gamma^{a,b}_{k+rk+sk+tk+u}(h_0)\right]+\\
&\frac{(\gamma_3+2)}{3}\mu_{k+rk+s}(h_0)\,\Delta^{ab}\left(\gamma^{ab}_{k+tk+u}(h_0)+\gamma^{a,b}_{k+tk+u}(h_0)\right)\Bigg]\,\delta_{k+rk+s}\delta_{k+tk+u},\\
c_{3\gamma}(\rho,K) &= [15]\,(2+\gamma_3)^2\,\mu_{k+rk+sk+t}(h_0)\,\mu_{k+uk+vk+w}(h_0)\,\delta_{k+rk+s}\delta_{k+tk+u}\delta_{k+vk+w}.
\end{aligned}
$$

Theorem 2 shows that, as with the exactly identified case, the terms $c_2(K)$ and $c_3(K)$ are 0 for EL, which implies that the EL ratio statistic is Bartlett correctable. Similarly, one can consider the modified statistic $S^m_\gamma\left(\widehat{\lambda},\widehat{h}\right)$ analog to that given in (17). Note that the modified statistic $S^m_\gamma\left(\widehat{\lambda},\widehat{h}\right)$ could also be computed using the original local estimating equations, since, as shown in Theorem 2, the reparametrization (23) is useful only for the EL ratio because it implies its Bartlett correctability.

Let $BC(K)=1+c_{1EL}(K)/\left(knb^{dz}\right)$ denote the Bartlett correction; then

Corollary 2.1 *Under the same assumptions of Theorem 2*

$$
\begin{aligned}
P\left(\frac{S_{EL}\left(\widehat{\lambda},\widehat{h}\right)}{BC(K)}\geq c_\alpha\right) &= \alpha+O\left(\frac{1}{\left(nb^{dz}\right)^2}\right) \qquad (26)\\
P\left(S^m_\gamma\left(\widehat{\lambda},\widehat{h}\right)\geq c_\alpha\right) &= \alpha+O\left(\frac{1}{\left(nb^{dz}\right)^2}\right).
\end{aligned}
$$

Finally, (25) can be used to obtain the asymptotic expansion of the upper quantile of the distribution of $S_\gamma\left(\widehat{\lambda},\widehat{h}\right)$, that is for

$$d_\gamma = c\left(1+2\sum_{j=1}^{3}\left(\sum_{l=j}^{3}C_j\left(K\right)\right)\right)\frac{c^{j-1}}{nb^{dz}\Gamma_j}, \tag{27}$$

one has that

$$P\left(S_\gamma\left(\widehat{\lambda},\widehat{h}\right)\geq d_\gamma\right)=1-G_{l-k}\left(c\right)+O\left(\frac{1}{\left(nb^{dz}\right)^2}\right). \tag{28}$$

.

Consistent estimators for the terms $C_j\left(K\right)$ $\left(j=0,1,2,3\right)$ can be obtained using their sample analogs, as in the case of Corollary (1.3). However, given their complex form and the lengthy number of calculations involved, it is useful to have an alternative approach that delivers the same level of second order asymptotic accuracy. Brown and Newey's (2002) efficient bootstrap (EB) provides such an alternative and is discussed in the next subsection.

3.1 Efficient local bootstrap

EB differs from the standard bootstrap in that the observations are resampled according to a distribution that is constrained to satisfy a given constraint, which in the context of this paper is given by $E\left[g\left(h_0\left(z\right),W\right)|Z=z\right]f(z)=0$. Let

$$\widehat{\pi}_i\left(z\right)=\frac{\gamma_1\left(\widehat{v}_i\left(z\right)\right)}{\sum_{j=1}^{n}\gamma_1\left(\widehat{v}_j\left(z\right)\right)} \tag{29}$$

denote the implied local GEL probabilities [3] and let $\left\{W_i^{*T},Z_i^{*T}\right]\}_{i=1}^{n}$ denote a sample drawn from the resulting constrained distribution $P^*\left(W=W_i\right)=\widehat{\pi}_i\left(z\right)$. Let $g\left(h,W_i^*\right)=g_i^*\left(h\right)$ denote the bootstrap analog of $g_i\left(h\right)$ and note that

$$\begin{aligned} E^*\left[\frac{g^{*K}\left(h\right)}{b^{dz}}\right] &= \sum_{i=1}^{n}\widehat{\pi}_i\left(z\right)\frac{g_i^K\left(h\right)}{b^{dz}} = \\ &\frac{1}{nb^{dz}}\sum_{i=1}^{n}g_i^K\left(h\right)-\frac{1}{nb^{dz}}\sum_{i=1}^{n}g_i^K\left(\widehat{h}\right)+o_p\left(1\right), \end{aligned}$$

which shows that the resampled local estimating equation does not need recentering - a key point to obtain second order asymptotic refinements (see for example Hall and Horowitz (1996)).

Let

$$S_\gamma^*\left(\widehat{\lambda}^*,\widehat{h}^*\right)=2nb^{dz}\left(\widehat{\Gamma}_\gamma\left(\widehat{\lambda}^*,\widehat{h}^*\right)-\widehat{\Gamma}_\gamma\left(0,\widehat{h}^*\right)\right) \tag{30}$$

denote the bootstrap analogs of (9), where $\widehat{\lambda}^*$ and $\widehat{h}^*$ are the bootstrap analogs of $\widehat{\lambda}$ and $\widehat{h}$, and note that, because of the automatic recentering, the probability limit of $\widehat{\lambda}^*$ is 0 and not $\widehat{\lambda}$ (see

[3]Note that (29) can also be computed as $\widetilde{\pi}_i(z)=\gamma_1(\widehat{\lambda}^T g_i^K(\widetilde{h}))/\sum_{i=1}^{n}\gamma_1(\widehat{\lambda}^T g_i^K(\widetilde{h}))$, where $\widetilde{h}$ is any consistent estimator for h_0.

the proof of Theorem 3 for further details). Let c_{α}^{*} denote the $1-\alpha$ quantile from the empirical distribution of $S_{\gamma}^{*}\left(\widehat{\lambda}^{*}, \widehat{h}^{*}\right)$ obtained by computing (30) using B $\{W_i^*, Z_i^*\}_{i=1}^{n}$ samples.

Theorem 3 *Under the same assumptions of Theorem 2,*

$$P\left(S_{\gamma}\left(\widehat{\lambda}, \widehat{h}\right) \geq c_{\alpha}^{*}\right) = \alpha + O\left(\frac{1}{\left(nb^{d_Z}\right)^{2}}\right).$$

Theorem 3 shows that bootstrapping $S_{\gamma}\left(\widehat{\lambda}, \widehat{h}\right)$ delivers the same order of accurate inference as that delivered by the modified statistics $S_{\gamma}^{m}\left(\widehat{\lambda}, \widehat{h}\right)$.

4 Examples and Monte Carlo results

This section illustrates both analytically and numerically the results of the previous section with the two examples 1 and 2 and contains a discussion on the bandwidth selection.

4.1 Nonparametric binary choice model

For the nonparametric binary choice model (3) estimation of h_0 at the point z can based on the random sample $\left(W_i = \left[Y_i, Z_i^T\right]\right)_{i=1}^{n}$ using the local estimating equation

$$g_i^K\left(h\left(z\right)\right) = \left(Y_i - F_{\varepsilon}\left(h\left(z\right)\right)\right) K\left(\frac{Z_i - z}{b}\right).$$

Let $f_{\varepsilon}\left(h\right) = dF_{\varepsilon}\left(h\right)/dh$, $f_{\varepsilon}'\left(h\right) = d^2F_{\varepsilon}\left(h\right)/dh^2$, $f_{\varepsilon}''\left(h\right) = d^3F_{\varepsilon}\left(h\right)/dh^3$ and so on, and note that $\Omega\left(z\right) = E\left[F_{\varepsilon}\left(h_0\left(Z\right)\right)\left(1 - F_{\varepsilon}\left(h_0\left(Z\right)\right)\right)|Z = z\right]$ and $G\left(z\right) = -E\left[f_{\varepsilon}\left(h_0\left(Z\right)\right)|Z = z\right]$. Assume that:

NBC1 (i) $P\left(h\left(z\right) \neq h_0\left(z\right)\right) > 0$ for all $h \neq h_0 \in \mathcal{H}_C$ and all $z \in \mathcal{Z}$, (ii) **A1**(ii)-(iii) hold,

NBC2 (i) $f_{\varepsilon}\left(h\right)$, $f_{\varepsilon}'\left(h\right)$, $f_{\varepsilon}''\left(h\right)$ and $f_{\varepsilon}'''\left(h\right)$ exist and are continuous for each $h \in \mathcal{H}_C$ *a.s.*, (ii) $E\left[\sup_{z\in\mathcal{Z}}, \sup_{h\in\mathcal{H}_C}\left|f_{\varepsilon}\left(h\left(z\right)\right)\right|^{\beta}\right] < \infty$ for $\beta = 1, 2, 3$, $E\left[\sup_{z\in\mathcal{Z}}, \sup_{h\in\mathcal{H}_C}\left|f_{\varepsilon}'\left(h\left(z\right)\right)\right|\right] < \infty$, $E\left[\sup_{z\in\mathcal{Z}}, \sup_{h\in\mathcal{H}_C}\left|f_{\varepsilon}''\left(h\left(z\right)\right)\right|\right] < \infty$ and $E\left[\sup_{z\in\mathcal{Z}}, \sup_{h\in\mathcal{H}_C}\left|f_{\varepsilon}'\left(h\left(z\right)\right) f_{\varepsilon}\left(h\left(z\right)\right)\right|\right] < \infty$,

NBC3 for all $z \in \mathcal{Z}$ (i) $E\left|Y - F_{\varepsilon}\left(h\left(z\right)\right)\right|^{\delta} < \infty$ for all $h \in \mathcal{H}_C$ and some $\delta > 2$, (ii) $\Omega\left(z\right) > 0$,

NBC4 **A4'** and **A5** hold.

Assumption **NBC1**(i) implies the identification condition **A1**(i) by the monotonicity of $F_{\varepsilon}\left(h\right)$ in h. **NBC2**(i) requires F_{ε} to be sufficiently smooth; combined with **NBC2**(ii) implies that the Euclidean and envelope assumptions in **A2**(ii) and **A2'**(iii)-(iv) are satisfied, and it can be shown that as $\left(nb^{dz}\right)^{1/2} \to \infty$, the asymptotic distribution of the estimator $\widehat{h}$ is

$$\left(nb^{dz}\right)^{1/2}\left(\widehat{h}(z) - h_0(z)\right) \xrightarrow{d} N\left(B(z), \frac{E\left[F_\varepsilon(h_0(Z))(1 - F_\varepsilon(h_0(Z)))|Z = z\right]}{f(z)\left(E\left[f_\varepsilon(h_0(Z))|Z = z\right]\right)^2}\int K^2(u)\,du\right),$$

where

$$B(z) = -\frac{1}{2f(z)} trace\left[\left(h_0''(z) f(z) + 2h_0'(z) f'(z)^T\right)\int uu^T K(u)\,du\right].$$

For inference, we assume that:

NBC5 (i) $E\left[|Y - F_\varepsilon(h_0)|^{15}\right] < \infty$, (ii) **A6**(ii) holds;

then, the terms required in the Edgeworth expansion of $S_\gamma\left(\widehat{\lambda}, h_0\right)$ (15) and its modified version $S_\gamma^m\left(\widehat{\lambda}, h_0\right)$ (17) are

$$\omega_{jl}(h_0) := \omega_2(h_0) = E\left[\frac{(Y - F_\varepsilon(h_0(z)))^2 K\left(\frac{Z-z}{b}\right)^2}{b}\right] \to$$
$$f(z) E\left[F_\varepsilon(h_0(Z))(1 - F_\varepsilon(h_0(Z)))|Z = z\right]\int K^2(u)\,du + O(b),$$
$$\mu_{j_1\ldots j_k}(h_0) := \mu_k(h_0) = E\left[\frac{1}{b}\left(\frac{(Y - F_\varepsilon(h_0(z))) K\left(\frac{Z-z}{b}\right)}{\omega_2(h)^{1/2}}\right)^k\right] \quad k = 3, 4,$$
$$\mu_3(h_0) \to \frac{1}{f(z)^{1/2}}\frac{E\left[F_\varepsilon(h_0(Z))(1 - F_\varepsilon(h_0(Z)))(1 - 2F_\varepsilon(h_0(Z)))|Z = z\right]}{E\left[F_\varepsilon(h_0(Z))(1 - F_\varepsilon(h_0(Z)))|Z = z\right]^{3/2}} \times$$
$$\frac{\int K^3(u)\,du}{\left(\int K^2(u)\,du\right)^{3/2}} + O(b),$$
$$\mu_4(h_0) \to \frac{1}{f(z)}\frac{E\left[F_\varepsilon(h_0(Z))(1 - F_\varepsilon(h_0(Z)))\right.}{E\left[F_\varepsilon(h_0(Z))(1 - F_\varepsilon(h_0(Z)))|Z = z\right]^2} \times$$
$$\left.\left(3F_\varepsilon(h_0(Z))^2 - 3F_\varepsilon(h_0(Z)) + 1\right)|Z = z\right]\frac{\int K^4(u)\,du}{\left(\int K^2(u)\,du\right)^2} + O(b).$$

4.2 Smooth coefficients model with endogenous covariates

For the smooth coefficients model with endogenous covariates (4) estimation of h_0 can based on the random sample $\left(W_i = \left[Y_i, X_i^T, Z_i^T\right]^T\right)_{i=1}^n$ using the local estimating equations

$$g_i^K(h(z)) = r(Z_i)\left(Y_i - X_i^T h(z)\right) K\left(\frac{Z_i - z}{b}\right).$$

Let $\Omega(z) = E\left[\left(r(Z) r(Z)^T \varepsilon^2\right)|Z = z\right]$ and $G(z) = E\left(r(Z) X^T|Z = z\right)$, and assume that:

SCM1 **A1**(ii)-(iii) hold,

SCM2 (i) $E\left(\left\|r(Z)X^T\right\|^\beta\right)<\infty$ for $\beta=1,2,3$, (ii) $E\left\|r(Z)\left(Y-X^T h(Z)\right)\right\|^\delta<\infty$ for all $h\in\mathcal{H}_C$ and some $\delta>2$, (iii) $\Omega(z)$ is positive definite, (iv) $rank(G(z))=k$,

SCM3 **A4'** and **A5** hold.

Assumption **SCM2**(i) ensures that the Euclidean and envelope assumptions in **A2**(ii) and **A2'**(iii)-(iv) are satisfied; **SCM2**(iv) implies the identification condition **A1**(i) as in standard linear instrumental variables models. It is then possible to show that as $\left(nb^{d_Z}\right)^{1/2}\to\infty$ the asymptotic distribution of the estimator $\widehat{h}(z)$ is

$$\left(nb^{d_Z}\right)^{1/2}\left(\widehat{h}(z)-h_0(z)\right)\xrightarrow{d}N\left(B(z),\Sigma(z)\right),$$

where $B(z)$ is as given in the previous example and

$$\Sigma(z)=\frac{1}{f(z)}\left(E\left(r(Z)X^T|Z=z\right)^T E\left[\left(r(Z)\varepsilon\right)^{\otimes 2}|Z=z\right]^{-1}E\left(r(Z)X^T|Z=z\right)\right)^{-1}\int K^2(u)\,du.$$

For inference, we assume that:

SCM4 (i) $E\left[\left\|r(Z)\varepsilon\right\|^{15}\right]<\infty$, $E\left(\left\|r(Z)X^T\right\|^5\right)<\infty$, $E\left(\left\|r(Z)X^T\varepsilon\right\|^5\right)<\infty$ and $E\left(\left\|r(Z)X^T\varepsilon^2\right\|^5\right)$ $<\infty$, (ii) **A6'**(ii) holds;

then, by Theorem 2 and Corollary 2.1 the terms required in the Edgeworth expansion of

$S_\gamma\left(\widehat{\lambda}, h_0\right)$ (15) and its modified version $S^m_\gamma\left(\widehat{\lambda}, h_0\right)$ (17) are

$$\omega_{jl}\left(h_0\right) = E\left[\frac{1}{b^{d_Z}}\left(r_j\left(Z\right) r_l\left(Z\right) \varepsilon^2\right) K\left(\frac{Z-z}{b}\right)^2\right] \rightarrow$$
$$f\left(z\right) E\left[r_j\left(Z\right) r_l\left(Z\right) \varepsilon^2 | Z = z\right] \int K^2\left(u\right) du + O\left(b\right),$$
$$\mu_{jkl}\left(h_0\right) = E\left[\frac{1}{b^{d_Z}} T_{1jj'} \omega^{j'm}\left(h_0\right)^{1/2} r_m\left(Z\right) T_{1kk'} \omega^{k'o}\left(h_0\right)^{1/2} r_o\left(Z\right) T_1 \omega^{lp}\left(h_0\right)^{1/2} r_p\left(Z\right) \varepsilon^3 K\left(\frac{Z-z}{b}\right)^3\right] \rightarrow$$
$$\frac{1}{f\left(z\right)^{1/2}} E\left[\omega^{jm}\left(h_0\right)^{1/2} r_m\left(Z\right) \omega^{ko}\left(h_0\right)^{1/2} r_o\left(Z\right) \omega^{lp}\left(h_0\right)^{1/2} r_p\left(Z\right) \varepsilon^3 | Z = z\right] \times$$
$$\frac{\int K^3\left(u\right) du}{\left(\int K^2\left(u\right) du\right)^{3/2}} + O\left(b\right),$$
$$\mu_{jklm}\left(h_0\right) = E\left[\frac{1}{b^{d_Z}} T_{1jj'} \omega^{j'n}\left(h_0\right)^{1/2} r_n\left(Z\right) T_{1kk'} \omega^{k'o}\left(h_0\right)^{1/2} r_o\left(Z\right) T_{1ll'} \omega^{l'p}\left(h_0\right)^{1/2} r_p\left(Z\right) \times\right.$$
$$\left. T_{1mm'} \omega^{m'q}\left(h_0\right)^{1/2} r_q\left(Z\right) \varepsilon^4 K\left(\frac{Z-z}{b}\right)^4\right] \rightarrow$$
$$\frac{1}{f\left(z\right)} E\left[T_{1jj'} \omega^{j'n}\left(h_0\right)^{1/2} r_n\left(Z\right) T_{1kk'} \omega^{k'p}\left(h_0\right)^{1/2} r_p\left(Z\right) T_{1ll'} \omega^{l'p}\left(h_0\right)^{1/2} r_p\left(Z\right) \times\right.$$
$$\left. T_{1mm'} \omega^{m'q}\left(h_0\right)^{1/2} r_q\left(Z\right) \varepsilon^4 | Z = z\right] \frac{\int K^4\left(u\right) du}{\left(\int K^2\left(u\right) du\right)^2} + O\left(b\right),$$

$$\gamma^a_{k+r}\left(h_0\right) = E\left[-\frac{1}{b^{d_Z}} T_{1k+rj} \omega^{jm}\left(h_0\right)^{1/2} r_m\left(Z\right) X_a K\left(\frac{Z-z}{b}\right)\right] \rightarrow$$
$$f\left(z\right)^{1/2} E\left[-T_{1k+rj} \omega^{jm}\left(h_0\right)^{1/2} r_m\left(Z\right) X_a | Z = z\right] \frac{1}{\left(\int K^2\left(u\right) du\right)^{1/2}} + O\left(b\right),$$
$$\gamma^a_{k+rk+s}\left(h_0\right) = O\left(b^2\right),$$
$$\gamma^{a,b}_{k+rk+s}\left(h_0\right) = E\left[\frac{1}{b^{d_Z}} T_{1k+rj} \omega^{jm}\left(h_0\right)^{1/2} r_m\left(Z\right) X_a T_{1k+sk} \omega^{ko}\left(h_0\right)^{1/2} r_o\left(Z\right) X_b K\left(\frac{Z-z}{b}\right)^2\right] \rightarrow$$
$$E\left[T_{1k+rj} \omega^{jk}\left(h_0\right)^{1/2} r_k\left(Z\right) X_a T_{1k+sl} \omega^{lm}\left(h_0\right)^{1/2} r_m\left(Z\right) X_b | Z = z\right] + O\left(b\right).$$

4.3 Bandwidth selection

The discussion in Section (3) shows that the optimal (minimizing the asymptotic mean squared error) bandwidth $b^{d_Z}_*$ for $\widehat{h}\left(z\right)$ is of order $O\left(n^{-\frac{1}{d_Z+4}}\right)$, that is the standard nonparametric rate, and data driven methods, such as least squares cross validation (see for example Li and Racine (2004) for some optimality properties of such procedure) could be used to automatically select $b^{d_Z}_*$. On the other hand, the results of Theorems 1, 2 and 3 require undersmoothing, hence least squares cross validation or other bandwidth selection methods cannot be used directly to automatically choose the bandwidth. For inference, one possibility is to develop an Edgeworth expansion for the coverage errors given in (15) and (25) that explicitly depends on b^{d_Z}, as for example in Chen (1996) and Chen and Qin (2000) and then choose the bandwidth that minimizes the resulting coverage errors. However, the statistical models considered in this paper involve potentially many more estimating equations than those considered by Chen (1996) and Chen and Qin (2000) with the bandwidth based on a possibly d_Z dimensional covariate, which makes this possibility very difficult in practice - see also Otsu et al. (2015) on this point. Another possibility is to follow the ad-hoc procedure suggested by Otsu et al.'s (2015) and use least squares cross validation to estimate the bandwidth and then multiply the resulting bandwidth by a power of the sample size that is consistent with undersmoothing. In this paper we consider another method, that is similar to the ad-hoc cross validation method of Otsu et al.'s (2015) but is less computationally intensive. Specifically, we consider a simple sample splitting procedure, which consists of computing for a random subset[4], say S_τ with $0 < \tau < 1$ - the training set- and a pilot bandwidth b_p[5]

$$\widehat{h}_p(z) = \arg\min_{h \in \mathcal{H}_C} \frac{1}{n_\tau b_p^{d_Z}} \sum_{i \in S_\tau} \gamma\left(\widehat{\lambda}(z)^T g_i^{K_{b_p}}(h)\right),$$

where $K_{b_p} := K((Z_i - z)/b_p)$, $\widehat{\lambda}(z)$ is computed using the same pilot bandwidth b_p, and then using the remaining part of the sample $S_{1-\tau}$- the validation set- to select the bandwidth as

$$\widehat{b} = \arg\min_{b \in B} \frac{1}{n_{1-\tau} b^{d_Z}} \sum_{i \in S_{1-\tau}} \gamma\left(\widehat{\lambda}(z)^T g_i^K\left(\widehat{h}_p\right)\right), \tag{31}$$

where B is a grid of possible values of b. In the simulations results of next section we choose $\tau = 80\%$, which is a commonly used percentage for the sample sizes considered; in the Additional tables section available in the Online supplement we have also considered $\tau = 70\%$ - see Tables 7-8. Finally, as in Otsu et al. (2015) $\widehat{b}$ is multiplied by a power of the sample size that is consistent with undersmoothing.

[4]It is important to note that in the context of the models considered in this paper, the way the sample is split does not seem to affect the finite sample properties of the proposed test - see Tables 9-12 in the "Additional tables" section available in the Online supplement.

[5]One simple practical way to choose b_p is to use the standard deviation of the first principal component (PC) in the PC analysis of the $n \times d_Z$ matrix of conditioning variables $\mathbf{Z}$, in which case b_p corresponds to the square root of the largest eigenvalue of the sample covariance matrix of $\mathbf{Z}$; as a robustness check, one could also choose some values within the interval $\pm 20\%$ of the standard deviation of the first principal component to verify the stability of the chosen bandwidth - see Tables 13-17 in the "Additional tables" section available in the Online supplement.

4.4 Monte Carlo results

For the nonparametric binary choice model (3) we consider the following logit specification for F_ε

$$P(Y_i = 1|Z_i = z) = \frac{\exp(\sin(\pi z))}{1 + \exp(\sin(\pi z))},$$

with $Z_i \sim U(-2, 2)$. We use the second order bi-weight kernel

$$K(u) = \frac{15}{28}\left(1 - \frac{u^2}{2}\right)^2 I(|u| \leq 2), \tag{32}$$

where $I(\cdot)$ is the indicator function and generate 5000 replications with sample sizes $n = 100$, $n = 400$ and $n = 1000$. Note that for this kernel $\int K^2(u)\,du \simeq 0.535$, $\int K^3(u)\,du \simeq 0.158$ and $\int K^4(u)\,du \simeq 0.112$. We consider constructing confidence intervals for h_0 at the points $z = 0$ and $z = 1$ with nominal coverage 0.95, using three local GEL statistics, namely the local empirical likelihood $S_{EL}\left(\widehat{\lambda}, h_0\right)$, the local exponential tilting $S_{ET}\left(\widehat{\lambda}, h_0\right)$ and the local continuous updating $S_{CU}\left(\widehat{\lambda}, h_0\right)$, defined, respectively, as

$$\begin{aligned}
S_{EL}\left(\widehat{\lambda}, h_0\right) &= 2\sum_{i=1}^{n} \log\left(1 - \widehat{\lambda}(z)\, g_i^K(h_0(z))\right),\\
S_{ET}\left(\widehat{\lambda}, h_0\right) &= 2\sum_{i=1}^{n} \left(1 - \exp\left(\widehat{\lambda}(z)\, g_i^K(h_0(z))\right)\right),\\
S_{CU}\left(\widehat{\lambda}, h_0\right) &= -\frac{1}{2}\sum_{i=1}^{n} \left(1 + \widehat{\lambda}(z)\, g_i^K(h_0(z))\right)^2.
\end{aligned}$$

Tables 1-2 report the finite sample coverage and average length of the confidence intervals based on three local GEL statistics and the three modified statistics $S^m_{EL}\left(\widehat{\lambda}, h_0\right)$, $S^m_{ET}\left(\widehat{\lambda}, h_0\right)$ and $S^m_{CU}\left(\widehat{\lambda}, h_0\right)$ defined in (17), using three different choices of the bandwidth: $b = \widehat{b}n^{-a}$ with $a = [3/7, 1/3, 1/4]$, where $\widehat{b}$ is chosen using the sample splitting procedure described in Section 4.3, with the pilot bandwidth $b_p = 1.15$ - the standard deviation of Z_i and the grid B consisting of points equally spaced by 0.1 on the interval 0.7 and 1.4.

Tables 1- 2 approx here

The results of Tables 1-2 can be summarized as follows: first the three local original GEL statistics are all characterized by a certain degree of undercoverage, which is reduced as the sample size increases. The modified local statistics improve the coverage accuracy of the original ones, with those based on the Bartlett correction delivering a slightly larger improvement compared to the other two statistics based on (17). Tables 1-2 show also that smaller bandwidths increase the coverage whereas larger bandwidth decreases the length of the confidence intervals, which is to be expected since it is a simple reflection of the bias-variance trade-off of kernel estimators, where the bias affects the coverage and the variance affects the length. Finally, between the three local test statistics, the local EL statistic has the better coverage across the three different bandwidths.

For the smooth coefficient model (4) we consider the following specification

$$\begin{aligned} Y_i &= X_{1i}\exp(-0.5Z_i) + X_{2i}\cos(\pi Z_i/2) + \varepsilon_i \\ X_{1i} &= Z_i + u_i \end{aligned}$$

where $Z_i \sim U(-2,2)$, $X_{1i} \sim N(0,1)$ and

$$\begin{bmatrix} \varepsilon_i \\ u_i \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right).$$

The vector of instruments $r(Z)$ is specified as $r(Z) = [Z, Z^2, X_2, X_2^2,]^T$, that is the degree of overidentification is 2, and and generate 5000 replications with sample sizes $n = 100$, $n = 400$ and $n = 1000$. We consider testing for the correct specification of (4) at the points $z = -0.5$ and $z = 1$ at a 0.05 significance level, using the same bi-weight kernel function (32) and the same three different choices of the bandwidth as those used in the previous example with the pilot bandwidth b_p set to 0.9 and the grid B consisting of points equally spaced by 0.1 on the interval 0.5 and 1.3. The test statistics are

$$\begin{aligned} S_{EL}\left(\widehat{\lambda}, \widehat{h}\right) &= 2\sum_{i=1}^{n} \log\left(1 - \widehat{\lambda}(z)^T g_i^K\left(\widehat{h}(z)\right)\right), \\ S_{ET}\left(\widehat{\lambda}, \widehat{h}\right) &= 2\sum_{i=1}^{n}\left(1 - \exp\left(\widehat{\lambda}(z)^T g_i^K\left(\widehat{h}(z)\right)\right)\right), \\ S_{CU}\left(\widehat{\lambda}, \widehat{h}\right) &= -\frac{1}{2}\sum_{i=1}^{n}\left(1 + \widehat{\lambda}(z)^T g_i^K\left(\widehat{h}(z)\right)\right)^2, \end{aligned}$$

where the estimators $\widehat{\lambda}$ and $\widehat{h}$ appearing in each test statistic are based on the corresponding local GEL objective function, and their EB analogs $S_\gamma^*\left(\widehat{\lambda}^*, \widehat{h}^*\right)$ for $\gamma =$ "EL", "ET" and "CU", which are calculated using the local implied GEL probabilities (29) based on the local ET estimator, that is $\widehat{\pi}_i(z) = \sum_{i=1}^n \exp\left(\widehat{\lambda}(z)^T g_i^K\left(\widehat{h}(z)\right)\right) / \exp\left(\widehat{\lambda}(z)^T g_i^K\left(\widehat{h}(z)\right)\right)$. Tables 3 and 4 report the finite sample rejection probabilities (finite sample sizes) of the three local GEL statistics and their EB analogs, which are based on $B = 500$ replications.

Tables 3-4 approx here

The results of Tables 3 and 4 show that the three local GEL statistics are all characterized by rejection probabilities that are larger than the nominal size and that their EB version provides an effective method to improve the rejection probabilities. As with the previous example, the choice of bandwidth has some bearings on the finite sample properties of the test statistics, which however becomes less evident as the sample size increases.

5 Empirical application

We consider a varying coefficients specification of the so-called Mincer's (1956) equation, a cornerstone of modern empirical economics, which estimates the returns of schooling by analyzing the relationship between earnings, years of schooling and work experience. We use the Young Men Cohort of the National Longitudinal Survey (NLS-Y) data originally used by Griliches (1976), available in the R package `gmm` (`wage`) and also in `https://lachlandeer.github.io/hayashir/`. We first specify the model as an exactly identified instrumental variables varying coefficients model, since it has long being acknowledged that years of schooling and possibly work experience might be endogenous, that is they might be correlated with the unobservable errors. The model is

$$g_i(h(Z_i)) = [1, Z_i^T]^T (Y_i - h_1(Z_i) - X_{1i} h_2(Z_i) - X_{2i} h_3(Z_i)), \tag{33}$$

where Y_i is the log-wage, X_{1i} is years of education, X_{2i} is work experience and the instruments $Z_i = [Z_{1i}, Z_{2i}]^T$ are, respectively, IQ_i, an IQ test scores used as a measure of ability, and AGE_i.

Figure 1 approx. here

Figure 1 shows an important feature of the wage equation, which is well captured by the varying coefficients specification (33), that is that there are some nonlinearities in the smooth coefficients, a fact noted by Trostel (2005) among many others. As expected, all three varying coefficients are increasing in both IQ and AGE, with the first estimated coefficient representing the effects of AGE and IQ on the log wages. Next, we consider inference for the three varying coefficients model. The bandwidth is chosen using the sample splitting technique of Section 4.3, with pilot bandwidth $b_p = 1.026$ and $n^{-1/3}$, which corresponds to the square root of the largest eigenvalue of the sample covariance matrix of Z_i. Figures 2-4 show the 95% confidence regions for h_1, h_2 and h_3 using the EL and CU statistics $S_{EL}(\widehat{\lambda}, h)$ and $S_{CU}(\widehat{\lambda}, h)$ and their modified versions $S_{EL}^{\widehat{m}}(\widehat{\lambda}, h)$ and $S_{CU}^{\widehat{m}}(\widehat{\lambda}, h)$.

Figures 2-4 approx. here

Figures 2-4 show two interesting features of the confidence regions: first, those based on EL are data adaptive in the sense that their shape is determined by the data (hence are not necessarily ellipsoidal). Second the modified statistics are characterized by smaller confidence regions indicating a better accuracy. Finally, we consider two overidentified specifications of

(33). Specifically, we consider the vectors of instruments $r_1(Z_i) = [1, Z_i^T, Z_{1i}^2, Z_{2i}^2]^T$ and $r_2(Z_i) = [1, Z_i^T, Z_{1i}^2, Z_{2i}^2, Z_{1i}^3, Z_{2i}^3]^T$, which give the degrees of overidentification $l-k$ of 2 and 4, respectively. We use the test statistics $S_{EL}(\widehat{\lambda}, \widehat{h})$ and $S_{CU}(\widehat{\lambda}, \widehat{h})$ to test for the correct local specification of the two overidentified nonparametric models evaluated at the 25%, 50% and 75% quantiles of iq and age, that is over the grid [95.25, 104, 103.86]×[30, 33, 36]. The rejection probabilities are computed using the EB statistics $S_{EL}^*(\widehat{\lambda}^*, \widehat{h}^*)$ and $S_{CU}^*(\widehat{\lambda}^*, \widehat{h}^*)$ with $B = 500$.

Tables 5-6 approx.here

Tables 5 and 6 show that both overidentified specifications are rejected at the traditional 5% significance level, although for the $R_1(Z_i)$ specification with $IQ = 95.25$ and $AGE = [30, 33, 36]$ it cannot be rejected at the 10% significance level. Overall, the results of Tables 5-6 imply that adding additional instruments does not improve the original (exactly identified) specification, suggesting that the additional instruments are either irrelevant (that is they have no additional explanatory power or possibly a weak association with the endogenous covariates they are intended to instruments for) and/or not valid (that is they might have a direct effect on the dependent variable). In either cases Tables 5-6 suggest that the chosen exactly identified specification seems adequate. To check the robustness of these conclusions, we have calculated the same statistics for two alternative pilot bandwidths, $b_p^{(1)} = 0.85$ and $b_p^{(2)} = 1.25$. The results, which are reported in Tables 16-19 in the "Robustness checks for the empirical application " section available in the Online Appendix, provide additional evidence supporting the overall conclusions implied by Tables 5-6.

6 Conclusions

This paper considers inference for nonparametric estimating equations models. The paper proposes a general class of local GEL test statistics that can be used to make inferences about the unknown infinite dimensional parameter. Under a standard undersmoothing assumption, the paper shows that the proposed test statistics can be modified with a Bartlett correction (for the local empirical likelihood) or Bartlett-type corrections to improve their finite sample accuracy. The paper also considers a local version of the efficient bootstrap of Brown and Newey (2002) that can be used to obtain asymptotic refinements to the proposed test statistics. The results of the paper are illustrated both numerically and analytically with two examples and a real data application. These results are encouraging and suggest that the proposed methods improve the finite sample performances of the test statistics and they are useful in practice.

References

Barndorff-Nielsen, O. E. and Hall, P.: 1988, On the level-error after Bartlett adjustment of the likelihood ratio statistic, *Biometrika* **75**, 374–378.

Bhattacharya, R. and Ghosh, J.: 1978, On the validity of the formal edgeworth expansion, *Annals of Statistics* **6**, 434–451.

Bravo, F.: 2004, Empirical likelihood based inference with applications to some econometric models, *Econometric Theory* **20**, 231–264.

Bravo, F.: 2022, Second order expansions of estimators in nonparametric moment conditions models with weakly dependent data, *Econometric Reviews* **41**, 583–606.

Brown, B. and Newey, W.: 2002, Generalized method of moments, efficient bootstrapping, and improved inference, *Journal of Business and Economic Statistics* **20**, 507–517.

Cai, Z.: 2003, Nonparametric estimating equations for time series data, *Statistics and Probability Letters* **62**, 379–390.

Cai, Z., Das, M., Xiong, H. and Wu, X.: 2006, Functional coefficients instrumental variables models, *Journal of Econometrics* **133**, 207–241.

Cai, Z. and Li, Q.: 2008, Nonparametric estimation of varying coefficient dynamic panel data models, *Econometric Theory* **24**, 1321–1342.

Cai, Z., Ren, Y. and Sun, L.: 2015, Pricing kernel estimation: a local estimating equation approach, *Econometric Theory* **31**, 560–580.

Calonico, S., Cattaneo, M. and Farrell, M.: 2018, On th effect of bias estimation on coverage accuracy in nonparametric inference, *Journal of the American Statistical Association* **113**, 767–779.

Chandra, T. and Ghosh, J.: 1980, Valid asymptotic expansions for the likelihood ratio under continuous alternatives, *Sankhya, A.* **62**, 170–184.

Chen, S. and Cui, H.: 2006, On Bartlett correction of empirical likelihood in presence of nuisance parameters, *Biometrika* **93**, 215–220.

Chen, S. and Cui, H.: 2007, On second order properties of empirical likelihood with moment restrictions, *Journal of Econometrics* **107**, 492–516.

Chen, S. X.: 1996, Empirical likelihood confidence intervals for nonparametric density estimation, *Biometrika* **83**, 329–341.

Chen, X. and Qin, S.: 2000, Empirical likelihood confidence intervals for local linear smoothers, *Biometrika* **87**, 946–953.

Cochrane, J.: 2001, *Asset Pricing*, Princeton University Press, Princeton.

Cordeiro, G. M. and Ferrari, S. L. P.: 1991, A modified score statistic having chi-squared distribution to order n^{-1}, *Biometrika* **78**, 573–582.

Cox, D. R. and Reid, N.: 1987a, Approximations to noncentral distributions, *Canadian Journal of Statistics* **15**, 105–114.

Cox, D. and Reid, N.: 1987b, Parameter orthogonality and approximate conditional inference, *Journal of the Royal Statistical Society, B* **49**, 1–39.

DiCiccio, T. J. and Romano, J. P.: 1989, On adjustments based on the signed root of the empirical likelihood ratio statistic, *Biometrika* **76**, 447–456.

Fang, Y., Ren, Y. and Yuan, Y.: 2011, Nonparametric estimation and testing of stochastic discount factor, *Finance Research Letters* **8**, 196–205.

Gospodinov, N., Kan, R. and Robotti, C.: 2013, Chi-squared tests for evaluation and comparison of asset pricing models, *Journal of Econometrics* **173**, 108–125.

Griliches, Z.: 1976, Wages of very young men, *Journal of Political Economy* **84**, 69–85.

Hall, P.: 1991, Edgeworth expansions for nonparametric density estimators, with applications, *Statistics* **22**, 215–232.

Hall, P.: 1992a, *The Bootstrap and Edgeworth Expansion*, Lecture Notes in Statistics, New York: Springer & Verlag.

Hall, P.: 1992b, On bootstrap confidence interval in nonparametric regression, *Annals of Statistics* **20**, 695–711.

Hall, P. and Horowitz, J.: 1996, Bootstrap critical values for tests based on generalized-method of moment estimators, *Econometrica* **64**, 891–916.

Hansen, L.: 1982, Large sample properties of generalized method of moments estimators, *Econometrica* **50**, 1029–1054.

Hansen, L., Heaton, J. and Yaron, A.: 1996, Finite-sample properties of some alternative GMM estimators, *Journal of Business and Economic Statistics* **14**, 262–280.

Hansen, L. and Jagannathan, R.: 1997, Assessing specification errors in stochastic discount factors models, *Journal of Finance* **52**, 557–590.

Horowitz, J.: 1998, Bootstrap methods for median regression models, *Econometrica* **66**, 1327–1351.

Kitamura, Y. and Stutzer, M.: 1997, An information theoretic alternative to generalized method of moments estimation, *Econometrica* **65**, 861–874.

Lewbel, A.: 2007, A local generalized method of moments estimator, *Economics Letters* **94**, 124–128.

Li, Q. and Racine, J.: 2004, Cross validated local linear nonparametric regression, *Statistica Sinica* **14**, 485–512.

Matsushita, Y. and Otsu, T.: 2013, Second order refinement of empirical likelihood for testing overidentifying restrictions, *Econometric Theory* **29**, 324–353.

McCullagh, P.: 1987, *Tensor Methods in Statistics*, London: Chapman and Hall.

Mincer, J.: 1956, Investment in human capital and personal income, *Journal of Political Economy* **66**, 281–302.

Newey, W. and Smith, R.: 2004, Higher order properties of GMM and generalized empirical likelihood estimators, *Econometrica* **72**, 219–256.

Nickl, R. and Potscher, D.: 2007, Bracketing metric entropy rates and empirical central limit theoremsfor fuction classes of Besov and Sobolev-type, *Journal of Theoretical Probability* **20**, 177–199.

Otsu, T., Xu, K.-L. and Matsushita, Y.: 2015, Empirical likelihood for regression discontinuity design, *Journal of Econometrics* **186**, 94–112.

Owen, A.: 1988, Empirical likelihood ratio intervals for a single functional, *Biometrika* **75**.

Pakes, A. and Pollard, D.: 1989, Simulation and the asymptotics of optmization estimators, *Econometrica* **57**, 1027–1057.

Qu, A., Lindsay, B. G. and Li, B.: 2000, Improving generalised estimating equations using quadratic inference functions, *Biometrika* **87**(4), 823–836.

Severini, T. and Staniswalis, J.: 1994, Quasi-likelihood estimation in semiparametric models, *Journal of The American Statistical Association* **89**, 501–511.

Smith, R. J.: 1997, Alternative semi-parametric likelihood approaches to generalised method of moments estimation, *Economic Journal* **107**, 503–519.

Trostel, P.: 2005, Nonlinearities in the return of education, *Journal of Applied Economics* **VIII**, 191–202.

Wilks, S.: 1938, The large-sample distribution of the likelihood ratio for testing composite hypotheses, *Annals of Mathematical Statistics* **9**, 60–62.

Wong, W. and Severini, T.: 1991, On maximum likelihood estimation in infinite dimensional spaces, *Annals of Statistics* **19**, 603–632.

7 Supplemental appendix

Throughout the Appendix "CMT", "CLT", "LLN", denote Continuous Mapping Theorem, Central Limit Theorem, (possibly uniform) Law of Large Numbers and with Probability Approaching 1. For any square matrix say M, $\sigma_{\min}(M)$ and $\sigma_{\max}(M)$ denote its minimum and maximum eigenvalue. Finally, for simplicity of notation, let $d_Z := d$, $\sum_{i=1}^{n} := \sum$ and from Proposition (7.2) onward we drop the dependence on z of λ, h and related estimators.

7.1 Preliminary results

Proposition 7.1 *Let $|p| = j_1 + ... + j_p$ and $\partial^p \cdot = \partial^p \cdot / \partial h_{j_1} ... \partial h_{j_p}$ for $j = 1, ..., l$ and $j_s = 1, ..., k$ denote a partial differential operator; assume that there exist integrable functions $\varphi_1(W)$ and $\varphi_p(W)$ such that for any $h, \widetilde{h} \in \mathcal{H}_C$ and $z, \widetilde{z} \in \mathcal{Z}$*

$$\begin{aligned}
\left\| g^K(h(z)) - g^K\left(\widetilde{h}(\widetilde{z})\right) \right\| &\leq \varphi_1(W) \left\| h(z) - \widetilde{h}(\widetilde{z}) \right\|, \\
\left\| \partial^p g_j^K(h(z)) - \partial^p g_j^K\left(\widetilde{h}(\widetilde{z})\right) \right\| &\leq \varphi_p(W) \left\| h_{j_s}(z) - \widetilde{h}_{j_s}(\widetilde{z}) \right\|.
\end{aligned}$$

Then the classes of functions

$$\begin{aligned}
\mathcal{G}_1^K &= \left\{ g^K(h(z)), b > 0, z \in \mathcal{Z}, h \in \mathcal{H}_C \right\}, \\
\mathcal{G}_{\partial^p}^K &= \left\{ \partial^p g_j^K(h(z)), b > 0, z \in \mathcal{Z}, h \in \mathcal{H}_C \right\}.
\end{aligned}$$

are Euclidean with envelopes $(|g_j| + C\varphi_1(W)) \sup_{z \in \mathcal{Z}} K(Z - z) := G_1$ and $(|\partial^p g_j(h)| + C\varphi_p(W)) \times$

$\sup_{z \in \mathcal{Z}} K(Z - z) := G_{\partial^p}$, where g_j and $\partial^p g_j(h)$ are evaluated at any arbitrary $h \in \mathcal{H}_C$. Furthermore any componentwise products of the classes $\mathcal{G}_1^K$ and $\mathcal{G}_{\partial^p}$ are Euclidean as long as the corresponding envelopes are integrable.

Proof. Consider the classes of functions

$$\begin{aligned}
\mathcal{G}_1 &= \left\{ g(h(z)), b > 0, z \in \mathcal{Z}, h \in \mathcal{H}_C \right\}, \\
\mathcal{K} &= \left\{ K\left(\frac{Z - z}{b}\right), b > 0, z \in \mathcal{Z} \right\}.
\end{aligned}$$

Note that for any $h, \widetilde{h} \in \mathcal{H}_C$ and $z, \widetilde{z} \in \mathcal{Z}$

$$\begin{aligned}
\left\| h(z) - \widetilde{h}(\widetilde{z}) \right\| &\leq \sup_{z \in \mathcal{Z}} \left\| h(z) - \widetilde{h}(z) \right\| + \|z - \widetilde{z}\| \sup_{z \neq \widetilde{z} \in \mathcal{Z}} \frac{\left\| h(z) - \widetilde{h}(\widetilde{z}) \right\|}{\|z - \widetilde{z}\|} \\
&\leq \left\| h(z) - \widetilde{h}(z) \right\|_L + C \|z - \widetilde{z}\|,
\end{aligned}$$

for some $C > 0$, where $||\cdot||_L$ is the Lipschitz norm, and thus

$$\left\| g(h) - g\left(\widetilde{h}\right) \right\| \leq C\varphi_1(W)\left(\left\| h - \widetilde{h} \right\|_L + \|z - \widetilde{z}\| \right).$$

The class of functions $\mathcal{K}$ is also Euclidean with envelope $\sup_{z\in\mathcal{Z}} K(Z-z)$, hence the class of functions $\mathcal{G}_1^K$ is Euclidean with envelope $(\|g\| + C\varphi_1(W)) \sup_{z\in\mathcal{Z}} K(Z-z)$, since it is the pointwise product of two Euclidean classes Pakes and Pollard (1989). A similar argument can be used to show that the class of functions $\mathcal{G}_{\partial^p}^K$ is Euclidean. The last statement follows again by the fact that products of Euclidean classes are Euclidean given the assumed integrability of their envelops. ■

Note that the assumptions of Proposition 7.1 are satisfied for $\varphi_1(W) = \sup_{h\in\mathcal{H}_C, z\in\mathcal{Z}} \left\| \partial g^K(h) / \partial h^T \right\|$ and similarly for $\varphi_2(W)$.

The following proposition establishes an EB LNN (EBLLN). For any EB statistic T^{*K}, let $\left\| T^{*K} \right\| = o_{p^*-p}(1)$ denote

$$P\left\{ \left(P^* \left\| T^{*K} \right\| > \varepsilon \right) > \delta \right\} = o(1)$$

for any ε, $\delta > 0$ as $nb^d \to \infty$.

Proposition 7.2 *Assume that the classes of functions* $\mathcal{F}^K = \left\{ g^K(h), b > 0, h \in \mathcal{H}_C \right\}$ *and* $\mathcal{F}_{\partial^p}^K = \left\{ \partial^p g_j^K(h), b > 0, h \in \mathcal{H}_C \right\}$ $(j = 1, ..., l)$ *where* ∂^p *is the partial differential operator of Proposition 7.1 are Euclidean with integrable envelopes* F_1 *and* F_{∂^p}. *Then for all* $h \in \mathcal{H}_C$

$$\begin{aligned} \left\| \frac{1}{nb^d} \sum g_i^{*K}(h) - E^* \left[\frac{g_i^{*K}(h)}{b^d} \right] \right\| &= o_{p^*-p}(1), \\ \left\| \frac{1}{nb^d} \sum \partial^p g_{ij}^{*K}(h) - E^* \left[\frac{\partial^p g_{ij}^{*K}(h)}{b^d} \right] \right\| &= o_{p^*-p}(1). \end{aligned}$$

Proof. Note that

$$E^* \left\| \frac{g_i^{*K}(h)}{b^d} \right\| \leq \left| 1 + \max_i \left| \widehat{\lambda}^T g_i^K(h) \right| \right| \frac{1}{nb^d} \sum \left\| g_i^K(h) \right\| = O_p(1) + o_p(1)$$

and that by Chebychev's and Markov's inequalities, for any given $\overline{h} \in \mathcal{H}_C$ and $\varepsilon, \delta > 0$

$$P\left(P^* \left(\left\| \frac{1}{nb^d} \sum g_i^{*K}\left(\overline{h}\right) - E^* \left[\frac{g_i^{*K}\left(\overline{h}\right)}{b^d} \right] \right\| \geq \varepsilon \right) \geq \delta \right) = O\left(\frac{1}{nb^d} \right) = o(1)$$

as $nb \to \infty$, hence

$$\left\| \frac{1}{nb^d} \sum g_i^{*K}\left(\overline{h}\right) - E^* \left[\frac{g_i^{*K}\left(\overline{h}\right)}{b^d} \right] \right\| = o_{p^*-p}(1). \tag{34}$$

Furthermore, exploiting the fact that $\mathcal{H}_C$ is a compact set, we can cover $\mathcal{H}_C$ by a number $N(\zeta)$ of subsets $\mathcal{H}_C^k$ centered at h_k with radius $\zeta > 0$, which is bounded by a positive constant that depends only on ζ. Then by triangle inequality

$$\sup_{h\in\mathcal{H}_C}\left\|\frac{1}{nb^d}\sum g_i^{*K}(h)-E^*\left[\frac{g_i^{*K}(h)}{b^d}\right]\right\|\leq\max_k\left\|\frac{1}{nb^d}\sum g_i^{*K}(h_k)-E^*\left[\frac{g_i^{*K}(h_k)}{b^d}\right]\right\|+$$
$$\max_k\left\|\frac{1}{nb^d}\sum g_i^{*K}(h)-\frac{1}{nb^d}\sum g_i^{*K}(h_k)\right\|+\max_k\left\|E^*\left[\frac{g_i^{*K}(h)}{b^d}\right]-E^*\left[\frac{g_i^{*K}(h_k)}{b^d}\right]\right\| \tag{35}$$

and by (34) and the fact that the class of functions $\mathcal{F}_1^K$ is Euclidean, it can be shown that, for an appropriately chosen ζ, the first two terms in (35) are $o_{p^*-p}(1)$ and the last one is $o_p(1)$ (see for example Horowitz and Hall (1996) combined with Proposition 7.1). The same arguments can be used to show the second conclusion. ■

7.2 Proof of the main results

Proof of Theorem 1. Recall that

$$\mu_{jkl...}(h)=E\left[\frac{1}{b^d}\left[v^{jm}(h)\right]^{1/2}g_m^K(h)\left[\mu^{kn}\right]^{1/2}g_n^K(h)\left[\mu^{lo}(h)\right]^{1/2}g_o^K(h)\ldots\right],$$

and define

$$U_{jkl...}^K(h):=U_{jkl...}(h)=\frac{1}{(nb^d)^{1/2}}\sum\Big(\left[\omega^{jm}(h)\right]^{1/2}g_{im}^K(h)\left[\omega^{kn}(h)\right]^{1/2}g_{in}^K(h)\times$$
$$\left[\omega^{lo}(h)\right]^{1/2}g_{io}^K(h)\ldots-\mu_{jkl...}(h)\Big).$$

By a Taylor expansion of the FOCs

$$0=\frac{1}{nb^d}\sum\gamma_1\left(\widehat{\lambda}\right)g_{ij}^K(h_0), \tag{36}$$

and solving for $\left(nb^d\right)^{1/2}\widehat{\lambda}_j$, it follows that under $nb^{d+4}\to 0$

$$\begin{aligned}\left(nb^d\right)^{1/2}\widehat{\lambda}_j&=-U_j(h_0)+\frac{1}{(nb^d)^{1/2}}\Big(U_{jk}(h_0)U_k(h_0)+\frac{\gamma_3}{2}\mu_{jkl}U_k(h_0)U_l(h_0)\Big)+\\&\frac{1}{nb^d}\Big(-\frac{\gamma_3}{2}U_{jk}(h_0)\mu_{kno}U_n(h_0)U_o(h_0)+\\&\frac{\gamma_3}{2}U_{jlm}(h_0)U_l(h_0)U_m(h_0)-\gamma_3\mu_{jlm}(h_0)U_l(h_0)U_{mp}(h_0)U_p(h_0)-\\&\frac{\gamma_3^2}{2}\mu_{jlm}(h_0)U_l(h_0)\mu_{mno}U_n(h_0)U_o(h_0)-U_{jl}(h_0)U_{lm}(h_0)U_m(h_0)+\\&\frac{\gamma_4}{4!}\mu_{jklm}(h_0)U_k(h_0)U_l(h_0)U_m(h_0)\Big)+O_p\left(\frac{1}{(nb^d)^{3/2}}\right),\end{aligned} \tag{37}$$

A Taylor expansion of $S_\gamma\left(\widehat{\lambda}, h_0\right)$ about 0 yields

$$S_\gamma\left(\widehat{\lambda}, h_0\right) = 2\left(-\sum g_{ij}\left(h_0\right)\widehat{\lambda}_j - \frac{1}{2}\sum\left(g_{ij}\left(h_0\right)\widehat{\lambda}_j\right)^2 + \right. \tag{38}$$

$$\left.\frac{\gamma_3}{3!}\sum\left(g_{ij}\left(h_0\right)\widehat{\lambda}_j\right)^3 + \frac{\gamma_4}{4!}\sum\left(g_{ij}\left(h_0\right)\widehat{\lambda}_j\right)^4\right) + O_p\left(\left(nb^d\right)\left(nb^d\right)^{-5/2}\right) =$$

$$2\left(-U_j\left(h_0\right)\left(nb^d\right)^{1/2}\widehat{\lambda}_j - \frac{1}{2}\left(\frac{U_{jk}\left(h_0\right)}{\left(nb^d\right)^{1/2}} + \mu_{jk}\left(h_0\right)\right)\left(nb^d\right)\widehat{\lambda}_j\widehat{\lambda}_k + \right.$$

$$\frac{\gamma_3}{3!}\left(\frac{U_{jkl}\left(h_0\right)}{nb^d} + \frac{\mu_{jkl}\left(h_0\right)}{\left(nb^d\right)^{1/2}}\right)\left(nb^d\right)^{3/2}\widehat{\lambda}_j\widehat{\lambda}_k\widehat{\lambda}_l +$$

$$\left.\frac{\gamma_4}{4!nb^d}\mu_{jklm}\left(h_0\right)\left(nb^d\right)^2\widehat{\lambda}_j\widehat{\lambda}_k\widehat{\lambda}_l\widehat{\lambda}_m\right) + O_p\left(\frac{1}{\left(nb^d\right)^{3/2}}\right),$$

hence using (37) we have

$$S_\gamma\left(\widehat{\lambda}, h_0\right) = U_j\left(h_0\right)U_j\left(h_0\right) + \frac{1}{\left(nb^d\right)^{1/2}}\left\{-U_{jk}\left(h_0\right)U_j\left(h_0\right)U_k\left(h_0\right) - \right. \tag{39}$$

$$\left.\frac{\gamma_3}{3}\mu_{jkl}\left(h_0\right)U_j\left(h_0\right)U_k\left(h_0\right)U_l\left(h_0\right)\right\} +$$

$$\frac{1}{nb^d}\left\{-\frac{\gamma_3}{3}U_{jkl}\left(h_0\right)U_j\left(h_0\right)U_k\left(h_0\right)U_l\left(h_0\right) + \right.$$

$$\gamma_3\mu_{jkm}\left(h_0\right)U_j\left(h_0\right)U_{kp}\left(h_0\right)U_m\left(h_0\right)U_p\left(h_0\right) + U_{jl}\left(h_0\right)U_j\left(h_0\right)U_{ln}\left(h_0\right)U_n\left(h_0\right) +$$

$$\frac{\gamma_3^2}{4}\mu_{jkl}\left(h_0\right)U_j\left(h_0\right)U_k\left(h_0\right)\mu_{loh}\left(h_0\right)U_0\left(h_0\right)U_h\left(h_0\right) +$$

$$\left.\frac{\gamma_4}{12}\mu_{jklm}\left(h_0\right)U_j\left(h_0\right)U_k\left(h_0\right)U_l\left(h_0\right)U_m\left(h_0\right)\right\} + O_p\left(\frac{1}{\left(nb^d\right)^{3/2}}\right).$$

The signed square root of $S_\gamma\left(\widehat{\lambda}, h_0\right)$ is

$$S_{\gamma j}\left(\widehat{\lambda}, h_0\right) = S^1_{\gamma j}\left(\widehat{\lambda}, h_0\right) + S^2_{\gamma j}\left(\widehat{\lambda}, h_0\right) + S^3_{\gamma j}\left(\widehat{\lambda}, h_0\right) + O_p\left(\frac{1}{\left(nb^d\right)^{3/2}}\right), \tag{40}$$

where

$$S^1_{\gamma j}\left(\widehat{\lambda}, h_0\right) = U_j\left(h_0\right), \quad S^2_{\gamma j}\left(\widehat{\lambda}, h_0\right) = \frac{1}{\left(nb^d\right)^{1/2}}\left(-\frac{1}{2}U_{jk}\left(h_0\right)U_k\left(h_0\right) - \right.$$

$$\left.\frac{\gamma_3}{6}\mu_{jkl}\left(h_0\right)U_k\left(h_0\right)U_l\left(h_0\right)\right), \quad S^3_{\gamma j}\left(\widehat{\lambda}, h_0\right) = \frac{1}{nb^d}\left(-\frac{\gamma_3}{6}U_{jkl}\left(h_0\right)U_k\left(h_0\right)U_l\left(h_0\right) + \right.$$

$$\frac{5\gamma_3}{12}\mu_{jkl}\left(h_0\right)U_{lm}\left(h_0\right)U_k\left(h_0\right)U_m\left(h_0\right) + \frac{3}{8}U_{jk}\left(h_0\right)U_{kl}\left(h_0\right)U_l\left(h_0\right) +$$

$$\left.\frac{\gamma_3^2}{9}\mu_{jkl}\left(h_0\right)U_k\left(h_0\right)\mu_{lmn}\left(h_0\right)U_m\left(h_0\right)U_n\left(h_0\right) + \frac{\gamma_4}{24}\mu_{jklm}U_k\left(h_0\right)U_l\left(h_0\right)U_m\left(h_0\right)\right),$$

and note that

$$S_{\gamma}\left(\widehat{\lambda},h_0\right) = S^1_{\gamma j}\left(\widehat{\lambda},h_0\right)S^1_{\gamma j}\left(\widehat{\lambda},h_0\right)+2S^2_{\gamma j}\left(\widehat{\lambda},h_0\right)S^1_{\gamma j}\left(\widehat{\lambda},h_0\right)+ 2S^3_{\gamma j}\left(\widehat{\lambda},h_0\right)S^1_{\gamma j}\left(\widehat{\lambda},h_0\right)+S^2_{\gamma j}\left(\widehat{\lambda},h_0\right)S^2_{\gamma j}\left(\widehat{\lambda},h_0\right).$$

Some calculations show that the first four moments of $S_{\gamma j}\left(\widehat{\lambda},h_0\right)$ are

$$E\left[S_{\gamma j}\left(\widehat{\lambda},h_0\right)\right] = -\frac{1}{\left(nb^d\right)^{1/2}}\frac{(3+\gamma_3)}{6}\mu_{jkk}\left(h_0\right)+O_p\left(\frac{1}{\left(nb^d\right)^{3/2}}\right),$$

$$E\left[S_{\gamma j}\left(\widehat{\lambda},h_0\right)S_{\gamma k}\left(\widehat{\lambda},h_0\right)\right] = \delta_{jk}+\frac{1}{nb^d}\left\{\left(-\gamma_3+\frac{\gamma_4}{4}\right)\mu_{jkll}\left(h_0\right)+ \left(\frac{3}{4}+\frac{5\gamma_3}{6}+\frac{2\gamma_3^2}{9}\right)\mu_{jkl}\left(h_0\right)\mu_{lmm}\left(h_0\right)+\left(1+\frac{5\gamma_3}{3}+\frac{\gamma_3^2}{2}\right)\mu_{jlm}\left(h_0\right)\mu_{klm}\left(h_0\right)+ \left(\frac{1}{2}+\frac{\gamma_3}{6}\right)^2\mu_{jll}\left(h_0\right)\mu_{kll}\left(h_0\right)\right\}+O_p\left(\frac{1}{\left(nb^d\right)^2}\right),$$

$$E\left[S_{\gamma j}\left(\widehat{\lambda},h_0\right)S_{\gamma k}\left(\widehat{\lambda},h_0\right)S_{\gamma l}\left(\widehat{\lambda},h_0\right)\right] = \frac{\mu_{jkl}\left(h_0\right)}{\left(nb^d\right)^{1/2}}- \frac{[3]\left(3+\gamma_3\right)}{\left(nb^d\right)^{1/2}6}\delta_{jk}\mu_{lmm}\left(h_0\right)+O_p\left(\frac{1}{\left(nb^d\right)^{3/2}}\right),$$

$$E\left[S_{\gamma j}\left(\widehat{\lambda},h_0\right)S_{\gamma k}\left(\widehat{\lambda},h_0\right)S_{\gamma l}\left(\widehat{\lambda},h_0\right)S_{\gamma m}\left(\widehat{\lambda},h_0\right)\right] = [3]\,\delta_{jk}\delta_{lm}+ \frac{1}{nb^d}\left[\left(\gamma_4-4\gamma_3-2\right)\mu_{jklm}\left(h_0\right)+\left(\frac{3\gamma_4}{2}-6\gamma_3\right)\delta_{jk}\mu_{lmnn}\left(h_0\right)+ \left(4+\frac{10\gamma_3}{3}+\frac{2\gamma_3^2}{3}\right)\mu_{jkl}\left(h_0\right)\mu_{mnn}\left(h_0\right)+\left(12+14\gamma_3+4\gamma_3^2\right)\mu_{jkn}\left(h_0\right)\mu_{lmn}\left(h_0\right)+ \left(\frac{9}{2}+5\gamma_3+\frac{4\gamma_3^2}{3}\right)\delta_{jk}\mu_{lmn}\left(h_0\right)\mu_{noo}\left(h_0\right)+ \left(6+10\gamma_3+3\gamma_3^2\right)\delta_{jk}\mu_{lno}\left(h_0\right)\mu_{mno}\left(h_0\right)+ \frac{\left(3+\gamma_3\right)^2}{6}\delta_{jk}\mu_{lnn}\left(h_0\right)\mu_{moo}\left(h_0\right)\right]+O_p\left(\frac{1}{\left(nb^d\right)^2}\right),$$

with the higher approximate moments being of order $O_p\left(\left(nb^d\right)^{-3/2}\right)$ or smaller. Then using the following formulas for the cumulants

$$\begin{aligned}
&\kappa_{j,k}(h_0) = E\left[S_{\gamma j}\left(\widehat{\lambda},h_0\right)S_{\gamma k}\left(\widehat{\lambda},h_0\right)\right] - \kappa_j(h_0)\,\kappa_k(h_0)\,, \\
&\kappa_{j,k,l}(h_0) = E\left[S_{\gamma j}\left(\widehat{\lambda},h_0\right)S_{\gamma k}\left(\widehat{\lambda},h_0\right)S_{\gamma l}\left(\widehat{\lambda},h_0\right)\right] - \\
&[3]\,\kappa_j(h_0)\,E\left[S_{\gamma k}\left(\widehat{\lambda},h_0\right)S_{\gamma l}\left(\widehat{\lambda},h_0\right)\right] + 2\kappa_j(h_0)\,\kappa_k(h_0)\,\kappa_l(h_0)\,, \\
&\kappa_{j,k,l,m}(h_0) = E\left[S_{\gamma j}\left(\widehat{\lambda},h_0\right)S_{\gamma k}\left(\widehat{\lambda},h_0\right)S_{\gamma l}\left(\widehat{\lambda},h_0\right)S_{\gamma m}\left(\widehat{\lambda},h_0\right)\right] - \\
&[3]\,E\left[S_{\gamma j}\left(\widehat{\lambda},h_0\right)S_{\gamma k}\left(\widehat{\lambda},h_0\right)\right]E\left[S_{\gamma l}\left(\widehat{\lambda},h_0\right)S_{\gamma m}\left(\widehat{\lambda},h_0\right)\right] - \\
&[4]\,\kappa_j(h_0)\,E\left[S_{\gamma k}\left(\widehat{\lambda},h_0\right)S_{\gamma l}\left(\widehat{\lambda},h_0\right)S_{\gamma m}\left(\widehat{\lambda},h_0\right)\right] + \\
&2\,[6]\,\kappa_j(h_0)\,\kappa_k(h_0)\,E\left[S_{\gamma l}\left(\widehat{\lambda},h_0\right)S_{\gamma m}\left(\widehat{\lambda},h_0\right)\right] - \\
&6\kappa_j(h_0)\,\kappa_k(h_0)\,\kappa_l(h_0)\,\kappa_m(h_0)\,,
\end{aligned} \tag{41}$$

it follows that the first four approximate cumulants $\kappa_j(h_0)$, $\kappa_{j,k}(h_0)$, $\kappa_{j,k,l}(h_0)$ and $\kappa_{j,k,l,m}(h_0)$ of $S_{\gamma j}\left(\widehat{\lambda},h_0\right)$ are

$$\begin{aligned}
&\kappa_j(h_0) = -\frac{1}{\left(nb^d\right)^{1/2}}\frac{(3+\gamma_3)}{6}\mu_{jkk}(h_0) + O_p\left(\frac{1}{\left(nb^d\right)^{3/2}}\right), \\
&\kappa_{j,k}(h_0) = \delta_{jk} + \kappa_{j,k}^{(2)}(h_0)\,, \quad \kappa_{j,k}^{(2)}(h_0) = -\frac{1}{nb^d}\left(\left(-\gamma_3 + \frac{\gamma_4}{4}\right)\mu_{jkll}(h_0) + \right. \\
&\left(\frac{3}{4} + \frac{5\gamma_3}{6} + \frac{2\gamma_3^2}{9}\right)\mu_{jkl}(h_0)\,\mu_{lmm}(h_0) + \\
&\left.\left(1 + \frac{5\gamma_3}{3} + \frac{\gamma_3^2}{2}\right)\mu_{jlm}(h_0)\,\mu_{klm}(h_0)\right) + O_p\left(\frac{1}{\left(nb^d\right)^2}\right), \\
&\kappa_{j,k,l}(h_0) = -\frac{\mu_{jkl}(h_0)}{\left(nb^d\right)^{1/2}}(2+\gamma_3) + O_p\left(\frac{1}{\left(nb^d\right)^{3/2}}\right), \\
&\kappa_{j,k,l,m}(h_0) = \frac{1}{nb^d}\left[(\gamma_4 - 4\gamma_3 - 2)\,\mu_{jklm}(h_0) + \right. \\
&\left(4 + \frac{10\gamma_3}{3} + \frac{2\gamma_3^2}{3}\right)\mu_{jkl}(h_0)\,\mu_{mnn}(h_0) + \\
&\left.\left(12 + 14\gamma_3 + 4\gamma_3^2\right)\mu_{jkn}(h_0)\,\mu_{lmn}(h_0)\right] + O_p\left(\frac{1}{\left(nb^d\right)^2}\right).
\end{aligned}$$

A formal Edgeworth expansion for $S_{\gamma j}\left(\widehat{\lambda},h_0\right)$ is

$$\sup_{A\in\mathcal{A}}\left|P\left(S_{\gamma j}\left(\widehat{\lambda},h_0\right)\in A\right) - \int_A p_{1n}(u)\,\phi(u)\,du\right| = O\left(\frac{1}{\left(nb^d\right)^{3/2}}\right), \tag{42}$$

where $\mathcal{A}$ is a class of Borel measurable subsets of $\mathbb{R}^{\overline{k}}$, $\overline{k} = k + k(k+1)/2 + k^2(k+1)/2$,

$$\begin{aligned} p_{1n}(u) &= 1 + \kappa_j(h_0) H_j(u) + \frac{1}{6}\kappa_{j,k,l}(h_0) H_{jkl}(u) + \frac{1}{2}\left[\kappa_{j,k}^{(2)}(h_0) + \kappa_j(h_0)\kappa_k(h_0)\right] H_{jk}(u) + \\ & \frac{1}{24}\left[\kappa_{j,k,l,m}(h_0) + [4]\kappa_j(h_0)\kappa_{k,l,m}(h_0)\right] H_{jklm}(u) + \\ & \frac{1}{72}\kappa_{j,k,l}(h_0)\kappa_{m,n,o}(h_0) H_{jklmno}(u), \end{aligned}$$

and $H_j(\cdot), \ldots., H_{jklmno}(\cdot)$ are the first six multivariate Hermite polynomials (see for example McCullagh (1987)). Assuming that (42) is valid, using results of Bravo (2004), it follows that for any $c \in [c_0, \infty)$ $(c_0 > 0)$

$$\begin{aligned} P\left(S_\gamma\left(\widehat{\lambda}, h_0,\right) \le c\right) = G_k(c) + \int_{u^T u \le c} p_{1n}(u)\phi(u) + O\left(\frac{1}{(nb^d)^{3/2}}\right) = G_k(c) + \\ \frac{1}{nb^d}\left\{\frac{1}{2}b_{1\gamma}(K)\nabla G_k(c) + \frac{1}{24}b_{2\gamma}(K)\nabla^2 G_k(c) + \frac{1}{72}b_{3\gamma}(K)\nabla^3 G_k(c)\right\} + O\left(\frac{1}{(nb^d)^2}\right), \end{aligned} \tag{43}$$

where $\nabla^k G_k(\cdot)$ is the kth difference operator applied to the distribution $G_k(\cdot)$ i.e. $\nabla^k G_k(\cdot) = \sum_{j=0}^{k}(-1)^j \binom{k}{j} G_{k2(k-j)}(\cdot)$,

$$\begin{aligned} b_{1\gamma}(K) &= \left(\gamma_3 - \frac{\gamma_4}{4}\right)\mu_{jjkk}(h_0) + \left(1 + \frac{5\gamma_3}{3} + \frac{\gamma_3^2}{2}\right)\mu_{jkl}(h_0)\mu_{jkl}(h_0) + \frac{(\gamma_3+2)^2}{36}\mu_{jjk}(h_0)\mu_{kll}(h_0), \\ b_{2\gamma}(K) &= [3]\left[(\gamma_4 - 4\gamma_3 - 2)\mu_{jklm}(h_0) + 2\left(4 + \frac{10\gamma_3}{3} + \frac{2\gamma_3^2}{3}\right)\mu_{jkl}(h_0)\mu_{mnn}(h_0) + \right. \\ & \left.\left(12 + 14\gamma_3 + 4\gamma_3^2\right)\mu_{jkn}(h_0)\mu_{lmn}(h_0)\right]\delta_{jk}\delta_{lm}, \\ b_{3\gamma}(K) &= [15](2+\gamma_3)^2\mu_{jkl}(h_0)\mu_{mno}(h_0)\delta_{jk}\delta_{lm}\delta_{no}, \end{aligned}$$

and the order of the remaiinder follows by the oddness/evenness property of the Hermite polynomials (Barndorff-Nielsen and Hall 1988), hence the conclusion follows after some algebra. To justify the validity of (42), we can use the same arguments of Bhattacharya and Ghosh (1978) noting that (40) can be expressed as a smooth function of $U_j(h_0)$, $U_{jk}(h_0)$ and $U_{jkl}(h_0)$ and the Cramer's condition is replaced by the (similar) one given for example in Hall (1991). To be specific, note that

$$\begin{aligned} \left|E\exp\left(\iota t^T U_{\overline{k}}(h_0, K)\right)\right| &\le \left|E(1 - F(h|W)) + F(-h|W)\exp\left(\iota t^T U(h_0)\right)\right| + \\ & \left|\int\int_{-1}^{1}\exp\left(\iota t^T U_{\overline{k}}(h_0, K)\right) f(z|w)\,dz\,dF(w)\right| \end{aligned} \tag{44}$$

where

$$U_{\overline{k}}(h, K) = \left[U_1(h), \ldots, U_k(h), U_{11}(h), \ldots, U_{kk}, U_{111}(h), \ldots., U_{kkk}(h)\right]^T.$$

Then by a Taylor expansion the first term in (44) is less than $1-hE\left(f\left(0|W\right)\right)$ and by partitioning $[-1,1]$ so as to satisfy **A6**(ii), following Horowitz (1998) it is possible to show that for $V_{\overline{k}}\left(\xi\right)=t^{T}U_{\overline{k}}\left(h_0,K\right)/\left\|t\right\|$ and $\tau=\left\|t\right\|$

$$\sup_{\|t\|>\varepsilon}\int_{-1}^{1}\exp\left(\iota t^{T}U_{\overline{k}}\left(h_0,K\right)\right)f\left(0|w\right)dz=\sup_{\tau>\varepsilon}\int_{u_{m-1}}^{u_m}\exp\left(\iota\tau V_{\overline{k}}\left(\xi\right)\right)d\xi<C_1<1.$$

Also, since $P\left(U_{\overline{k}}\left(h_0,K\right)=0\right)<1$, there exist $\eta>0$ and $\gamma<1$ such that $\int_{\|w\|<\eta}f\left(0|w\right)dF\left(w\right)=\gamma C_2$ for some $C_2>0$ hence

$$\left|\int\int_{-1}^{1}\exp\left(\iota t^{T}U_{\overline{k}}\left(h_0,K\right)\right)f\left(0|w\right)dzdF\left(w\right)\right|<b\left(\gamma C_2+\left(1-\gamma\right)C_2C_1\right)=bC_3$$

where $C_3<1$. Since

$$\int_{-1}^{1}\left|f\left(h\xi|w\right)-f\left(0|w\right)\right|d\xi dF\left(w\right)\leq\varepsilon E\left(f\left(0|W\right)\right)$$

for b sufficiently small, we have by the triangle equality that

$$\sup_{\|t\|>\epsilon}\left|E\exp\left(\iota t^{T}U_{\overline{k}}\left(h_0,K\right)\right)\right|<1-bE\left(f\left(0|W\right)\right)\left(1-\varepsilon-C_3\right),\tag{45}$$

which is Hall's (1991) analog of Cramer's condition for nonparametric estimators. It remains to justify expansion (15); first note that by (39)

$$S_{\gamma}\left(\widehat{\lambda},h_0\right)=U_j\left(h_0\right)U_j\left(h_0\right)q_n\left(U_j\left(h_0\right),U_{jk}\left(h_0\right),U_{jkl}\left(h_0\right)\right),$$

where q_n is a polynomial in $U_j\left(h_0\right)$, $U_{jk}\left(h_0\right)$ and $U_{jkl}\left(h_0\right)$. Furthermore (45) and the validity of (42) imply the validity of the Edgeworth expansion (43) in the sense of Chandra and Ghosh (1980), that is

$$\sup_{c\in[c_0,\infty)}\left|P\left(S_{\gamma}\left(\widehat{\lambda},h_0,\right)\leq c\right)-G_k\left(c\right)-\int_{u^Tu\leq c}p_{2n}\left(u\right)\phi\left(u\right)\right|=O\left(\frac{1}{\left(nb^d\right)^2}\right),$$

which in turn implies the validity of (13). ■

Proof of Corollary 1.1. By the recurrence relation

$$G_{k+2}\left(c\right)=G_k\left(c\right)-\frac{2}{k}cg_k\left(c\right)\tag{46}$$

we have that

$$P\left(S_{EL}\left(\widehat{\lambda},h_0\right)\leq c_{\alpha}\right)=1-\alpha-\frac{b_{1EL}\left(K\right)c_{\alpha}}{\left(nb^d\right)k}g_k\left(c_{\alpha}\right)+O\left(\frac{1}{\left(nb^d\right)^2}\right),$$

so that by a straightforward application of the delta method

$$P\left(\frac{S_{EL}\left(\widehat{\lambda},h_0\right)}{BC\left(K\right)}\leq c_{\alpha}\right)=1-\alpha+O\left(\frac{1}{\left(nb^d\right)^2}\right).$$

■

Proof of Corollary 1.2. By (46) and noting that $\sum_{j=0}^{3} B_j(K) = 0$, it follows that expansion (13) in the main text can be written as

$$P\left(S_\gamma\left(\widehat{\lambda}, h_0\right) \leq c_\alpha\right) = 1 - \alpha - g_k(c_\alpha) \sum_{j=1}^{3} B_j'(K)(c_\alpha)^j + O_p\left(\frac{1}{(nb^d)^2}\right)$$

where

$$B_j'(K) = \frac{2\Gamma\left(\frac{1}{2k}\right)\sum_{l=j}^{3} B_l(K)}{2^j \Gamma\left(\frac{1}{2k} + j\right)},$$

and $\Gamma(\cdot)$ is the gamma function. The conclusion follows using the results of Cox and Reid (1987b) applied to the modified statistic $S_\gamma^m\left(\widehat{\lambda}, h_0\right)$. ■

Proof of Corollary 1.3. Let $\widehat{\omega}\left(\widehat{h}\right)$ denote the sample analog of $\omega(h_0)$. By the triangle inequality, a mean value expansion, the LLN (which hold by **A2'**(ii)-(iii)), we have for $j, l, m = 1, ..., k$

$$\begin{aligned}
&\left|\widehat{\omega}^{jl}\left(\widehat{h}\right)^{1/2} - \omega^{jl}(h_0)^{1/2}\right| \leq \left|\widehat{\omega}^{jl}\left(\widehat{h}\right)^{1/2} - \widehat{\omega}^{jl}(h_0)^{1/2}\right| + \left|\widehat{\omega}^{jl}(h_0)^{1/2} - \omega^{jl}(h_0)^{1/2}\right| \leq \\
&\inf_{h\in\mathcal{H}_C} \sigma_{\min}\left(\widehat{\omega}_{jl}(h)\right)^{-1/2} \sup_{h\in\mathcal{H}_C} \frac{1}{nb^d} \sum \left|\frac{\partial g_{ij}^K(h)\, g_{il}^K(h)}{\partial h_m}\right| \left|\widehat{h}_m - h_{0m}\right| + \\
&\frac{\sigma_{\min}\left(\omega_{jl}(h_0)\right)^{-3/2}}{2} \left|\frac{\Delta(h_0)}{(nb^d)^{1/2}}\right| + O_p\left(\left(nb^d\right)^{-1}\right) = O_p\left((nb)^{-1/2}\right)
\end{aligned} \tag{47}$$

where $\Delta(h_0) = \left(\sum g_{ij}^K(h_0)\, g_{il}^K(h_0) - E\left[g_j^K(h_0)\, g_l^K(h_0)\right]\right) / \left(nb^d\right)^{1/2} = O_p(1)$ by the CLT. Let $\widehat{\mu}_{jlm}$ denote the sample analog of μ_{jlm}; then as in (47)

$$\begin{aligned}
&\left|\widehat{\mu}_{jlm} - \mu_{jlm}\right| \leq \left|\widehat{\omega}^{jl}\left(\widehat{h}\right)^{1/2} - \omega^{jl}(h_0)^{1/2}\right|^3 \sup_{h\in\mathcal{H}_C} \frac{1}{nb^d} \sum \left|g_{ij}^K(h)\right|^3 + \\
&\inf_{h\in\mathcal{H}_C} \sigma_{\min}\left(\omega_{jl}(h)\right)^{3/2} \left(\sup_{h\in\mathcal{H}_C} \frac{1}{nb^d} \sum \left|\frac{\partial g_{ij}^K(h)\, g_{il}^K(h)\, g_{im}^K(h)}{\partial h_n}\right| \left|\widehat{h}_n - h_{0n}\right| + \right. \\
&\left.\left|\frac{1}{nb^d} \sum g_{ij}^K(h_0) - E\left[\frac{g_j^K(h_0)}{b^d}\right]\right|^3\right) = O_p\left(\left(nb^d\right)^{-3/2}\right) + O_p\left(\left(nb^d\right)^{-1/2}\right) + O_p\left(\left(nb^d\right)^{-3/2}\right).
\end{aligned}$$

Similarly, it can be shown that $\widehat{\mu}_{jlmn} = \mu_{jlmn} + O_p\left((nb)^{-1/2}\right)$, thus by the delta method $P(S_\gamma^{\widehat{m}}(\widehat{\lambda}, h_0) \leq c_\alpha) = P(S_\gamma^m(\widehat{\lambda}, h_0) \leq c_\alpha) + O(1/(nb^d)^{3/2}$ and the first conclusion follows by the oddness/eveness property of the Hermite polynomials (Barndorff-Nielsen and Hall (1988)). A similar argument can be used to prove the second conclusion. ■

Proof of Theorem 2. Recall that the indices $a, b, ...$ run from 1 to k, $j, k, ..$ run from 1 to l and $r, s, ...$ run from 1 to $l-k$, and recall also that

$$G_{jkl..}^{a_1...a_{m_1},b_1...b_{m_2}}(h) = \frac{1}{(nb^d)^{1/2}} \sum \left(\frac{\partial^{m_1} U_{ij}(h)}{\partial h_{a_1} ... \partial h_{a_{m_1}}} \frac{\partial^{m_2} U_{ik}(h)}{\partial h_{b_1} ... \partial h_{b_{m_2}}} U_{il}(h) - \gamma(h)_{jkl..}^{a_1...a_{m_1},b_1...b_{m_2}} \right); \tag{48}$$

then the same Taylor expansion as that given in (38) combined with a further Taylor expansion about h_0 yields

$$\begin{aligned}
&S_\gamma\left(\widehat{\lambda}, \widehat{h}\right) = 2\left(-\sum U_{ij}\left(\widehat{h}\right)\widehat{\lambda}_j - \frac{1}{2}\sum\left(U_{ij}\left(\widehat{h}\right)\widehat{\lambda}_j\right)^2 + \frac{\gamma_3}{3!}\sum\left(U_{ij}\left(\widehat{h}\right)\widehat{\lambda}_j\right)^3 + \right.\\
&\left.\frac{\gamma_4}{4!}\sum\left(U_{ij}\left(\widehat{h}\right)\widehat{\lambda}_j\right)^4\right) + O_p\left(\frac{1}{(nb^d)^{3/2}}\right) = \\
&2\left\{-\widehat{\lambda}_j\left[\sum U_{ij}(h_0) + \sum \frac{\partial U_{ij}(h_0)}{\partial h_a}\left(\widehat{h}-h_0\right)_a + \right.\right.\\
&\frac{1}{2}\sum \frac{\partial^2 U_{ij}(h_0)}{\partial h_a \partial h_b}\left(\widehat{h}-h_0\right)_a\left(\widehat{h}-h_0\right)_b + \\
&\left.\frac{1}{3!}\sum \frac{\partial^3 U_{ij}(h_0)}{\partial h_a \partial h_b \partial h_c}\left(\widehat{h}-h_0\right)_a\left(\widehat{h}-h_0\right)_b\left(\widehat{h}-h_0\right)_c\right] - \\
&\frac{\widehat{\lambda}_j\widehat{\lambda}_k}{2}\left[\sum U_{ij}(h_0)U_{ik}(h_0) + [2]\sum\frac{\partial U_{ij}(h_0)}{\partial h_a}U_{ik}(h_0)\left(\widehat{h}-h_0\right)_a + \right.\\
&\frac{[2]}{2}\sum\left(\frac{\partial^2 U_{ij}(h_0)}{\partial h_a \partial h_b}U_{ik}(h_0) + \frac{\partial U_{ij}(h_0)}{\partial h_a}\frac{\partial U_{ik}(h_0)}{\partial h_b}\right)\times\\
&\left.\left(\widehat{h}-h_0\right)_a\left(\widehat{h}-h_0\right)_b\right] + \frac{\gamma_3}{3!}\widehat{\lambda}_j\widehat{\lambda}_k\widehat{\lambda}_l\sum\left[U_{ij}(h_0)U_{ik}(h_0)U_{il}(h_0) + \right.\\
&\left.[3]\frac{\partial U_{ij}(h_0)}{\partial h_a}U_{ik}(h_0)U_{il}(h_0)\left(\widehat{h}-h_0\right)_a\right] + \\
&\left.\frac{\gamma_4}{4!}\widehat{\lambda}_j\widehat{\lambda}_k\widehat{\lambda}_l\widehat{\lambda}_m\sum U_{ij}(h_0)U_{ik}(h_0)U_{il}(h_0)U_{im}(h_0)\right\} + O_p\left(\frac{1}{(nb^d)^{3/2}}\right).
\end{aligned}$$

By (48) it then follows that

$$S_\gamma\left(\widehat{\lambda},\widehat{h}\right)=2\left\{-\left(nb^d\right)^{1/2}\widehat{\lambda}_j\left[U_j\left(h_0\right)+\frac{G_j^a\left(h_0\right)}{\left(nb^d\right)^{1/2}}\left(nb^d\right)^{1/2}\left(\widehat{h}-h_0\right)_a+\right.\right. \tag{49}$$
$$\gamma_j^a\left(h_0\right)\left(nb\right)^{1/2}\left(\widehat{h}-h_0\right)_a+\frac{G_j^{ab}\left(h_0\right)}{2nb^d}\left(nb^d\right)\left(\widehat{h}-h_0\right)_a\left(\widehat{h}-h_0\right)_b+$$
$$\frac{\gamma_j^{ab}\left(h_0\right)}{2\left(nb^d\right)^{1/2}}\left(nb^d\right)\left(\widehat{h}-h_0\right)_a\left(\widehat{h}-h_0\right)_b+$$
$$\left.\frac{\gamma_j^{abc}\left(h_0\right)}{3!nb^d}\left(nb^d\right)^{3/2}\left(\widehat{h}-h_0\right)_a\left(\widehat{h}-h_0\right)_b\left(\widehat{h}-h_0\right)_c\right]-$$
$$\left(nb^d\right)\frac{\widehat{\lambda}_j\widehat{\lambda}_k}{2}\left[\frac{U_{jk}\left(h_0\right)}{\left(nb^d\right)^{1/2}}+\delta_{mn}+[2]\frac{G_{jk}^a\left(h_0\right)}{nb^d}\left(nb^d\right)^{1/2}\left(\widehat{h}-h_0\right)_a+[2]\frac{\gamma_{jk}^a\left(h_0\right)}{\left(nb^d\right)^{1/2}}\times\right.$$
$$\left.\left(nb^d\right)^{1/2}\left(\widehat{h}-h_0\right)_a+\frac{[2]}{2}\left(\frac{\gamma_{jk}^{ab}\left(h_0\right)}{nb^d}+\frac{\gamma_{jk}^{a,b}\left(h_0\right)}{nb^d}\right)\left(nb^d\right)\left(\widehat{h}-h_0\right)_a\left(\widehat{h}-h_0\right)_b\right]+$$
$$\frac{\gamma_3}{3!}\left(nb^d\right)^{3/2}\widehat{\lambda}_j\widehat{\lambda}_k\widehat{\lambda}_l\left(\frac{U_{jkl}\left(h_0\right)}{nb^d}+\frac{\mu_{jkl}\left(h_0\right)}{\left(nb^d\right)^{1/2}}+[3]\frac{\gamma_{jkl}^a\left(h_0\right)}{nb^d}\left(nb^d\right)^{1/2}\left(\widehat{h}-h_0\right)_a\right)+$$
$$\left.\frac{\gamma_4}{4!}\left(nb^d\right)^2\widehat{\lambda}_j\widehat{\lambda}_k\widehat{\lambda}_l\widehat{\lambda}_m\frac{\mu_{jklm}\left(h_0\right)}{nb^d}\right\}+O_p\left(\frac{1}{\left(nb^d\right)^{3/2}}\right).$$

In order to simultaneously consider derivatives with respect to h and λ we introduce the Greek letters indices $\alpha,\beta,...$, which run from from 1 to $l+k$; let

$$\begin{aligned}\sigma_{\alpha\beta}\left(h\right) &= E\left(\frac{\partial^2\Gamma_\gamma\left(\lambda,h\right)}{\partial\eta_\alpha\partial\eta_\beta}\right)|_{\lambda=0}= \\ & \left[\begin{array}{cc}-\delta_{jk} & -\gamma_j^a\left(h\right)\\ -\gamma_j^a\left(h\right) & O_{ab}\end{array}\right]\end{aligned}$$

where $\eta_\alpha=\left[\lambda_j,\left(h-h_0\right)_a\right]$ and

$$\begin{aligned}v_{\alpha_1\alpha_3\ldots\alpha_k}\left(h\right) &= E\left(\frac{\sigma^{\alpha_1\alpha_2}\left(h\right)\partial^k\Gamma_\gamma\left(\lambda,h\right)}{\partial\eta_{\alpha_2\ldots\alpha_k}}\right)|_{\lambda=0}, \\ V_{\alpha_1\alpha_3\ldots\alpha_k}\left(\lambda,h\right) &= \left(nb^d\right)^{1/2}\left(\frac{\sigma^{\alpha_1\alpha_2}\left(h\right)\partial^k\widehat{\Gamma}_\gamma\left(\lambda,h\right)}{\partial\eta_{\alpha_2\ldots\alpha_k}}-v_{\alpha_1\alpha_3\ldots.\alpha_k}\right),\ k\geq 2,\end{aligned}$$

where $\sigma^{\alpha\beta}(h)$ denotes the matrix inverse of $\sigma_{\alpha\beta}(h)$. Note that

$$\begin{aligned}
\frac{\partial^3\Gamma_\gamma(\lambda,h)}{\partial\lambda_j\partial\lambda_k\partial\lambda_l} &= \gamma_3(v(z))\,g_j(h)\,g_k(h)\,g_l(h)\,,\\
\frac{\partial^3\Gamma_\gamma(\lambda,h)}{\partial\lambda_j\partial\lambda_k\partial h_a} &= [2]\,\gamma_2(v(z))\,\frac{\partial g_j(h)}{\partial h_a}g_k(h)+\gamma_3(v(z))\,g_j(h)\,g_k(h)\,\frac{\partial g_l(h)}{\partial h_a}\lambda_l,\\
\frac{\partial^3\Gamma_\gamma(\lambda,h)}{\partial\lambda_j\partial h_a\partial h_b} &= \gamma_1(v(z))\,\frac{\partial^2 g_j(h)}{\partial h_a\partial h_b}+[2]\,\gamma_2(v(z))\,\frac{\partial g_j(h)}{\partial h_a}\frac{\partial g_k(h)}{\partial h_b}\lambda_k+\\
&\quad \gamma_2(v(z))\,g_j(h)\,\frac{\partial^2 g_k(h)}{\partial h_a\partial h_b}\lambda_k+\gamma_3(v(z))\,g_j(h)\,\frac{\partial g_k(h)}{\partial h_a}\lambda_k\frac{\partial g_l(h)}{\partial h_b}\lambda_l,\\
\frac{\partial^3\Gamma_\gamma(\lambda,h)}{\partial h_a\partial h_b\partial h_c} &= \gamma_1(v(z))\,\frac{\partial^3 g_j(h)}{\partial h_a\partial h_b\partial h_c}\lambda_j+[3]\,\gamma_2(v(z))\,\frac{\partial^2 g_j(h)}{\partial h_a\partial h_b}\lambda_j\frac{\partial g_k(h)}{\partial h_c}\lambda_k+\\
&\quad \gamma_3(v(z))\,\frac{\partial g_j(h)}{\partial h_a}\lambda_j\frac{\partial g_k(h)}{\partial h_b}\lambda_k\frac{\partial g_l(h)}{\partial h_c}\lambda_l,
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^4\Gamma_\gamma(\lambda,h)}{\partial\lambda_j\partial\lambda_k\partial\lambda_l\partial\lambda_m} &= \gamma_4(v(z))\,g_j(h)\,g_k(h)\,g_l(h)\,g_m(h)\,,\\
\frac{\partial^4\Gamma_\gamma(\lambda,h)}{\partial\lambda_j\partial\lambda_k\partial\lambda_l\partial h_a} &= [3]\,\gamma_3(v(z))\,g_j(h)\,g_k(h)\,\frac{\partial g_l(h)}{\partial h_a}+\gamma_4(v(z))\,g_j(h)\,g_k(h)\,g_l(h)\,\frac{\partial g_m(h)}{\partial h_a}\lambda_m,\\
\frac{\partial^4\Gamma_\gamma(\lambda,h)}{\partial\lambda_j\partial\lambda_k\partial h_a\partial h_b} &= [2]\left(\gamma_2(v(z))\,\frac{\partial g_j(h)}{\partial h_a}\frac{\partial g_k(h)}{\partial h_b}+\gamma_2(v(z))\,\frac{\partial^2 g_j(h)}{\partial h_a\partial h_b}g_k(h)+\right.\\
&\quad \left.\gamma_3(v(z))\,\frac{\partial g_j(h)}{\partial h_a}g_k(h)\,\frac{\partial g_l(h)}{\partial h_b}\lambda_l\right)+\\
&\quad \gamma_3(v(z))\left([2]\,\frac{\partial g_j(h)}{\partial h_a}g_k(h)\,\frac{\partial g_l(h)}{\partial h_b}\lambda_l+g_j(h)\,g_k(h)\,\frac{\partial^2 g_l(h)}{\partial h_a\partial h_b}\lambda_l\right)+\\
&\quad \gamma_4(v(z))\,g_j(h)\,g_k(h)\,\frac{\partial g_l(h)}{\partial h_a}\lambda_l\frac{\partial g_m(h)}{\partial h_b}\lambda_m,
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial^4 \Gamma_\gamma(\lambda,h)}{\partial\lambda_j\partial h_a\partial h_b\partial h_c} &= \gamma_1(v(z))\frac{\partial^3 g_j(h)}{\partial h_a\partial h_b\partial h_c} + [6]\,\gamma_2(v(z))\frac{\partial^2 g_j(h)}{\partial h_a\partial h_b}\frac{\partial g_k(h)}{\partial h_c}\lambda_k + \\
&\gamma_2(v(z))\,g_j(h)\frac{\partial^3 g_k(h)}{\partial h_a\partial h_b\partial h_c}\lambda_k + [3]\,\gamma_3(v(z))\frac{\partial g_j(h)}{\partial h_a}\frac{\partial g_k(h)}{\partial h_b}\lambda_k\frac{\partial g_l(h)}{\partial h_c}\lambda_l + \\
&[3]\,\gamma_3(v(z))\,g_j(h)\frac{\partial^2 g_k(h)}{\partial h_a\partial h_b}\lambda_k\frac{\partial g_l(h)}{\partial h_c}\lambda_l + \\
&\gamma_4(v(z))\,g_j(h)\frac{\partial g_k(h)}{\partial h_b}\lambda_k\frac{\partial g_l(h)}{\partial h_c}\lambda_l\frac{\partial g_m(h)}{\partial h_b}\lambda_m, \\
\frac{\partial^4 \Gamma_\gamma(\lambda,h)}{\partial h_a\partial h_b\partial h_c\partial h_d} &= \gamma_1(v(z))\frac{\partial^4 g_j(h)}{\partial h_a\partial h_b\partial h_c\partial h_d}\lambda_j + \gamma_2(v(z))\frac{\partial^3 g_j(h)}{\partial h_a\partial h_b\partial h_c}\lambda_j\frac{\partial g_k(h)}{\partial h_d}\lambda_k + \\
&[6]\,\gamma_2(v(z))\frac{\partial^2 g_j(h)}{\partial h_a\partial h_b}\lambda_j\frac{\partial^2 g_k(h)}{\partial h_c\partial h_d}\lambda_k + \\
&[6]\,\gamma_3(v(z))\frac{\partial^2 g_j(h)}{\partial h_a\partial h_b}\lambda_j\frac{\partial g_k(h)}{\partial h_c}\lambda_k\frac{\partial g_l(h)}{\partial h_d}\lambda_l + \\
&\gamma_4(v(z))\frac{\partial g_j(h)}{\partial h_a}\lambda_j\frac{\partial g_k(h)}{\partial h_b}\lambda_k\frac{\partial g_l(h)}{\partial h_c}\lambda_l\frac{\partial g_m(h)}{\partial h_d}\lambda_m,
\end{aligned}
$$

so that

$$
V_\alpha(0,h) = \left[O_a, U_{k+r}(h), \Delta^{ab}U_b(h)\right], \tag{50}
$$

$$
V_{\alpha\beta}(0,h) = \begin{bmatrix} \Delta^{ab}G_c^b(h) & \Delta^{ab}G_{k+r}^b(h) & O_{ab} \\ U_{k+ra}(h) & U_{k+rk+s}(h) & -G_{k+r}^a(h) \\ \Delta^{ab}\Delta^{cd}G_b^d(h) - \Delta^{ab}U_{bc}(h) & \Delta^{ab}\Delta^{cb}G_{k+r}^c(h) - \Delta^{ab}U_{bk+r}(h) & \Delta^{ab}G_c^b(h) \end{bmatrix}, \tag{51}
$$

for $j = 1, ..., l$

$$
V_{j\alpha\beta}(0,h) = \begin{bmatrix} [2]\,\Delta^{ac}G_{jb}^c(h) & [2]\,\Delta^{ac}G_{jk+r}^c(h) & \Delta^{ab}G_j^{bc}(h) \\ -\gamma_3 U_{jk+ra}(h) & -\gamma_3 U_{jk+rk+s}(h) & [2]\,G_{jk+r}^a(h) \\ \gamma_3\left(-\Delta^{ab}U_{jbc}(h) + [2]\,\Delta^{ab}\Delta^{cd}G_{jb}^d(h)\right) & V_{jk+ra}(0,h) & V_{jab}(0,h) \end{bmatrix}, \tag{52}
$$

where

$$
\begin{aligned}
V_{jk+ra}(0,h) &= \gamma_3\left(-\Delta^{ab}U_{jk+rb}(h) + [2]\,\Delta^{ab}\Delta^{bc}G_{jk+r}^c(h)\right), \\
V_{jab}(0,h) &= -\Delta^{ac}\Delta^{bd}G_{jd}^c(h) + [2]\,\Delta^{ab}G_j^{bc}(h)
\end{aligned}
$$

and for l fixed and $a = 1, ..., k$

$$
V_{l+a\alpha\beta}(0,h) = \begin{bmatrix} \Delta^{ac}G_b^{cd}(h) & \Delta^{ac}G_{k+r}^{cd}(h) & O_{ab} \\ [2]\,G_{k+rb}^a(h) & [2]\,G_{k+rk+s}^a(h) & G_{k+r}^{ab}(h) \\ -\Delta^{ac}\Delta^{bd}G_b^{cd}(h) - [2]\,\Delta^{ad}G_{bc}^d(h) & V_{l+ak+rb}(0,h) & \Delta^{ad}G_d^{bc}(h) \end{bmatrix}, \tag{53}
$$

where

$$
V_{l+ak+rb}(0,h) = -\Delta^{ac}\Delta^{bd}G_{k+r}^{cd}(h) + [2]\,\gamma_2\Delta^{ad}G_{k+rc}^d(h).
$$

By expanding the FOCs

$$0=\frac{\sigma^{\alpha\beta}\left(h_0\right)\partial\widehat{\Gamma}_{\gamma}\left(\widehat{\eta}\right)}{\partial\eta_{\beta}},$$

solving for $\left(nb^d\right)^{1/2}\left(\widehat{\eta}-\eta_0\right)_a$ and noting that $v_{\alpha\beta}\left(0,h_0\right)=\delta_{\alpha\beta}$, we have

$$\begin{aligned}
&\left(nb^d\right)^{1/2}\left(\widehat{\eta}-\eta_0\right)_{\alpha}=-V_{\alpha}\left(0,h_0\right)+\frac{1}{\left(nb^d\right)^{1/2}}\left(V_{\alpha\beta}\left(0,h_0\right)V_{\beta}\left(0,h_0\right)-\right.\\
&\left.\frac{v_{\alpha\beta\gamma}\left(h_0\right)}{2}V_{\beta}\left(0,h_0\right)V_{\gamma}\left(0,h_0\right)\right)+\\
&\frac{1}{nb^d}\left(-V_{\alpha\beta}\left(0,h_0\right)V_{\beta\gamma}\left(0,h_0\right)V_{\gamma}\left(0,h_0\right)+\frac{v_{\beta\gamma\delta}\left(h_0\right)}{2}V_{\alpha\beta}\left(0,h_0\right)V_{\gamma}\left(0,h_0\right)V_{\delta}\left(0,h_0\right)+\right.\\
&v_{\alpha\beta\gamma}\left(h_0\right)V_{\beta\delta}\left(0,h_0\right)V_{\gamma}\left(0,h_0\right)V_{\delta}\left(0,h_0\right)-\\
&\frac{v_{\alpha\beta\gamma}\left(h_0\right)v_{\beta\delta\varepsilon}\left(h_0\right)}{2}V_{\gamma}\left(0,h_0\right)V_{\delta}\left(0,h_0\right)V_{\varepsilon}\left(0,h_0\right)-\frac{1}{2}V_{\alpha\beta\gamma}\left(0,h_0\right)V_{\beta}\left(0,h_0\right)V_{\gamma}\left(0,h_0\right)+\\
&\left.\frac{v_{\alpha\beta\gamma\delta}\left(h_0\right)}{6}V_{\beta}\left(0,h_0\right)V_{\gamma}\left(0,h_0\right)V_{\delta}\left(0,h_0\right)\right)+O_p\left(\frac{1}{\left(nb^d\right)^{3/2}}\right),
\end{aligned} \tag{54}$$

Using (54) in (49) yields

$$
\begin{aligned}
&S_\gamma\left(\widehat{\lambda},\widehat{h}\right)=2\left\{-\left[-V_j\left(0,h_0\right)+\frac{1}{\left(nb^d\right)^{1/2}}\left(V_{j\alpha}\left(0,h_0\right)V_\alpha\left(0,h_0\right)-\right.\right.\right. \qquad (55)\\
&\left.-\frac{v_{j\alpha\beta}\left(h_0\right)}{2}V_\alpha\left(0,h_0\right)V_\beta\left(0,h_0\right)\right)+\\
&\frac{1}{nb^d}\left(-V_{j\alpha}\left(0,h_0\right)V_{\alpha\beta}\left(0,h_0\right)V_\beta\left(0,h_0\right)+\frac{v_{\alpha\beta\gamma}\left(h_0\right)}{2}V_{j\alpha}\left(0,h_0\right)V_\beta\left(0,h_0\right)V_\gamma\left(0,h_0\right)+\right.\\
&v_{j\alpha\beta}\left(h_0\right)V_{\alpha\gamma}\left(0,h_0\right)V_\beta\left(0,h_0\right)V_\gamma\left(0,h_0\right)-\frac{v_{j\alpha\beta}\left(h_0\right)v_{\alpha\gamma\delta}\left(h_0\right)}{2}V_\beta\left(0,h_0\right)V_\gamma\left(0,h_0\right)V_\delta\left(0,h_0\right)-\\
&\left.\left.\frac{1}{2}V_{j\alpha\beta}\left(0,h_0\right)V_\alpha\left(0,h_0\right)V_\beta\left(0,h_0\right)+\frac{v_{j\alpha\beta\gamma}\left(h_0\right)}{6}V_\alpha\left(0,h_0\right)V_\beta\left(0,h_0\right)V_\gamma\left(0,h_0\right)\right]\times\\
&\left\{U_j\left(h_0\right)+\left(\frac{G_j^a\left(h_0\right)}{\left(nb^d\right)^{1/2}}+\gamma_j^a\left(h_0\right)\right)\left[-V_{l+a}\left(0,h_0\right)+\right.\right.\\
&\frac{1}{\left(nb^d\right)^{1/2}}\left(V_{l+a\beta}\left(0,h_0\right)V_\beta\left(0,h_0\right)-\frac{v_{l+a\beta\gamma}\left(h_0\right)}{2}V_\beta\left(0,h_0\right)V_\gamma\left(0,h_0\right)\right)+\\
&\frac{1}{nb^d}\left(-V_{l+a\beta}\left(0,h_0\right)V_{\beta\gamma}\left(0,h_0\right)V_\gamma\left(0,h_0\right)+\frac{v_{\alpha\beta\gamma}\left(h_0\right)}{2}V_{l+a\alpha}\left(0,h_0\right)V_\beta\left(0,h_0\right)V_\gamma\left(0,h_0\right)+\right.\\
&v_{l+a\alpha\beta}\left(h_0\right)V_{\alpha\gamma}\left(0,h_0\right)V_\beta\left(0,h_0\right)V_\gamma\left(0,h_0\right)-\frac{v_{l+a\alpha\beta}\left(h_0\right)v_{\alpha\gamma\delta}\left(h_0\right)}{2}V_\beta\left(0,h_0\right)V_\gamma\left(0,h_0\right)V_\delta\left(0,h_0\right)-\\
&\left.\left.\frac{1}{2}V_{l+a\alpha\beta}\left(0,h_0\right)V_\alpha\left(0,h_0\right)V_\beta\left(0,h_0\right)+\frac{v_{l+a\alpha\beta\gamma}\left(h_0\right)}{6}V_\alpha\left(0,h_0\right)V_\beta\left(0,h_0\right)V_\gamma\left(0,h_0\right)\right)\right]+\\
&\frac{G_j^{ab}\left(h_0\right)}{2nb^d}\left(-V_{l+a}\left(0,h_0\right)+\frac{1}{\left(nb^d\right)^{1/2}}\left(V_{l+a\alpha}\left(0,h_0\right)V_\beta\left(0,h_0\right)-\frac{v_{l+a\alpha\beta}\left(h_0\right)}{2}V_\alpha\left(0,h_0\right)V_\beta\left(0,h_0\right)\right)\right)\times\\
&\left(-V_{l+b}\left(0,h_0\right)+\frac{1}{\left(nb^d\right)^{1/2}}\left(V_{l+b\gamma}\left(0,h_0\right)V_\gamma\left(0,h_0\right)-\frac{v_{l+b\gamma\delta}\left(h_0\right)}{2}V_\gamma\left(0,h_0\right)V_\delta\left(0,h_0\right)\right)\right)+\\
&\left.\gamma_j^{ab}\left(h_0\right)\frac{V_{l+a}\left(0,h_0\right)V_{l+b}\left(0,h_0\right)}{2\left(nb^d\right)^{1/2}}+\gamma_j^{abc}\left(h_0\right)\frac{V_{l+a}\left(0,h_0\right)V_{l+b}\left(0,h_0\right)V_{l+c}\left(0,h_0\right)}{3!nb^d}\right\}\\
&-\frac{1}{2}\left(-V_j\left(h_0\right)+\frac{1}{\left(nb^d\right)^{1/2}}\left(V_{j\alpha}\left(0,h_0\right)V_\alpha\left(0,h_0\right)-\frac{v_{j\alpha\beta}\left(h_0\right)}{2}V_\alpha\left(0,h_0\right)V_\beta\left(0,h_0\right)\right)+\right.\\
&\frac{1}{nb^d}\left(-V_{j\alpha}\left(0,h_0\right)V_{\beta\alpha}\left(0,h_0\right)V_\beta\left(0,h_0\right)+\frac{v_{\alpha\beta\gamma}\left(h_0\right)}{2}V_{j\alpha}\left(0,h_0\right)V_\beta\left(0,h_0\right)V_\gamma\left(0,h_0\right)+\right.\\
&v_{j\alpha\beta}\left(h_0\right)V_{\alpha\gamma}\left(0,h_0\right)V_\beta\left(0,h_0\right)V_\gamma\left(0,h_0\right)-\frac{v_{j\alpha\beta}\left(h_0\right)v_{\alpha\gamma\delta}\left(h_0\right)}{2}V_\beta\left(0,h_0\right)V_\gamma\left(0,h_0\right)V_\delta\left(0,h_0\right)-\\
&\left.\frac{1}{2}V_{j\alpha\beta\gamma}\left(0,h_0\right)V_\beta\left(0,h_0\right)V_\gamma\left(0,h_0\right)+\frac{v_{j\alpha\beta\gamma}\left(h_0\right)}{6}V_\alpha\left(0,h_0\right)V_\beta\left(0,h_0\right)V_\gamma\left(0,h_0\right)\right)\times
\end{aligned}
$$

$$
\left(-V_k\left(0,h_0\right)+\frac{1}{\left(nb^d\right)^{1/2}}\left(V_{k\delta}\left(0,h_0\right)V_\zeta\left(0,h_0\right)-\frac{v_{k\delta\zeta}\left(h_0\right)}{2}V_\delta\left(0,h_0\right)V_\zeta\left(0,h_0\right)\right)+\right.
$$

$$
\frac{1}{nb^d}\left(-V_{k\delta}\left(0,h_0\right)V_{\delta\zeta}\left(0,h_0\right)V_\zeta\left(0,h_0\right)+\frac{v_{\delta\zeta\eta}\left(h_0\right)}{2}V_{k\delta}\left(0,h_0\right)V_\zeta\left(0,h_0\right)V_\eta\left(0,h_0\right)+\right.
$$

$$
v_{k\delta\zeta}\left(h_0\right)V_{\delta\eta}\left(0,h_0\right)V_\zeta\left(0,h_0\right)V_\eta\left(0,h_0\right)-\frac{v_{k\delta\zeta}\left(h_0\right)v_{\delta\eta\xi}\left(h_0\right)}{2}V_\zeta\left(0,h_0\right)V_\eta\left(0,h_0\right)V_\xi\left(0,h_0\right)+
$$

$$
v_{k\delta\zeta}\left(h_0\right)V_{\delta\xi}\left(0,h_0\right)V_\zeta\left(0,h_0\right)V_\xi\left(0,h_0\right)-\frac{v_{k\delta\zeta}\left(h_0\right)v_{\delta\xi\varsigma}\left(h_0\right)}{2}V_\zeta\left(0,h_0\right)V_\xi\left(0,h_0\right)V_\varsigma\left(0,h_0\right)+
$$

$$
\left.\frac{v_{k\delta\zeta\xi}\left(h_0\right)}{6}V_\delta\left(0,h_0\right)V_\zeta\left(0,h_0\right)V_\xi\left(0,h_0\right)\right)\times\left[\frac{U_{jk}\left(h_0\right)}{\left(nb^d\right)^{1/2}}+\delta_{jk}+\right.
$$

$$
[2]\,\frac{G^e_{jk}\left(h_0\right)V_{l+e}\left(0,h_0\right)}{nb^d}+[2]\,\frac{\gamma^e_{jk}\left(h_0\right)}{\left(nb^d\right)^{1/2}}\left(-V_{l+e}\left(0,h_0\right)+\right.
$$

$$
\left.\left.\frac{1}{\left(nb^d\right)^{1/2}}\left(V_{l+e\kappa}\left(0,h_0\right)V_\kappa\left(0,h_0\right)-\frac{v_{l+e\kappa\mu}\left(h_0\right)}{2}V_\kappa\left(0,h_0\right)V_\mu\left(0,h_0\right)\right)\right)+
$$

$$
\left.\frac{[2]}{2}\left(\frac{\gamma^{ab}_{jk}\left(h_0\right)}{nb^d}+\frac{\gamma^{a,b}_{jk}\left(h_0\right)}{nb^d}\right)V_{l+a}\left(0,h_0\right)V_{l+b}\left(0,h_0\right)\right]+
$$

$$
\frac{\gamma_3}{3!}\left(-V_j\left(0,h_0\right)+\frac{1}{\left(nb^d\right)^{1/2}}\left(V_{j\kappa}\left(0,h_0\right)V_\kappa\left(0,h_0\right)-\frac{v_{j\kappa\mu}\left(h_0\right)}{2}V_\kappa\left(0,h_0\right)V_\mu\left(0,h_0\right)\right)\right)\times
$$

$$
\left(-V_k\left(h_0\right)+\frac{1}{\left(nb^d\right)^{1/2}}\left(V_{k\kappa}\left(0,h_0\right)V_\kappa\left(0,h_0\right)-\frac{v_{k\kappa\tau}\left(h_0\right)}{2}V_\kappa\left(0,h_0\right)V_\tau\left(0,h_0\right)\right)\right)\times
$$

$$
\left(-V_l\left(h_0\right)+\frac{1}{\left(nb^d\right)^{1/2}}\left(V_{o\varepsilon}\left(0,h_0\right)V_\varepsilon\left(0,h_0\right)-\frac{v_{m\varepsilon\zeta}\left(h_0\right)}{2}V_\varepsilon\left(0,h_0\right)V_\zeta\left(0,h_0\right)\right)\right)\times
$$

$$
\left(\frac{\mu_{jkl}\left(h_0\right)}{\left(nb^d\right)^{1/2}}+[3]\,\frac{\gamma^e_{jkl}\left(h_0\right)}{nb^d}V_{l+e}\left(0,h_0\right)\right)+
$$

$$
\left.\frac{\gamma_4}{4!}\frac{\mu_{jklm}\left(h_0\right)V_j\left(h_0\right)V_k\left(h_0\right)V_l\left(h_0\right)V_m\left(h_0\right)}{nb^d}\right\}+O_p\left(\frac{1}{\left(nb^d\right)^{3/2}}\right).
$$

Using (50)-(53), the signed squared root decomposition is

$$
S_{\gamma k+r}\left(\widehat{\lambda},\widehat{h}\right)=S^1_{\gamma k+r}\left(\widehat{\lambda},\widehat{h}\right)+S^2_{\gamma k+r}\left(\widehat{\lambda},\widehat{h}\right)+S^3_{\gamma k+r}\left(\widehat{\lambda},\widehat{h}\right)+O_p\left(\frac{1}{\left(nb^d\right)^{3/2}}\right)
$$

where

$$S^1_{\gamma k+r}\left(\widehat{\lambda},\widehat{h}\right)=U_{k+r}\left(h_0\right),\quad S^2_{\gamma k+r}\left(\widehat{\lambda},\widehat{h}\right)=\frac{1}{\left(nb^d\right)^{1/2}}\left(\frac{1}{2}U_{k+rk+s}\left(h_0\right)U_{k+s}\left(h_0\right)-\right.$$
$$\frac{\gamma_3}{6}\mu_{k+rk+sk+t}\left(h_0\right)U_{k+s}\left(h_0\right)U_{k+t}\left(h_0\right)-\Delta^{ab}G^a_{k+r}\left(h_0\right)U_b\left(h_0\right)+$$
$$\left.\frac{1}{2}\Delta^{ac}\Delta^{bd}\gamma^{ab}_{k+r}\left(h_0\right)U_c\left(h_0\right)U_d\left(h_0\right)+\Delta^{ab}\gamma^a_{k+rk+s}\left(h_0\right)U_{k+s}\left(h_0\right)U_b\left(h_0\right)\right),$$
$$S^3_{\gamma k+r}\left(\widehat{\lambda},\widehat{h}\right)=\frac{1}{nb^d}G^a_{k+t}\left(h_0\right)\left(\Delta^{ab}\Delta^{bc}G^c_{k+r}\left(h_0\right)U_{k+r}\left(h_0\right)-\frac{1}{2}\left([2]\,\gamma^a_{k+rk+s}\left(h_0\right)U_{k+r}\left(h_0\right)U_{k+s}\left(h_0\right)+\right.\right.$$
$$\left.\Delta^{bc}\gamma^{ab}_{k+r}\left(h_0\right)U_{k+r}\left(h_0\right)U_c\left(h_0\right)+[2]\,\Delta^{bc}\gamma^a_{k+rb}\left(h_0\right)U_{k+r}\left(h_0\right)U_c\left(h_0\right)\right)\Big)+$$
$$\frac{1}{2}\Delta^{ac}\Delta^{bd}G^{ab}_{k+t}\left(h_0\right)U_c\left(h_0\right)U_d\left(h_0\right)-$$
$$\frac{1}{3!}\Delta^{ad}\Delta^{be}\Delta^{cf}\gamma^{abc}_{r+k}\left(h_0\right)U_d\left(h_0\right)U_e\left(h_0\right)U_f\left(h_0\right)+$$
$$\Delta^{ad}\left(U_{k+rk+s}\left(h_0\right)-\frac{1}{2}\left(-\gamma_3\mu_{k+rk+s}\left(h_0\right)+[2]\,\Delta^{bc}\gamma^b_{k+rk+s}\left(h_0\right)U_c\left(h_0\right)+\right.\right.$$
$$\left.\left.\gamma_3\left(-\Delta^{ef}U_{k+rk+sf}\left(h_0\right)+[2]\,\Delta^{eh}\Delta^{hi}G^i_{k+rk+s}\left(h_0\right)\right)\Delta^{eg}U_g\left(h_0\right)\right)\right)\times$$
$$G^d_{k+s}\left(h_0\right)U_d\left(h_0\right)+\Big\{-V_{k+r}\left(0,h_0\right)\Delta^{ab}G^b_{k+s}\left(h_0\right)-V_{k+rk+t}\left(0,h_0\right)U_{k+sk+t}\left(h_0\right)+$$
$$V_{k+rl+a}\left(0,h_0\right)\left(\Delta^{ab}\Delta^{bc}G^c_{k+s}\left(h_0\right)-\Delta^{ab}U_b\left(h_0\right)U_{k+rk+s}\left(h_0\right)\right)+$$
$$\frac{1}{2}V_{k+rk}\left(0,h_0\right)\left[-\gamma_3\mu_{k+rk+sk+t}\left(h_0\right)U_{k+t}\left(h_0\right)+\Delta^{ab}\left([2]\,\gamma^a_{k+rk+s}\left(h_0\right)-\gamma_3\mu_{k+rk+ra}\left(h_0\right)\right)U_{k+r}\left(h_0\right)U_b\left(h_0\right)\right]$$
$$\frac{1}{2}V_{k+rl+a}\left(0,h_0\right)\left[-\gamma_3\mu_{ak+sk+t}\left(h_0\right)U_{k+t}\left(h_0\right)+[2]\left(\gamma^b_{ak+s}\left(h_0\right)+\gamma^b_{k+sa}\left(h_0\right)\right)U_b\left(h_0\right)\right]+$$
$$\left([2]\,\Delta^{ab}\gamma^b_j\left(h_0\right)U_{k+t}\left(h_0\right)+\Delta^{ab}\Delta^{cd}\gamma^{bc}_{k+t}\left(h_0\right)U_d\left(h_0\right)\right)\Delta^{ae}G^e_{k+s}\left(h_0\right)+$$
$$[2]\,\gamma^b_{ak+s}\left(h_0\right)\Delta^{bc}U_c\left(h_0\right)\Big)+\left(-\gamma_3\mu_{k+rk+sk+t}\left(h_0\right)U_{k+t}\left(h_0\right)+[2]\,\gamma^a_{k+rk+s}\left(h_0\right)\Delta^{ab}U_b\left(h_0\right)\right)\times$$
$$\left(-\gamma_3\mu_{k+rk+sk+t}\left(h_0\right)U_{k+t}\left(h_0\right)+[2]\,\Delta^{ab}\gamma^a_{k+s}\left(h_0\right)U_b\left(h_0\right)\right)U_{k+sk+u}\left(h_0\right)U_{k+u}\left(h_0\right)-$$
$$\frac{1}{2}\left(-\gamma_3U_{k+rk+sk+t}\left(h_0\right)U_{k+t}\left(h_0\right)+[3]\,\Delta^{ab}G^a_{k+rk+s}\left(h_0\right)U_b\left(h_0\right)-\right.$$
$$\left.\Delta^{ab}\mu_{k+rk+sa}\left(h_0\right)U_b\left(h_0\right)\right)+\frac{1}{6}\left[-\gamma_4U_{k+rk+sk+tk+u}\left(h_0\right)U_{k+t}\left(h_0\right)U_{k+u}\left(h_0\right)-\right.$$
$$[3]\,\gamma_3\Delta^{ab}\gamma^a_{k+rk+sk+t}\left(h_0\right)U_{k+t}\left(h_0\right)U_b\left(h_0\right)-\gamma_4\Delta^{ab}\Delta^{bc}U_{k+rak+sk+t}\left(h_0\right)U_{k+t}\left(h_0\right)U_c\left(h_0\right)+$$
$$[3]\,\gamma_3\Delta^{ab}\Delta^{bc}\Delta^{cd}\gamma^a_{k+rk+sk+t}\left(h_0\right)U_{k+t}\left(h_0\right)U_d\left(h_0\right)-$$
$$[3]\,\gamma_3\Delta^{ab}\Delta^{cd}\Delta^{de}\gamma^a_{k+rck+s}\left(h_0\right)U_b\left(h_0\right)U_e\left(h_0\right)-[2]\,\Delta^{ab}\Delta^{bc}\Delta^{de}\left(\gamma^{a,d}_{k+rk+s}+\gamma^{ad}_{k+rjk+s}\right)U_c\left(h_0\right)U_e\left(h_0\right)-$$
$$[3]\,\gamma_3\Delta^{ab}\gamma^a_{k+rk+sk+t}\left(h_0\right)U_{k+t}\left(h_0\right)U_b\left(h_0\right)-\Delta^{ac}\Delta^{bd}\gamma^{abk+s}_{k+r}\left(h_0\right)U_c\left(h_0\right)U_d\left(h_0\right)-$$
$$\left.[3]\,\Delta^{ab}\Delta^{ae}\Delta^{cd}\gamma^b_{k+rck+s}\left(h_0\right)U_d\left(h_0\right)U_e\left(h_0\right)\right]\Big\}\,\gamma^e_{k+s}\left(h_0\right)U_e\left(h_0\right)+$$
$$\frac{[2]}{2}\Delta^{ab}G^a_{k+rk+s}\left(h_0\right)V_{k+s}\left(0,h_0\right)U_b\left(h_0\right)-\frac{[2]}{2}\gamma^a_{k+rk}\left(h_0\right)V_k\left(0,h_0\right)U_{l+a}\left(h_0\right)-$$
$$\frac{[2]}{2}\left(\gamma^{ab}_{k+rk}\left(h_0\right)+\gamma^{a,b}_{k+rk}\left(h_0\right)\right)V_k\left(0,h_0\right)V_{l+a}\left(0,h_0\right)V_{l+b}\left(0,h_0\right)+\frac{[2]}{2}U_{k+rk}\left(h_0\right)U_k\left(h_0\right)-$$

$\frac{\gamma_3}{3!}[3]\,\Delta^{ab}\gamma^a_{k+rkl}(h_0)\,V_k(0,h_0)\,V_l(0,h_0)\,U_b(h_0)+\frac{\gamma_3}{3!}\mu_{k+rkl}(h_0)\,V_k(0,h_0)\,(U_{k+rk+s}(h_0)\,U_{k+s}(h_0)-$
$\frac{\gamma_3}{2}\left(\mu_{k+rk+sk+t}(h_0)\,U_{k+s}(h_0)\,U_{k+t}(h_0)-\Delta^{ab}\mu_{k+rk+sb}(h_0)\,U_{k+s}(h_0)\,\Delta^{af}U_f(h_0)\right)\times$
$\left(\Delta^{ac}\Delta^{bd}\gamma^d_{jc}(h_0)+[2]\,\Delta^{ac}\gamma^{bc}_j(h_0)\right)\Delta^{be}U_e(h_0)+$
$\frac{\gamma_4}{4!}\mu_{k+rklm}(h_0)\,V_k(0,h_0)\,V_l(0,h_0)\,V_m(0,h_0)\,.$

Lengthy calculations show that

$$E\left[S_{\gamma k+r}\left(\widehat{\lambda},\widehat{h}\right)\right]=-\frac{(3+\gamma_3)}{(nb^d)^{1/2}\,6}\mu_{k+rk+sk+sj}(h_0)-\Delta^{ab}\gamma^a_{k+rb}(h_0)+\frac{1}{2}\Delta^{ac}\Delta^{bc}\gamma^{ab}_{k+r}(h_0)+O_p\left(\frac{1}{(nb^d)^{3/2}}\right),$$

$$\begin{aligned}
&E\left[S_{\gamma k+r}\left(\widehat{\lambda},\widehat{h}\right)S_{\gamma k+s}\left(\widehat{\lambda},\widehat{h}\right)\right]=\delta_{k+rk+s}+\frac{1}{nb^d}\left\{\left(\frac{\gamma_4}{4}+2\right)\mu_{k+rk+sk+tk+t}(h_0)+\right.\\
&\left(\frac{\gamma_3+3}{6}\right)^2\mu_{k+rk+tk+t}\mu_{k+sk+uk+u}(h_0)+\left(\frac{5\gamma_3}{6}+\frac{59}{36}\right)\mu_{k+rk+sk+t}(h_0)\,\mu_{k+tk+uk+u}(h_0)+\\
&\left(\frac{25}{9}+\frac{5\gamma_3}{3}+\frac{\gamma_3^2}{18}\right)\mu_{k+rk+tk+u}(h_0)\,\mu_{k+sk+tk+u}(h_0)+[2]\,\Delta^{ab}\gamma^a_{k+rk+sb}(h_0)+\\
&[2]\,\frac{\gamma_3}{2}\Delta^{ab}\mu_{k+rk+sb}(h_0)\,\gamma^a_{k+uk+u}(h_0)-\frac{1}{2}\Delta^{ab}\mu_{k+sk+sb}(h_0)\,\gamma^a_{k+rk+t}(h_0)-\\
&\frac{1}{2}\Delta^{ab}\Delta^{bc}\Delta^{df}\Delta^{ef}\gamma^{ad}_{k+r}(h_0)\,\gamma^{be}_{k+s}(h_0)-\Delta^{ac}\Delta^{bc}\gamma^{ab}_{k+rk+s}(h_0)+\\
&\frac{1}{6}\Delta^{ab}\Delta^{bc}\mu_{k+rk+sk+t}\gamma^{ab}_{k+t}(h_0)+\left(\frac{2\gamma_3+1}{3}\right)\Delta^{ab}\mu_{k+sk+sk+t}(h_0)\,\gamma^a_{k+tb}(h_0)+\\
&[2]\,\Delta^{ac}\Delta^{bc}\Delta^{de}\gamma^{ab}_{k+r}(h_0)\,\gamma^d_{k+se}(h_0)+\Delta^{ac}\Delta^{bc}\gamma^a_{k+rk+t}(h_0)\,\gamma^b_{k+sk+t}(h_0)-\\
&\frac{[2]}{2}\Delta^{ab}\Delta^{cd}\gamma^{ab}_{k+r}(h_0)\,\mu_{k+scd}(h_0)+\Delta^{ab}\Delta^{cd}\gamma^a_{k+rb}(h_0)\,\gamma^c_{k+rd}(h_0)+\\
&\frac{1}{4}\Delta^{ac}\Delta^{bc}\Delta^{df}\Delta^{ef}\gamma^{ab}_{k+r}(h_0)\,\gamma^{cd}_{k+s}(h_0)+[2]\left(\frac{\gamma_3+3}{6}\right)\Delta^{ab}\mu_{k+rk+tk+t}(h_0)\,\gamma^a_{k+sb}(h_0)-\\
&\left.[2]\left(\frac{\gamma_3+3}{12}\right)\Delta^{ac}\Delta^{bc}\mu_{k+rk+tk+t}(h_0)\,\gamma^{ab}_{k+s}(h_0)-\frac{[2]}{2}\Delta^{ab}\Delta^{ce}\Delta^{de}\gamma^a_{k+rb}(h_0)\,\gamma^{cd}_{k+s}(h_0)\right\}+O_p\left(\frac{1}{(nb^d)^2}\right),
\end{aligned}$$

$$\begin{aligned}
&E\left[S_{\gamma k+r}\left(\widehat{\lambda},\widehat{h}\right)S_{\gamma k+s}\left(\widehat{\lambda},\widehat{h}\right)S_{\gamma k+t}\left(\widehat{\lambda},\widehat{h}\right)\right]=\frac{\mu_{k+rk+sk+t}(h_0)}{(nb^d)^{1/2}}+\\
&\frac{[3]}{(nb^d)^{1/2}}\left(\frac{(3+\gamma_3)}{6}\mu_{k+rk+uk+u}(h_0)\,\delta_{k+sk+t}-\frac{(3+\gamma_3)}{3}\mu_{k+rk+sk+t}(h_0)-\right.\\
&\left.\Delta^{ab}\gamma^a_{k+rb}(h_0)\,\delta_{k+sk+t}+\frac{1}{2}\Delta^{ac}\Delta^{bc}\gamma^{ab}_{k+r}(h_0)\,\delta_{k+sk+t}\right)+O_p\left(\frac{1}{(nb^d)^{3/2}}\right).
\end{aligned}$$

Let

$$\begin{aligned}
\tau_1&=\mu_{k+rk+sk+tk+u}(h_0)\,,\quad \tau_2=[3]\,\delta_{k+rk+s}\delta_{k+tk+u},\\
\tau_3&=[4]\,\mu_{k+rk+sk+t}(h_0)\,\mu_{k+uk+vk+v}(h_0)\,,\quad \tau_4=[3]\,\mu_{k+rk+sk+v}(h_0)\,\mu_{k+tk+uk+v}(h_0)\,;
\end{aligned}$$

note that

$$E\left[S_{\gamma k+r}\left(\widehat{\lambda},\widehat{h}\right)S_{\gamma k+s}\left(\widehat{\lambda},\widehat{h}\right)S_{\gamma k+t}\left(\widehat{\lambda},\widehat{h}\right)S_{\gamma k+u}\left(\widehat{\lambda},\widehat{h}\right)\right]=$$
$$E\left[S^1_{\gamma k+r}\left(\widehat{\lambda},\widehat{h}\right)S^1_{\gamma k+s}\left(\widehat{\lambda},\widehat{h}\right)S^1_{\gamma k+t}\left(\widehat{\lambda},\widehat{h}\right)S^1_{\gamma k+u}\left(\widehat{\lambda},\widehat{h}\right)\right]+$$
$$[4]\,E\left[S^2_{\gamma k+r}\left(\widehat{\lambda},\widehat{h}\right)S^1_{\gamma k+s}\left(\widehat{\lambda},\widehat{h}\right)S^1_{\gamma k+t}\left(\widehat{\lambda},\widehat{h}\right)S^1_{\gamma k+u}\left(\widehat{\lambda},\widehat{h}\right)\right]+$$
$$[6]\,E\left[S^2_{\gamma k+r}\left(\widehat{\lambda},\widehat{h}\right)S^2_{\gamma k+s}\left(\widehat{\lambda},\widehat{h}\right)S^1_{\gamma k+t}\left(\widehat{\lambda},\widehat{h}\right)S^1_{\gamma k+u}\left(\widehat{\lambda},\widehat{h}\right)\right]+$$
$$[4]\,E\left[S^3_{\gamma k+r}\left(\widehat{\lambda},\widehat{h}\right)S^1_{\gamma k+s}\left(\widehat{\lambda},\widehat{h}\right)S^1_{\gamma k+t}\left(\widehat{\lambda},\widehat{h}\right)S^1_{\gamma k+u}\left(\widehat{\lambda},\widehat{h}\right)\right]$$

and that

$$E\left[S^1_{\gamma k+r}\left(\widehat{\lambda},\widehat{h}\right)S^1_{\gamma k+s}\left(\widehat{\lambda},\widehat{h}\right)S^1_{\gamma k+t}\left(\widehat{\lambda},\widehat{h}\right)S^1_{\gamma k+u}\left(\widehat{\lambda},\widehat{h}\right)\right]=\frac{\tau_1}{nb^d}+\tau_2$$
$$[4]\,E\left[S^2_{\gamma k+r}\left(\widehat{\lambda},\widehat{h}\right)S^1_{\gamma k+s}\left(\widehat{\lambda},\widehat{h}\right)S^1_{\gamma k+t}\left(\widehat{\lambda},\widehat{h}\right)S^1_{\gamma k+u}\left(\widehat{\lambda},\widehat{h}\right)\right]-$$
$$[12]\,E\left[S^2_{\gamma k+r}\left(\widehat{\lambda},\widehat{h}\right)S^2_{\gamma k+s}\left(\widehat{\lambda},\widehat{h}\right)\right]E\left[S^1_{\gamma k+t}\left(\widehat{\lambda},\widehat{h}\right)S^1_{\gamma k+u}\left(\widehat{\lambda},\widehat{h}\right)\right]=$$
$$-6\tau_1-\left(\frac{\gamma_3+3}{6}\right)\tau_3-\frac{2}{3}\left(2\gamma_3-3\right)\tau_4-$$
$$[4]\,\Delta^{ab}\gamma^{ab}_{k+rb}\left(h_0\right)\mu_{k+sk+tk+u}\left(h_0\right)+\frac{[4]}{2}\Delta^{ac}\Delta^{bc}\gamma^{ab}_{k+r}\left(h_0\right)\mu_{k+sk+tk+u}\left(h_0\right),$$
$$[6]\,E\left[S^2_{\gamma k+r}\left(\widehat{\lambda},\widehat{h}\right)S^2_{\gamma k+s}\left(\widehat{\lambda},\widehat{h}\right)S^1_{\gamma k+t}\left(\widehat{\lambda},\widehat{h}\right)S^1_{\gamma k+u}\left(\widehat{\lambda},\widehat{h}\right)\right]-$$
$$[6]\,E\left[S^2_{\gamma k+r}\left(\widehat{\lambda},\widehat{h}\right)S^2_{\gamma k+s}\left(\widehat{\lambda},\widehat{h}\right)\right]E\left[S^1_{\gamma k+t}\left(\widehat{\lambda},\widehat{h}\right)S^1_{\gamma k+u}\left(\widehat{\lambda},\widehat{h}\right)\right]=$$
$$3\tau_1-\tau_2+\left(\gamma_3+3\right)^2\frac{\tau_3}{6}+\left(4\gamma_3^2+24\gamma_3+27\right)\frac{\tau_4}{9}-$$
$$\frac{[6]}{3}\left(\gamma_3+3\right)\left(\Delta^{ac}\Delta^{bc}\mu_{k+rk+tk+u}\left(h_0\right)\gamma^{ab}_{k+s}\left(h_0\right)-2\Delta^{ab}\mu_{k+rk+tk+u}\left(h_0\right)\gamma^{a}_{k+sb}\left(h_0\right)\right),$$
$$[4]\left[S^3_{\gamma k+r}\left(\widehat{\lambda},\widehat{h}\right)S^1_{\gamma k+s}\left(\widehat{\lambda},\widehat{h}\right)S^1_{\gamma k+t}\left(\widehat{\lambda},\widehat{h}\right)S^1_{\gamma k+u}\left(\widehat{\lambda},\widehat{h}\right)\right]-$$
$$[12]\,E\left[S^3_{\gamma k+r}\left(\widehat{\lambda},\widehat{h}\right)S^1_{\gamma k+s}\left(\widehat{\lambda},\widehat{h}\right)\right]E\left[S^1_{\gamma k+t}\left(\widehat{\lambda},\widehat{h}\right)S^1_{\gamma k+u}\left(\widehat{\lambda},\widehat{h}\right)\right]=$$
$$\frac{[4]}{4}\left(\gamma_4+8\right)\tau_1+\left(30\gamma_3+59\right)\frac{\tau_4}{9}+3\,[4]\left(\gamma_3+2\right)\Delta^{ab}\mu_{k+rk+sb}\left(h_0\right)\gamma^{a}_{k+tk+u}\left(h_0\right).$$

Since

$$E\left[S^2_{\gamma k+r}\left(\widehat{\lambda},\widehat{h}\right)\right]\Big\{[4]\,E\left[S^1_{\gamma k+s}\left(\widehat{\lambda},\widehat{h}\right)S^1_{\gamma k+t}\left(\widehat{\lambda},\widehat{h}\right)S^1_{\gamma k+u}\left(\widehat{\lambda},\widehat{h}\right)\right]+$$
$$[12]\,E\left[S^2_{\gamma k+s}\left(\widehat{\lambda},\widehat{h}\right)S^1_{\gamma k+t}\left(\widehat{\lambda},\widehat{h}\right)S^1_{\gamma k+u}\left(\widehat{\lambda},\widehat{h}\right)\right]-$$
$$2\,[6]\,E\left[S^2_{\gamma k+s}\widehat{\lambda},\widehat{h}\right]E\left[S^1_{\gamma k+t}\left(\widehat{\lambda},\widehat{h}\right)S^1_{\gamma k+u}\left(\widehat{\lambda},\widehat{h}\right)\right]\Big\}=-\,[4]\left(\gamma_3+2\right)\mu_{k+rk+sk+t}\left(h_0\right)\times$$
$$\left[\left(\frac{3+\gamma_3}{6}\right)\mu_{k+uk+vk+v}\left(h_0\right)-\Delta^{ab}\gamma^{a}_{k+ub}\left(h_0\right)+\frac{1}{2}\Delta^{ac}\Delta^{bc}\gamma^{ab}_{k+u}\left(h_0\right)\right]$$

using (41) we have that

$$\kappa_{k+r}(h_0) = \frac{1}{(nb^d)^{1/2}}\left[-\frac{(3+\gamma_3)}{6}\mu_{k+rk+sk+s}(h_0) - \Delta^{ab}\gamma^a_{k+rb}(h_0) + \frac{1}{2}\Delta^{ac}\Delta^{bc}\gamma^{ab}_{k+r}(h_0)\right] + O_p\left(\frac{1}{(nb^d)^{3/2}}\right)$$

$$\kappa_{k+r,k+s}(h_0) = \delta_{jk} + \kappa^{(2)}_{j,k}(h_0), \quad \kappa^{(2)}_{j,k}(h_0) = \frac{1}{nb^d}\left\{\left(\frac{\gamma_4}{4}+2\right)\mu_{k+rk+sk+tk+t}(h_0) + \right.$$
$$\left(\frac{5\gamma_3}{6}+\frac{59}{36}\right)\mu_{k+rk+sk+t}(h_0)\,\mu_{k+tk+uk+u}(h_0) + \left(\frac{25}{9}+\frac{5\gamma_3}{3}+\frac{\gamma_3^2}{18}\right)\mu_{k+rk+tk+u}(h_0)\,\mu_{k+sk+tk+u}(h_0) +$$
$$[2]\,\Delta^{ab}\gamma^a_{k+rk+sb}(h_0) + [2]\,\frac{\gamma_3}{2}\Delta^{ab}\mu_{k+rk+sb}(h_0)\,\gamma^a_{k+uk+u}(h_0) -$$
$$\frac{1}{2}\Delta^{ab}\mu_{k+sk+sb}(h_0)\,\gamma^a_{k+rk+t}(h_0) - \frac{1}{2}\Delta^{ab}\Delta^{bc}\Delta^{df}\Delta^{ef}\gamma^{ad}_{k+r}(h_0)\,\gamma^{be}_{k+s}(h_0) - \Delta^{ac}\Delta^{bc}\gamma^{ab}_{k+rk+s}(h_0) +$$
$$\frac{1}{4}\Delta^{ac}\Delta^{bc}\Delta^{df}\Delta^{ef}\gamma^{ab}_{k+r}(h_0)\,\gamma^{cd}_{k+s}(h_0) + [2]\left(\frac{\gamma_3+3}{6}\right)\Delta^{ab}\mu_{k+rk+tk+t}(h_0)\,\gamma^a_{k+sb}(h_0) -$$
$$\left.[2]\left(\frac{\gamma_3+3}{12}\right)\Delta^{ac}\Delta^{bc}\mu_{k+rk+tk+t}(h_0)\,\gamma^{ab}_{k+s}(h_0) - \frac{[2]}{2}\Delta^{ab}\Delta^{ce}\Delta^{de}\gamma^a_{k+rb}(h_0)\,\gamma^{cd}_{k+s}(h_0)\right\}$$
$$+O_p\left(\frac{1}{(nb^d)^2}\right),$$

$$\kappa_{k+r,k+s,k+t}(h_0) = -\frac{\mu_{k+rk+sk+t}(h_0)}{(nb^d)^{1/2}}(2+\gamma_3) + O_p\left(\frac{1}{(nb^d)^{3/2}}\right),$$

$$\kappa_{k+r,k+s,k+t,k+u}(h_0) = \frac{1}{nb^d}\left\{(\gamma_4+6)\,\tau_4 + \frac{(\gamma_3^2+5\gamma_3+6)}{6}\tau_3 + \frac{2}{9}\left(2\gamma_3^2+21\gamma_3+34\right)\tau_4 - \right.$$
$$[4]\,\mu_{k+rk+sk+t}(h_0)\left[\left(\frac{3+\gamma_3}{6}\right)\mu_{k+uk+vk+v}(h_0) - \Delta^{ab}\gamma^a_{k+ub}(h_0) + \Delta^{ac}\Delta^{bc}\gamma^{ab}_{k+u}(h_0)\right] +$$
$$\frac{(\gamma_3+2)}{2}\left[\Delta^{ac}\Delta^{bc}\gamma^a_{k+rk+s}(h_0)\,\gamma^b_{k+tk+u}(h_0) + \Delta^{ab}\gamma^{a,b}_{k+rk+sk+tk+u}(h_0)\right] + \frac{(\gamma_3+2)}{3}\mu_{k+rk+s}(h_0)\,\Delta^{ab}\times$$
$$\left.\left(\gamma^{ab}_{k+tk+u}(h_0) + \gamma^{a,b}_{k+tk+u}(h_0)\right)\right\} + O_p\left(\frac{1}{(nb^d)^2}\right).$$

The same arguments as those used to prove Theorem 1 show that the Edgeworth expansion for $S_\gamma\left(\widehat{\lambda},\widehat{h}\right)$

$$\sup_{c\in[c_0,\infty)}\left|P\left(S_\gamma\left(\widehat{\lambda},\widehat{h}\right)\geq c\right) - G^-_{l-k}(c) - \int_{u^Tu\geq c}p_{2n}(u)\,\phi(u)\right| = O\left(\frac{1}{(nb^d)^2}\right)$$

is valid in the sense of Chandra and Ghosh (1980), where $G^-_{l-k}(c) = 1 - G_{l-k}(c)$,

$$\int_{u^Tu\geq c}p_{2n}(u)\,\phi(u) = \frac{1}{nb^d}\left\{\frac{1}{2}c_{1\gamma}(K)\,\nabla G^-_{l-k}(c) + \frac{1}{24}c_{2\gamma}(K)\,\nabla^2 G^-_{l-k}(c_\alpha) + \frac{1}{72}c_{3\gamma}(\rho,K)\,\nabla^3 G^-_{l-k}(c_\alpha)\right\}$$

and

$$\begin{aligned}
c_{1\gamma}(\rho, K) &= \left(\frac{\gamma_4}{4}+2\right)\mu_{k+rk+rk+tk+t}(h_0)+\left(\frac{5\gamma_3}{6}+\frac{59}{36}\right)\mu_{k+rk+rk+t}(h_0)\,\mu_{k+tk+uk+u}(h_0)+ \qquad (56)\\
&\quad \left(\frac{25}{9}+\frac{5\gamma_3}{3}+\frac{\gamma_3^2}{18}\right)\mu_{k+rk+tk+u}(h_0)\,\mu_{k+rk+tk+u}(h_0)+\\
&\quad \frac{(3+\gamma_3)^2}{36}\mu_{k+rk+sk+s}(h_0)\,\mu_{k+rk+tk+t}(h_0)+[2]\,\Delta^{ab}\gamma^{a}_{k+rk+sb}(h_0)+\\
&\quad [2]\,\frac{\gamma_3}{2}\Delta^{ab}\mu_{k+rk+sb}(h_0)\,\gamma^{a}_{k+uk+u}(h_0)-\frac{1}{2}\Delta^{ab}\mu_{k+rk+sb}(h_0)\,\gamma^{a}_{k+rk+t}(h_0)\\
&\quad -\frac{1}{2}\Delta^{ab}\Delta^{bc}\Delta^{df}\Delta^{ef}\gamma^{ad}_{k+r}(h_0)\,\gamma^{be}_{k+r}(h_0)-\Delta^{ac}\Delta^{bc}\gamma^{ab}_{k+rk+r}(h_0)+\\
&\quad \frac{1}{2}\Delta^{ac}\Delta^{bc}\Delta^{df}\Delta^{ef}\gamma^{ab}_{k+r}(h_0)\,\gamma^{cd}_{k+r}(h_0)+\Delta^{ab}\Delta^{cd}\gamma^{ab}_{k+r}(h_0)\,\gamma^{cd}_{k+r}(h_0)\\
&\quad -\left(\frac{\gamma_3+3}{3}\right)\Delta^{ac}\Delta^{bc}\mu_{k+rk+tk+t}(h_0)\,\gamma^{ab}_{k+r}(h_0)-\\
&\quad [2]\,\Delta^{ab}\Delta^{ce}\Delta^{de}\gamma^{a}_{k+rb}(h_0)\,\gamma^{cd}_{k+s}(h_0),\\
c_{2\gamma}(\rho, K) &= [3]\left[(\gamma_4+6)\,\tau_4+\frac{(\gamma_3^2+5\gamma_3+6)}{6}\tau_3+\frac{2}{9}\left(2\gamma_3^2+21\gamma_3+34\right)\tau_4-\right.\\
&\quad [4]\,(\gamma_3+2)\,\mu_{k+tk+sk+t}(h_0)\left(-\Delta^{ab}\gamma^{a}_{k+ub}(h_0)+\Delta^{ac}\Delta^{bc}\gamma^{ab}_{k+u}(h_0)\right)\\
&\quad \frac{(\gamma_3+2)}{2}\left[\Delta^{ac}\Delta^{bc}\gamma^{a}_{k+rk+s}(h_0)\,\gamma^{b}_{k+tk+u}(h_0)+\Delta^{ab}\gamma^{a,b}_{k+rk+sk+tk+u}(h_0)\right]+\\
&\quad \left.\frac{(\gamma_3+2)}{3}\mu_{k+rk+s}(h_0)\,\Delta^{ab}\left(\gamma^{ab}_{k+tk+u}(h_0)+\gamma^{a,b}_{k+tk+u}(h_0)\right)\right]\delta_{k+rk+s}\delta_{k+tk+u},\\
c_{3\gamma}(\rho, K) &= [15]\,(2+\gamma_3)^2\,\mu_{k+rk+sk+t}(h_0)\,\mu_{k+uk+vk+w}(h_0)\,\delta_{k+rk+s}\delta_{k+tk+u}\delta_{k+vk+w}.
\end{aligned}$$

■

Proof of Corollary 1.4. The proof is similar to that of Corollary 1.3, so is omitted. ■

Proof of Theorem 3. The consistency of $\widehat{\lambda}^*$ and $\widehat{h}^*$ follows by adapting the arguments of Newey and Smith (2004). First notice that by Proposition (7.2) and the envelope assumption A2(iii), we have that

$$\sup_{h\in\mathcal{H}_C}\left\|\frac{1}{nb^d}\sum g_i^{*K}(h)\,g_i^{*K}(h)^T-E^*\left[\frac{g_i^{*K}(h)\,g_i^{*K}(h)^T}{b^d}\right]\right\|=o_{p^*-p}(1), \qquad (57)$$

and that $\max_{1\le i\le n}\gamma_2(\lambda_n^{*T}g_i^{*K}(\widehat{h}^*))=o_{p^*-p}(1)$ and $\gamma_2(\lambda_n^{*T}g_i^{*K}(\widehat{h}^*))\geq -C$ with probability approaching 1 and $\lambda_n^*=(nb^d)^{-\delta}g_i^{*K}(\widehat{h}^*)/\|g_i^{*K}(\widehat{h}^*)\|$. A second order Taylor expansion yields

$$\begin{aligned}
\widehat{\Gamma}_\gamma(\lambda^*,\widehat{h}^*) &\geq \gamma_0+\frac{\lambda_n^{*T}}{nb^d}\sum g_i^{*K}(\widehat{h}^*)+\frac{\lambda_n^{*T}}{2nb^b}\sum\gamma_2(\overline{\lambda}_n^{*T}g_i^{*K}(\widehat{h}^*))g_i^{*K}(\widehat{h}^*)g_i^{*K}(\widehat{h}^*)^T\lambda_n^*\\
&\geq \gamma_0+\frac{1}{(nb^d)^\delta}\left\|\frac{\sum g_i^{*K}(\widehat{h})}{nb^d}\right\|-\frac{C}{(nb^d)^{2\delta}},
\end{aligned}$$

where $\overline{\lambda}_n^*$ is on the line joining 0 and λ_n^*, and the second inequality follows because by (57), Propositions 7.1 and 7.2

$$\|\frac{E^*(g_i^{*K}(\widehat{h}^*)g_i^{*K}(\widehat{h}^*)^T)}{b^d} \leq \sup_{h\in\mathcal{H}_C} \|\frac{1}{nb^d}\sum g_i^K(h)g_i^K(h)^T - \Omega(h)\int K(u)^2 du f(z)\| = o_p(1)$$

so that $\sigma_{\min}(E^*(g_i^{*K}(\widehat{h}^*)g_i^{*K}(\widehat{h}^*)^T)/b^d > C > 0$. Since

$$\|\frac{1}{nb^d}\sum g_i^{*K}(\widehat{h})\| \leq \|\frac{1}{nb^d}\sum g_i^{*K}(\widehat{h}) - E^*(\frac{g_i^{*K}(\widehat{h})}{b^d})\| + \|E^*(\frac{g_i^{*K}(\widehat{h})}{b^d})\| = o_{p*-p}(1) \tag{58}$$

the saddlepoint property of $\widehat{h}^*$ implies that

$$\gamma_0 + \frac{1}{(nb^d)^{\delta}}\|\frac{\sum g_i^{*K}(\widehat{h}^*)}{nb}\| - \frac{C}{(nb^d)^{(2\delta)}} \leq \widehat{\Gamma}_\gamma(\lambda_n^*, \widehat{h}^*) \leq \sup_{\lambda\in\Lambda_n(\widehat{h})} \widehat{\Gamma}_\gamma(\lambda, \widehat{h}) \leq \|\|\frac{1}{nb^d}\sum g_i^{*K}(\widehat{h})\|^2 = o_{p^*-p}(1), \tag{59}$$

hence as in Lemma 3 of Newey and Smith (2004) we have $\|\sum g_i^{*K}(\widehat{h}^*)\| = o_{p^*-p}(1)$. By Proposition (7.2) it follows that $\|E^*(g_i^{*K}(\widehat{h}^*))\| = o_p(1)$ so $\|E^*(g_i^*(\widehat{h}^*))|Z_i^* = z)f(z)\| = o_p(1)$. By **A1**(i), for $\eta > 0$, there exists an $\epsilon > 0$ such that for $\|h^* - \widehat{h}\| > \epsilon$ $\|E^*(g_i^*(h^*))|Z_i^* = z)f(z)\| > \eta$ with probability approaching 1, which impliies $\|\widehat{h}^* - \widehat{h}\| = o_{p^*-p}(1)$. Let $\widehat{\lambda}^* := \arg\max_{\lambda\in\Lambda_n\widehat{h}^*} \widehat{\Gamma}_n(\lambda, , \widehat{h}^*)$; given the consistency of $\widehat{h}^*$, by a second order Taylor expansion with Lagrange remainder, there exists a $\overline{\lambda}^*$ on the line joining 0 and $\widehat{h}^*$ such that

$$\begin{aligned}
\gamma_0 = \widehat{\Gamma}_\gamma(0, \widehat{h}^*) &\leq \gamma_0 - \widehat{\lambda}^{*T}\frac{\sum g_i^{*K}(\widehat{h}^*)}{nb^d} + \widehat{\lambda}^{*T}\frac{\sum \gamma_2(\overline{\lambda}^{*T} g_i^{*K}(\widehat{h}^*))g_i^{*K}(\widehat{h}^*)g_i^{*K}(\widehat{h}^*)^T\widehat{\lambda}^*}{2nb^d} \\
&\leq \gamma_0 + \frac{\|\widehat{\lambda}^*\|}{nb^d}\sum g_i^{*K}(\widehat{h}^*) - \widehat{\lambda}^{*T}\frac{\sum g_i^{*K}(\widehat{h}^*)g_i^{*K}(\widehat{h}^*)^T\widehat{\lambda}^*}{4nb^d} \\
&\leq \gamma_0 + \|\widehat{\lambda}^*\|\|\frac{\sum g_i^{*K}(\widehat{h}^*)}{nb^d}\| - C\|\widehat{\lambda}^*\|^2,
\end{aligned}$$

which implies $\|\widehat{\lambda}^*\| = o_{p^*-p}(1)$. Given the consistency of $\widehat{\lambda}^*$ and $\widehat{h}^*$, a mean value expansion and the above arguments can be used to show that

$$\left(nb^d\right)^{1/2}\begin{bmatrix}\widehat{\lambda}^* \\ \widehat{h}^* - \widehat{h}\end{bmatrix} = -\begin{bmatrix}\Omega(z)f(z)\int K^2(u)\,du & G(z)f(z) \\ G(z)f(z)^T & 0\end{bmatrix}^{-1}\left(nb^d\right)^{1/2}\begin{bmatrix}\widehat{g}^{*K}\left(\widehat{h}\right) \\ 0\end{bmatrix} + o_{p^*-p}(1). \tag{60}$$

Note that

$$\begin{aligned}
&\left(nb^d\right)^{1/2}\left(\widehat{g}^{*K}\left(\widehat{h}\right) - \widehat{g}^K(h_0)\right) = \left(nb^d\right)^{1/2}\left(\widehat{g}^{*K}(h_0) - E^*\left(\frac{g_i{}^{*K}(h_0)}{b^d}\right)\right) - \qquad (61)\\
&\left(nb^d\right)^{1/2}\left(\widehat{g}^K(h_0) - \widehat{g}^{*K}\left(\widehat{h}\right)\right) + \left(nb^d\right)^{1/2}\left(E^*\left(\frac{g_i{}^{*K}(h_0)}{b^d}\right) - \widehat{g}^K\left(\widehat{h}\right)\right) = \\
&\left(nb^d\right)^{1/2}\left(\widehat{g}^{*K}(h_0) - E^*\left(\frac{g_i{}^{*K}(h_0)}{b^d}\right)\right) + o_{p^*-p}(1),
\end{aligned}$$

where the second equality follows by a mean value expansion and EBLLN of Proposition 7.2, that is

$$
\begin{aligned}
&\left(nb^d\right)^{1/2}\left(\widehat{g}^K\left(h_0\right)-\widehat{g}^{*K}\left(\widehat{h}\right)\right)-\left(nb^d\right)^{1/2}\left(E^*\left(\frac{g_i{}^{*K}\left(h_0\right)}{b^d}\right)-\widehat{g}^K\left(\widehat{h}\right)\right)= \\
&\frac{1}{nb^d}\left(\sum\frac{\partial g_i^{*K}\left(\overline{h}\right)}{\partial h^T}-\sum\frac{\partial g_i^K\left(\widetilde{h}\right)}{\partial h^T}\right)\left(1+o_p\left(1\right)\right)\left(nb^d\right)^{1/2}\left(\widehat{h}-h_0\right)= \\
&o_{p^*-p}\left(1\right)\left(1+o_p\left(1\right)\right)O_p\left(1\right),
\end{aligned}
$$

where $\overline{h}$ and $\widetilde{h}$ are possibly different mean values. To establish the asymptotic normality of $\left(nb^d\right)^{1/2}\left(\widehat{g}^{*K}\left(h_0\right)-E^*\left(g_i{}^{*K}\left(h_0\right)/b^d\right)\right)$, it suffices to consider the univariate case and verify the Lyapunov condition $\lim_{n\to\infty}\sum E^*\left|x_{ni}^*\left(h_0\right)\right|^{2+\delta}\xrightarrow{p}0$, for

$$
\begin{aligned}
x_{ni}^*\left(h_0\right) &= \frac{1}{s_n^*\left(h_0\right)}\left(g_i^{*K}\left(h_0\right)-E^*\left[\frac{g_i^{*K}\left(h_0\right)}{b^d}\right]\right), \\
s_n^{*2}\left(h_0\right) &= \sum\left(g_i^K\left(h_0\right)-E^*\left[\frac{g_i^{*K}\left(h_0\right)}{b^d}\right]\right)^2.
\end{aligned}
$$

By LLN $s_n^{*2}\left(h_0\right)\xrightarrow{p}nb^dVar\left(g_i^K\left(h_0\right)\right)$ and

$$
E^*\frac{\left|g_i^{*K}\left(h_0\right)-E^*\left(g_i{}^{*K}\left(h_0\right)/b^d\right)\right|^{2+\delta}}{b^d}\xrightarrow{p}E\frac{\left|g_i^K\left(h_0\right)-E\left(g_i{}^K\left(h_0\right)\right)\right|^{2+\delta}}{b^d},
$$

so that by the c^r inequality

$$
\begin{aligned}
&\lim_{n\to\infty}\sum E^*\left|x_{ni}^*\left(h_0\right)\right|^{2+\delta}\leq\lim_{n\to\infty}\left(nb^bVar\left(g^K\left(h_0\right)\right)\right)^{-(1+\delta/2)}2^{1+\delta}nE\left|g^K\left(h_0\right)\right|^{2+\delta}+o_p\left(1\right)= \\
&\lim_{n\to\infty}Var\left(g^K\left(h_0\right)\right)^{-1+\delta/2}\left(nb^d\right)^{-\delta/2}2^{1+\delta}E\frac{\left|g_i^K\left(h_0\right)\right|^{2+\delta}}{b^d}=o_p\left(1\right).
\end{aligned}
$$

Thus by the Cramer-Wold device $\left(nb^d\right)^{1/2}\left(\widehat{g}^{*K}\left(h_0\right)-E^*\left(g_i{}^{*K}\left(h_0\right)/b^d\right)\right)\xrightarrow{d^*}N\left(0,\Omega\left(z\right)\right)$ with probability approaching 1, hence using (60),

$$
\sup_{x\in\mathbb{R}^{l+k}}\left|P^*\left(\left(nb^d\right)^{1/2}\left(\widehat{\eta}^*-\widehat{\eta}\right)\leq x\right)-P\left(N\left(0,diag\left[P\left(z\right),\Sigma\left(z\right)\right]\right)\leq x\right)\right|=o_p\left(1\right),
$$

where $\widehat{\eta}^*-\widehat{\eta}=\left[\widehat{\lambda}^{*T},\left(\widehat{h}^*-\widehat{h}\right)^T\right]^T$; consequently

$$
\sup_{x\in\mathbb{R}^{l+k}}\left|P^*\left(\left(nb^d\right)^{1/2}\left(\widehat{\eta}^*-\widehat{\eta}\right)\leq x\right)-P\left(\left(nb^d\right)^{1/2}\left(\widehat{\eta}-\eta_0\right)\leq x\right)\right|=o_p\left(1\right).
$$

Next, following the proof of Theorem 2, we can obtain the EB analog of the stochastic expansion of $\left(nb^d\right)^{1/2}(\widehat{\eta}-\eta_0)_\alpha$ given in (54), that is

$$\left(nb^d\right)^{1/2}(\widehat{\eta}^*-\widehat{\eta})_\alpha=-V_\alpha\left(0,\widehat{h}\right)+\frac{1}{\left(nb^d\right)^{1/2}}\left(V^*_{\alpha\beta}\left(0,\widehat{h}\right)V^*_\beta\left(0,\widehat{h}\right)-\right.$$
$$\left.\frac{v^*_{\alpha\beta\gamma}\left(\widehat{h}\right)}{2}V^*_\beta\left(0,\widehat{h}\right)V^*_\gamma\left(0,\widehat{h}\right)\right)+$$
$$\frac{1}{nb^d}\left(-V^*_{\alpha\beta}\left(0,\widehat{h}\right)V^*_{\beta\gamma}\left(0,\widehat{h}\right)V^*_\gamma\left(0,\widehat{h}\right)+\frac{v^*_{\beta\gamma\delta}\left(\widehat{h}\right)}{2}V^*_{\alpha\beta}\left(0,\widehat{h}\right)V^*_\gamma\left(0,\widehat{h}\right)V^*_\delta\left(0,\widehat{h}\right)+\right.$$
$$v^*_{\alpha\beta\gamma}\left(\widehat{h}\right)V^*_{\beta\delta}\left(0,\widehat{h}\right)V^*_\gamma\left(0,\widehat{h}\right)V^*_\delta\left(0,\widehat{h}\right)-$$
$$\frac{v^*_{\alpha\beta\gamma}\left(\widehat{h}\right)v^*_{\beta\delta\varepsilon}\left(\widehat{h}\right)}{2}V^*_\gamma\left(0,\widehat{h}\right)V^*_\delta\left(0,\widehat{h}\right)V^*_\varepsilon\left(0,\widehat{h}\right)-\frac{1}{2}V^*_{\alpha\beta\gamma}\left(0,\widehat{h}\right)V^*_\beta\left(0,\widehat{h}\right)V^*_\gamma\left(0,\widehat{h}\right)+$$
$$\left.\frac{v^*_{\alpha\beta\gamma\delta}\left(\widehat{h}\right)}{6}V^*_\beta\left(0,\widehat{h}\right)V^*_\gamma\left(0,\widehat{h}\right)V^*_\delta\left(0,\widehat{h}\right)\right)+O_{p^*-p}\left(\frac{1}{\left(nb^d\right)^{3/2}}\right),$$

and of $S_\gamma\left(\widehat{\lambda}^*,\widehat{h}^*\right)$ given in (55). Then using the same signed squared root decomposition and the cumulants calculations of Theorem 2 we obtain the empirical Edgeworth expansion of $S^*_\gamma\left(\widehat{\lambda}^*,\widehat{h}^*\right)$, that is

$$\sup_{c\in[c_0,\infty)}\left|P^*\left(S^*_\gamma\left(\widehat{\lambda}^*,\widehat{h}^*\right)\geq c\right)-G^-_{l-k}(c)+\int_{u^Tu\geq c_\alpha}\widehat{p}_{2n}(u)\,\phi(u)\right|=O_p\left(\frac{1}{\left(nb^d\right)^2}\right),\tag{62}$$

where

$$\int_{u^Tu\geq c_\alpha}\widehat{p}_{2n}(u)\,\phi(u)=\frac{1}{nb^d}\left\{\frac{1}{2}\widehat{c}^*_{1\gamma}(K)\nabla G_{l-k}(c)+\frac{1}{24}\widehat{c}^*_{2\gamma}(K)\nabla^2G_{l-k}(c)+\frac{1}{72}\widehat{c}^*_{3\gamma}(K)\nabla^3G_{l-k}(c)\right\}$$

and $\widehat{c}^*_{j\gamma}(K)$ $(j=1,2,3)$ are the bootstrap analog of the $c_{j\gamma}(K)$'s given in (56). The validity of (62) follows by the same arguments as those used in the proofs of Theorems 1 and 2, noting that

$$\begin{aligned}\sup_{\varepsilon<\|t\|\leq n^c}\left|E^*\exp\left(\iota t^TU_{\overline{k}}\left(\widehat{h},K\right)\right)\right| &\leq \sup_{\varepsilon<\|t\|\leq n^c}\left|E\exp\left(\iota t^TU_{\overline{k}}(h_0,K)\right)\right|+\\ &\quad \sup_{\varepsilon<\|t\|\leq n^c}\left|E^*\exp\left(\iota t^TU_{\overline{k}}\left(\widehat{h},K\right)\right)-E\exp\left(\iota t^TU_{\overline{k}}(h_0,K)\right)\right|\\ &\leq \sup_{\varepsilon<\|t\|\leq n^c}\left|E\exp\left(\iota t^TU_{\overline{k}}(h_0,K)\right)\right|+o_p(1)\\ &= 1-C_4b+o_p(1),\end{aligned}$$

by the same arguments as those used by Horowitz (1998), noting that the class of functions $\mathcal{U}^K=\left\{U_{\overline{k}}(h,K),b>0,h\in\mathcal{H}_C\right\}$ is Euclidean as it consists of products of Euclidean classes of functions, hence the class $\left\{\exp\left(\iota t^T U_{\overline{k}}(h,K)\right), b>0, h\in\mathcal{H}_C, \varepsilon<\|t\|\le n^c\right\}$ is also Euclidean, since exponentiation is Lipschitz continuous on compact sets. Similar arguments as those used in the proof of Corollary 1.3 can be used to show that $\left|\widehat{\mu}^*_{k+r_1\ldots.k+r_l}\left(\widehat{h}^*\right)-\widehat{\mu}_{k+r_1\ldots.k+r_l}\left(\widehat{h}\right)\right| = O_{p^*-p}\left(\left(nb^d\right)^{-1/2}\right)$, $\left|\widehat{\gamma}^{*a_1\ldots a_l}_{k+r_1\ldots k+r_l}\left(\widehat{h}^*\right)-\widehat{\gamma}^{a_1\ldots a_l}_{k+r_1\ldots k+r_l}\left(\widehat{h}\right)\right| = O_{p^*-p}\left(\left(nb^d\right)^{-1/2}\right)$ and $\left|\widehat{\mu}_{k+r_1\ldots.k+r_l}\left(\widehat{h}\right)-\mu_{k+r_1\ldots.k+r_l}(h_0)\right| = O_p\left(\left(nb^d\right)^{-1/2}\right)$, $\left|\widehat{\gamma}^{a_1\ldots a_l}_{k+r_1\ldots k+r_l}\left(\widehat{h}\right)-\gamma^{a_1\ldots a_l}_{k+r_1\ldots k+r_l}(h_0)\right| = O_p\left(\left(nb^d\right)^{-1/2}\right)$, hence by CMT we have that $\left|\widehat{c}^*_{j\gamma}(K)-c_{j\gamma}(K)\right| = O_p\left(\left(nb^d\right)^{-1/2}\right)$ and by the the oddness/evenness property of the Hermite polynomials (Barndorff-Nielsen and Hall 1988)

$$\sup_{c\in[c_0,\infty)}\left|P^*\left(S^*_\gamma\left(\widehat{\lambda}^*,\widehat{h}^*\right)\ge c\right)-P\left(S_\gamma\left(\widehat{\lambda},\widehat{h}\right)\ge c\right)\right| = O_p\left(\frac{1}{\left(nb^d\right)^2}\right).$$

Finally, the asymptotic expansion of the upper quantile of $S_\gamma\left(\widehat{\lambda},\widehat{h}\right)$ given in (25) shows that

$$\begin{aligned} d_\gamma &= c_\alpha\left(1+2\sum_{j=1}^{3}\left(\sum_{l=j}^{3}C_j(K)\right)\right)\frac{c_\alpha^{j-1}}{nb^d\Gamma_j}+O\left(\frac{1}{\left(nb^d\right)^2}\right), \\ c^*_\alpha &= c_\alpha\left(1+2\sum_{j=1}^{3}\left(\sum_{l=j}^{3}\widehat{C}_j(K)\right)\right)\frac{c_\alpha^{j-1}}{nb^d\Gamma_j}+O_p\left(\frac{1}{\left(nb^d\right)^2}\right). \end{aligned}$$

The above results show that $\left|\widehat{C}_j(K)-C_j(K)\right| = O_p\left(\left(nb^d\right)^{-1/2}\right)$, hence the Edgeworth expansion for $S_\gamma\left(\widehat{\lambda},\widehat{h}\right)$ and the delta method imply that

$$\begin{aligned} P\left(S_\gamma\left(\widehat{\lambda},\widehat{h}\right)\ge c^*_\alpha\right) &= P\left(S_\gamma\left(\widehat{\lambda},\widehat{h}\right)-\left(c^*_\alpha-d_\gamma\right)\ge d_\gamma\right) \\ &= P\left(S_\gamma\left(\widehat{\lambda},\widehat{h}\right)-O_p\left(\left(nb^d\right)^{-3/2}\right)\ge c_\alpha\left(1+2\sum_{j=1}^{3}\left(\sum_{l=j}^{3}C_j(K)\right)\right)\frac{c_\alpha^{j-1}}{nb^d\Gamma_j}\right) \\ &= 1-G_{l-k}\left(c_\alpha\left(1+2\sum_{j=1}^{3}\left(\sum_{l=j}^{3}C_j(K)\right)\right)\frac{c_\alpha^{j-1}}{nb^d\Gamma_j}\right)+ \\ &\quad g_{l-k}\left(c_\alpha\left(1+2\sum_{j=1}^{3}\left(\sum_{l=j}^{3}C_j(K)\right)\right)\frac{c_\alpha^{j-1}}{nb^d\Gamma_j}\right)+O\left(\frac{1}{\left(nb^d\right)^2}\right), \end{aligned}$$

since the $O_p\left(\left(nb^d\right)^{-3/2}\right)$ term is actually $O\left(\left(nb^d\right)^{-2}\right)$ by the oddness/evenness property of the Hermite polynomials. ■

8 Tables

Table 1 Finite sample coverage and average length of confidence intervals for $h_0(0)$ in the nonparametric binary choice model

$b=\widehat{b}n^{-a}$		$n=100$		$n=400$		$n=1000$	
	Stat	COV^a	LEN^b	COV^a	LEN^b	COV^a	LEN^b
$a=\frac{1}{4}$	S_{EL}	0.897	0.338	0.910	0.325	0.926	0.296
	S_{EL}^m	0.914	0.339	0.921	0.326	0.939	0.299
	S_{ET}	0.900	0.340	0.907	0.336	0.924	0.300
	S_{ET}^m	0.910	0.341	0.917	0.330	0.937	0.302
	S_{CU}	0.901	0.343	0.910	0.332	0.925	0.297
	S_{CU}^m	0.913	0.342	0.915	0.320	0.937	0.303
$a=\frac{1}{3}$	S_{EL}	0.902	0.344	0.917	0.339	0.930	0.293
	S_{EL}^m	0.918	0.345	0.924	0.340	0.942	0.298
	S_{ET}	0.903	0.364	0.910	0.343	0.928	0.296
	S_{ET}^m	0.915	0.366	0.920	0.346	0.940	0.300
	S_{CU}	0.907	0.360	0.912	0.346	0.927	0.295
	S_{CU}^m	0.915	0.362	0.920	0.348	0.939	0.301
$a=\frac{3}{7}$	S_{EL}	0.908	0.382	0.918	0.343	0.934	0.287
	S_{EL}^m	0.920	0.384	0.929	0.345	0.946	0.292
	S_{ET}	0.906	0.383	0.916	0.345	0.933	0.291
	S_{ET}^m	0.918	0.386	0.925	0.348	0.944	0.296
	S_{CU}	0.910	0.386	0.914	0.350	0.932	0.290
	S_{CU}^m	0.913	0.387	0.926	0.353	0.943	0.295

a Coverage, b Average length

Table 2 Finite sample coverage and average length of confidence intervals for $h_0(1)$ in the nonparametric binary choice model

$b = \widehat{b}n^{-a}$		$n = 100$		$n = 400$		$n = 1000$	
	Stat	COV^a	LEN^b	COV^a	LEN^b	COV^a	LEN^b
$a = \frac{1}{4}$	S_{EL}	0.893	0.424	0.908	0.392	0.928	0.294
	S_{EL}^m	0.907	0.425	0.922	0.394	0.941	0.297
	S_{ET}	0.890	0.434	0.904	0.396	0.924	0.298
	S_{ET}^m	0.904	0.435	0.918	0.398	0.940	0.300
	S_{CU}	0.894	0.434	0.903	0.400	0.926	0.295
	S_{CU}^m	0.903	0.435	0.917	0.402	0.944	0.301
$a = \frac{1}{3}$	S_{EL}	0.901	0.423	0.915	0.385	0.932	0.385
	S_{EL}^m	0.920	0.423	0.926	0.387	0.943	0.387
	S_{ET}	0.899	0.425	0.911	0.387	0.930	0.387
	S_{ET}^m	0.915	0.427	0.921	0.389	0.940	0.389
	S_{CU}	0.900	0.425	0.914	0.390	0.931	0.390
	S_{CU}^m	0.916	0.427	0.923	0.393	0.941	0.393
$a = \frac{3}{7}$	S_{EL}	0.906	0.402	0.919	0.376	0.934	0.291
	S_{EL}^m	0.922	0.404	0.929	0.377	0.943	0.295
	S_{ET}	0.903	0.412	0.915	0.383	0.930	0.293
	S_{ET}^m	0.918	0.413	0.926	0.385	0.942	0.298
	S_{CU}	0.905	0.415	0.916	0.388	0.931	0.296
	S_{CU}^m	0.919	0.417	0.925	0.390	0.944	0.298

a Coverage, b Average length

Table 3 Finite sample rejection probabilities for the specification test in the IV smooth coefficients model for $h_0(-0.5)$

$b = \widehat{b}n^{-a}$		$n = 100$	$n = 400$	$n = 1000$
	Stat	RP	RP	RP
$a = \frac{1}{4}$	S_{EL}	0.090	0.082	0.068
	S_{EL}^{*}	0.065	0.054	0.051
	S_{ET}	0.089	0.081	0.071
	S_{ET}^{*}	0.066	0.056	0.052
	S_{CU}	0.091	0.081	0.070
	S_{CU}^{*}	0.067	0.057	0.051
$a = \frac{1}{3}$	S_{EL}	0.090	0.081	0.064
	S_{EL}^{*}	0.068	0.053	0.050
	S_{ET}	0.088	0.080	0.065
	S_{ET}^{*}	0.070	0.054	0.051
	S_{CU}	0.089	0.080	0.065
	S_{CU}^{*}	0.067	0.055	0.052
$a = \frac{3}{7}$	S_{EL}	0.087	0.079	0.062
	S_{EL}^{*}	0.067	0.052	0.050
	S_{ET}	0.086	0.080	0.063
	S_{ET}^{*}	0.068	0.053	0.050
	S_{CU}	0.085	0.080	0.064
	S_{CU}^{*}	0.066	0.053	0.052

Table 4 Finite sample rejection probabilities for the specification test in the IV smooth coefficients model for $h_0(1)$

$b = \widehat{b}n^{-a}$		$n = 100$	$n = 400$	$n = 1000$
	Stat	RP	RP	RP
$a = \frac{1}{4}$	S_{EL}	0.091	0.081	0.065
	S_{EL}^{*}	0.067	0.058	0.051
	S_{ET}	0.088	0.084	0.066
	S_{ET}^{*}	0.068	0.059	0.051
	S_{CU}	0.090	0.086	0.066
	S_{CU}^{*}	0.066	0.057	0.052
$a = \frac{1}{3}$	S_{EL}	0.087	0.079	0.063
	S_{EL}^{*}	0.065	0.056	0.052
	S_{ET}	0.087	0.079	0.064
	S_{ET}^{*}	0.066	0.057	0.052
	S_{CU}	0.085	0.078	0.063
	S_{CU}^{*}	0.067	0.057	0.052
$a = \frac{3}{7}$	S_{EL}	0.085	0.076	0.062
	S_{EL}^{*}	0.067	0.055	0.050
	S_{ET}	0.083	0.078	0.063
	S_{ET}^{*}	0.068	0.056	0.051
	S_{CU}	0.086	0.077	0.062
	S_{CU}^{*}	0.063	0.056	0.051

Table 5 Rejection probabilities for the $r_1(Z_i)$ specification

$IQ \backslash AGE$	30		33		36	
95.25	4.64^{a}	4.59^{b}	4.70^{a}	4.68^{b}	4.80^{a}	4.86^{b}
	0.098^{c}	0.100^{c}	0.095^{c}	0.096^{c}	0.090^{c}	0.088^{c}
104	5.01^{a}	5.5^{b}	6.24^{a}	6.66^{c}	7.16^{a}	7.06^{c}
	0.078^{c}	0.063^{c}	$0.04^{c}5$	0.035^{c}	0.027^{c}	0.029^{c}
113.75	6.39^{a}	6.56^{b}	6.71^{a}	6.8^{b}	7.83^{a}	7.77^{b}
	0.040^{c}	0.032^{c}	0.034^{c}	0.033^{c}	0.019^{c}	0.020^{c}

$a\, S_{EL}^{*}$, $b\, S_{CU}^{*}$, c p-value

Table 6 Rejection probabilities for the $r_2(Z_i)$ specification

iq \age	30		33		36	
95.25	13.96^a	13.37^b	14.25^a	14.39^b	14.54^a	14.25^b
	0.007^c	0.009^c	0.006^c	0.006^c	0.005^c	0.006^c
104	$14,58^a$	14.31^b	14.52^a	14.60^c	14.71^a	14.61^c
	0.006^c	0.006^c	0.005^c5	0.005^c	0.005^c	0.005^c
113.75	14.32^a	14.42^b	14.48^a	14.56^b	14.82^a	14.73^b
	0.006^c	0.006^c	0.005^c	0.005^c	0.005^c	0.005^c

$a\, S^*_{EL}$, $b\, S^*_{CU}$, c p-value

sectionFigures

Figure 1: Estimated $h's$ as a function of Iq and Age

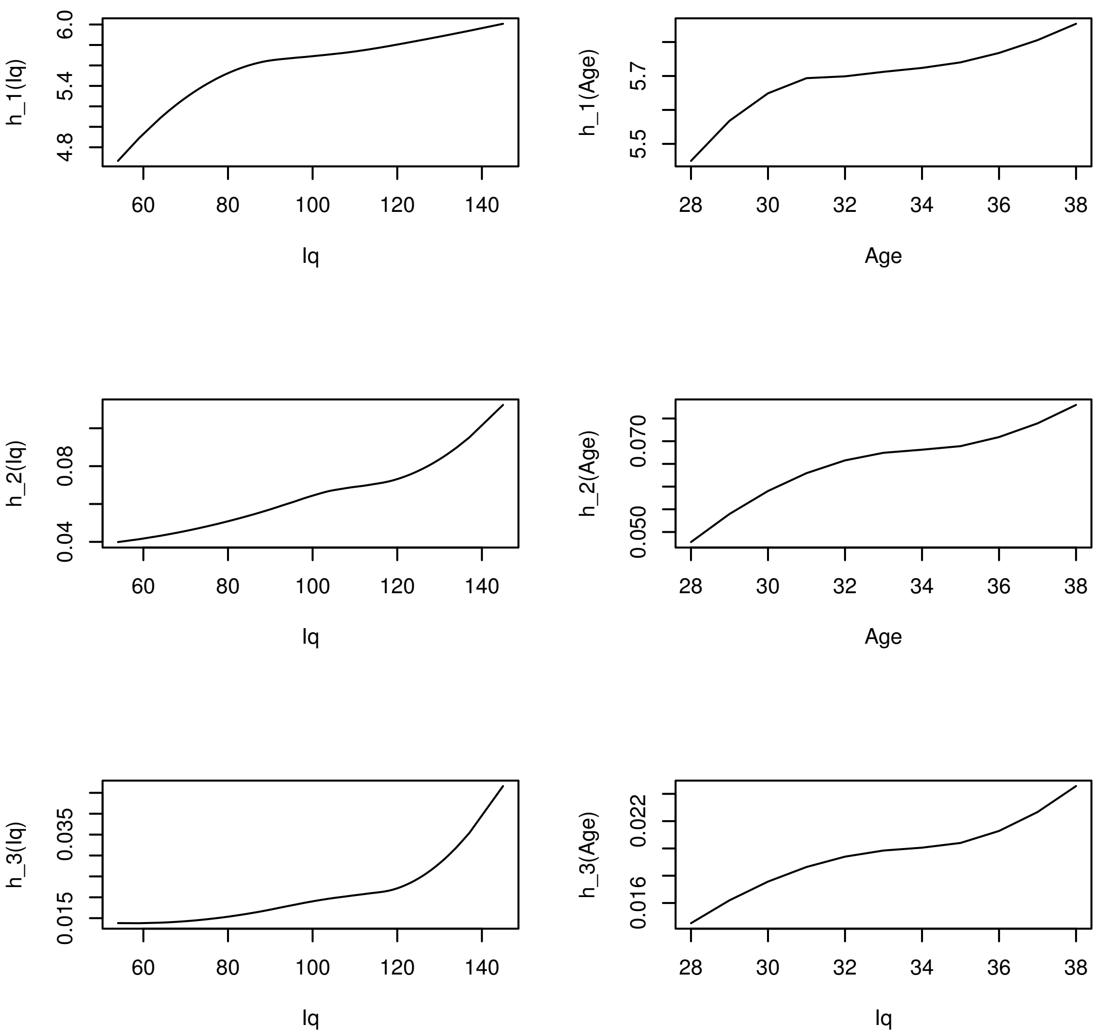

Figure 2: 95% confidence regions for h_1 and h_2. The red and red dotted contours correspond to S_{EL} and S_{EL}^m; the blue and blue dotted contours correspond to S_{CU} and S_{CU}^m.

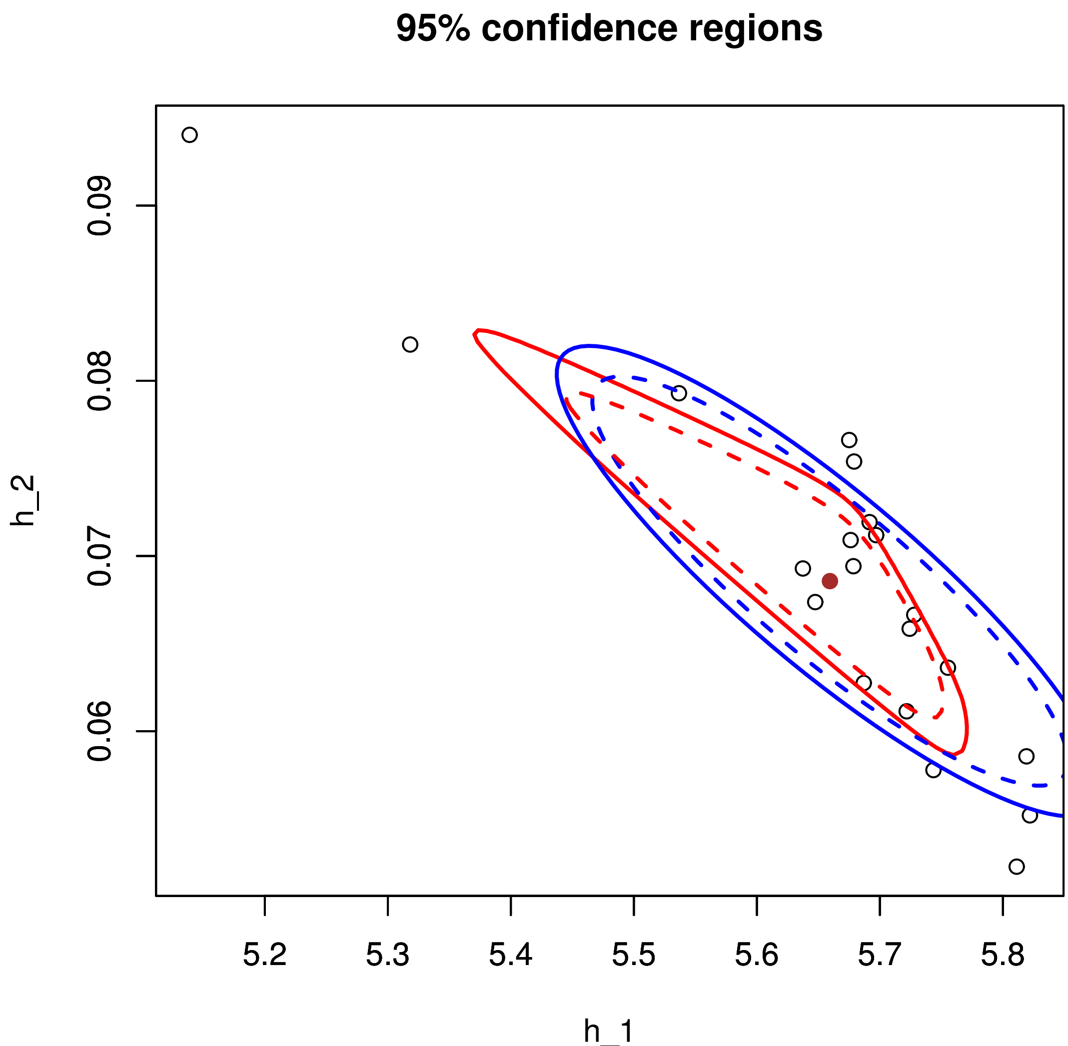

Figure 3: 95% confidence regions for h_1 and h_3. The red and red dotted contours correspond to S_{EL} and S_{EL}^m; the blue and blue dotted contours correspond to S_{CU} and S_{CU}^m.

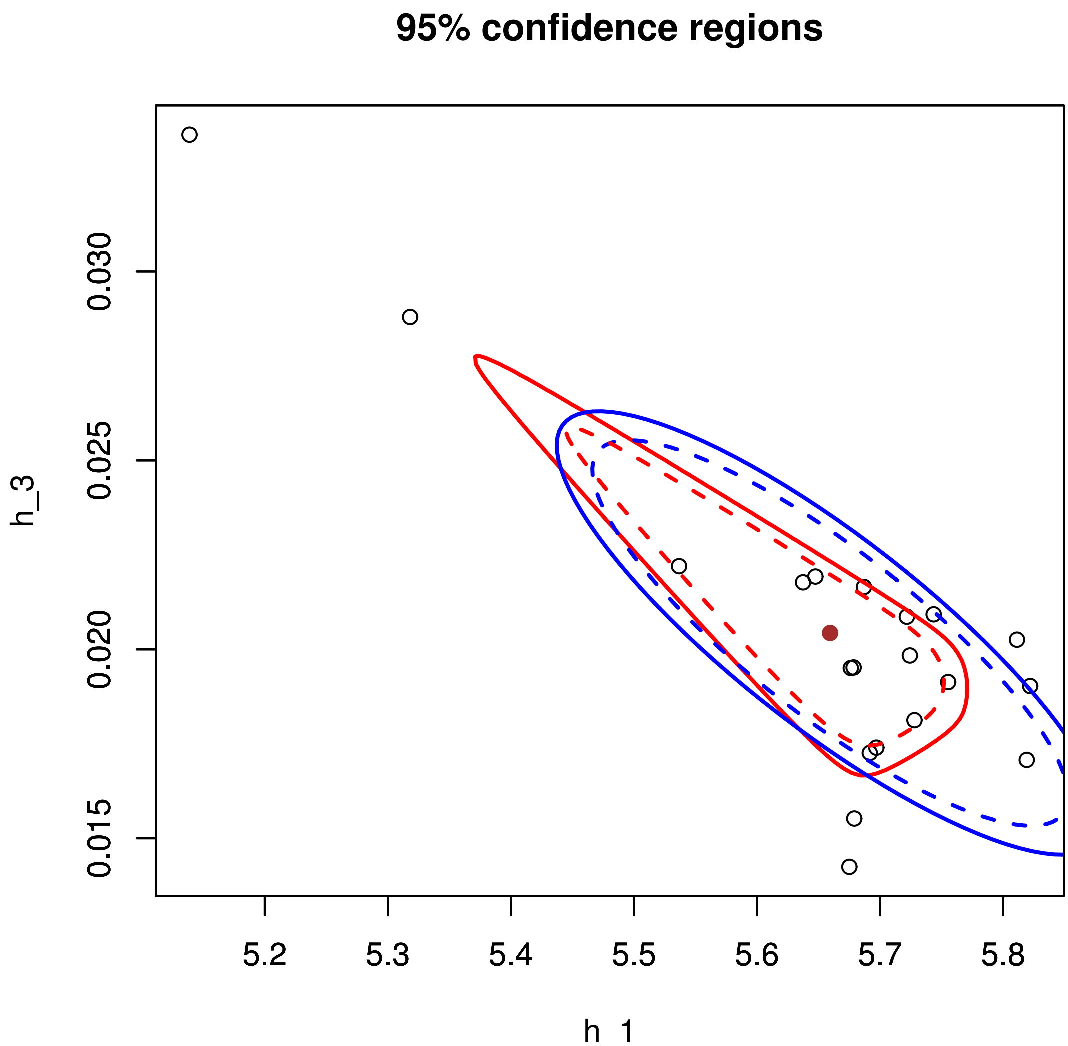

Figure 4: 95% confidence regions for h_2 and h_3. The red and red dotted contours correspond to S_{EL} and S_{EL}^m; the blue and blue dotted contours correspond to S_{CU} and S_{CU}^m.