CIRJE-F-124

Empirical Likelihood-Based Inference in Conditional Moment Restriction Models

Yuichi Kitamura
University of Wisconsin / CIRJE

Gautam Tripathi
University of Wisconsin

Hyungtaik Ahn
Dongguk University

July 2001

EMPIRICAL LIKELIHOOD-BASED INFERENCE IN CONDITIONAL MOMENT RESTRICTION MODELS

YUICHI KITAMURA, GAUTAM TRIPATHI, AND HYUNGTAIK AHN

Abstract. This paper proposes an asymptotically efficient method for estimating models with conditional moment restrictions. Our estimator generalizes the maximum empirical likelihood estimator (MELE) of Qin and Lawless (1994). Using a kernel smoothing method, we efficiently incorporate the information implied by the conditional moment restrictions into our empirical likelihood-based procedure. This yields a one-step estimator which avoids estimating optimal instruments. Our likelihood ratio-type statistic for parametric restrictions does not require the estimation of variance, and achieves asymptotic pivotalness implicitly. The estimation and testing procedures we propose are normalization invariant. Simulation results suggest that our new estimator works remarkably well in finite samples.

1. Introduction

Estimation of econometric models via moment restrictions has been extensively investigated in the literature. Perhaps the most popular technique for estimating models under unconditional moment restrictions is Hansen's (1982) Generalized Method of Moments (GMM). Recently, some alternatives have been suggested by Qin and Lawless (1994), Kitamura and Stutzer (1997), and Imbens, Spady, and Johnson (1998). All these estimators are based on unconditional moment restrictions.

Economic theory, however, often provides conditional moment restrictions. A leading example is the theory of dynamic optimizing agents with time-separable utility. This theory typically predicts implications in terms of martingale differences. GMM and its variants can handle such models, because a conditional moment restriction can be used to derive a set of unconditional moment restrictions using instrumental variables (IV's) that are arbitrary measurable functions of the conditioning variables. However, it is advantageous to efficiently use the information contained in the conditional moment restrictions for better statistical inference. Earlier in the literature, Amemiya (1974) derived the optimal instrumental variables for conditional moment models with homoscedastic errors. Chamberlain (1987) allowed heteroscedasticity of unknown form and showed that the semiparametric efficiency bound for conditional moment restriction models is attained by the optimal IV estimator.

Date: This Version: July 03, 2001.

Keywords: Conditional Moment Restrictions, Empirical Likelihood, Kernel Smoothing.

JEL Classification Number: C14.

We thank Don Andrews, Bruce Hansen, Keisuke Hirano, John Kennan, Ken West and participants at various seminars for many valuable discussions. Shane Sherlund provided excellent research assistance. The first author acknowledges financial support from the Alfred P. Sloan Foundation Research Fellowship, CIRJE, and from the National Science Foundation via grants SBR-9632101 and SES-9905247. The second author thanks the University of Wisconsin Graduate School and the NSF via grant SES-0111917 for research support.

The implementation of the above efficient estimation concepts has been discussed, among others, by Robinson (1987) and Newey (1990, 1993). Robinson and Newey use nonparametric methods to estimate the optimal instruments. Such a procedure yields an asymptotically efficient estimator under quite general and flexible conditions. It can be viewed as a feasible version of Chamberlain's efficient estimator. Although the feasible optimal IV estimator possesses good asymptotic properties in terms of its generality, nonparametric estimation of optimal instruments may require very large samples, thereby affecting the finite sample performance of the feasible estimator.

This paper extends the method of empirical likelihood, introduced by Owen (1988, 1990,1991), to the estimation of conditional moment models. Our approach is similar to the one taken by Robinson and Newey in that it uses a nonparametric method to allow for maximal generality. However, it circumvents the problem of the nonparametric estimation of the optimal instruments. By using a localized version of empirical likelihood we derive a new estimator that achieves the semiparametric efficiency bound automatically; i.e. without estimating the optimal instruments explicitly.

Empirical likelihood is a useful tool of finding estimators, constructing confidence regions, and testing hypotheses. It competes very convincingly with other methods, such as the bootstrap. It is quite general and its applications can be found in a wide range of areas. See, for instance, the review papers by Owen (1995) and Hall and LaScala (1990). In particular, Qin and Lawless (1994) demonstrated that empirical likelihood extends to unconditional moment restriction model with iid samples and that it yields an efficient estimator. Imbens (1993) and Imbens, Spady, and Johnson (1998) discuss similar methods. Kitamura (1997b) showed the weak consistency of the maximum empirical likelihood estimator and further extended the Qin and Lawless approach to weakly dependent data series.

The framework of empirical likelihood is natural and appealing. While it is a nonparametric procedure, it has likelihood-theoretic foundations. Many desirable features of parametric likelihood methods carry over to empirical likelihood. For example, MELE is transformation invariant. A nonparametric analogue of Wilks' theorem also holds: by taking the difference between the constrained and unconstrained empirical loglikelihood and multiplying it by -2, we obtain the empirical likelihood ratio statistic (ELR) that converges to a χ^2 distribution. This point has an important practical implication; namely, ELR-based tests achieve asymptotic pivotalness without explicit studentization. "Implicit pivotalness" may be useful when estimating the variance for studentization is difficult (Chen 1996). This feature is particularly attractive when applying the bootstrap, where pivoting is theoretically important (Beran 1988) but may lead to poor results in practice due to the difficulty of estimating the variance. See, for instance, Fisher, Hall, Jing, and Wood (1996) and Kitamura (1997a).

ELR has other interesting and potentially useful theoretical properties. For example, as shown by DiCiccio, Hall, and Romano (1991), ELR is Bartlett-correctable. Also, Kitamura (2000) recently showed that ELR tests have an optimal power property in terms of a Hoeffding (1963) type asymptotic efficiency criterion. See Hall (1990) for other desirable properties of empirical likelihood.

Our approach builds upon the empirical likelihood method for unconditional moment models discussed above, though our goal is to achieve efficiency gain by exploiting the conditional moment restriction $\mathbb{E}\{g(z,\theta_0)|x\} = 0$, where x denotes the vector of conditioning variables. The estimation strategy follows a two-step method. In the first step we fix θ and obtain the localized version of empirical likelihood at each realization of x under the conditional moment restriction $\mathbb{E}\{g(z,\theta)|x\} = 0$. These are used to construct a global profile likelihood function. In the second step we maximize the profile likelihood from the first step to obtain an estimate of θ_0. More details on this procedure are provided in the next section.

In this paper we show that our approach uses information from the conditional moments effectively, and allows us to obtain an estimate for θ_0 that achieves the semiparametric efficiency bound. This approach emerges naturally as an extension of the classical likelihood paradigm, and is theoretically quite appealing. It seems to be useful in practice as well. For example, as mentioned before, our method has the implicit pivotalness that can be important in a situation where the estimation of the asymptotic variance is difficult.

Before we close this section, let us mention additional papers which may also be related to our investigation. In an independent study Brown and Newey (1998) investigate the same class of conditional moment models as ours. They consider the bootstrap for a conventional optimal instrumental variables estimator such as Newey's. They propose to resample data series according to a distribution estimate obtained from the local empirical likelihood, evaluated at the optimal instrumental variables estimator in question. Their approach seems to be promising, but their goal is quite different from ours in that they considered the bootstrap of conventional estimators, whereas we propose to construct a new efficient estimator. Following Brown and Newey's suggestion, it should be interesting to examine the performance of the bootstrap for our estimator.

LeBlanc and Crowley (1995) propose to use local empirical likelihood to estimate a "conditional functional." However, the class of models they consider is narrower than ours, because they only examine regression functionals. They do not provide formal results on the consistency and asymptotic normality as we do, nor do they note that the local empirical likelihood estimator achieves the semiparametric efficiency bound. Donald, Imbens, and Newey (2001) develop an interesting empirical likelihood-based estimator for conditional moment restriction models. As Donald et al note, their approach is very different from ours in that their estimator achieves the semiparametric efficiency bound by letting the dimension of the unconditional moments grow with sample size. The impact of having high dimensional moment conditions on the finite sample performance of their estimator remains to be seen.

Zhang and Gijbels (2001) independently develop a methodology close to ours. They consider parametric and nonparametric regression models, whereas we consider parametric conditional moment models that nest regression as a special case. Unlike us, they rule out unbounded regressors by assuming that the conditioning variables x are compactly supported such that the density of x is bounded away from zero on its support. Furthermore, their identification relies on the following condition (in our notation): $\inf_{\|\theta-\theta_0\|\geq\delta} \|\mathbb{E}g(z,\theta)\| \neq 0$ for any $\delta > 0$. For this condition to hold, $g(z,\theta)$ cannot be a regression residual. For instance, their identification condition does not cover the example we consider in Section 5. To identify the regression function through such an unconditional moment restriction, an appropriate instrument vector has to be specified. But a part of our original motivation was to avoid using arbitrary instruments. Finally, we develop a likelihood ratio-type test for parametric hypotheses and give a formal derivation of its asymptotic distribution. We also provide extensive simulation results. These results have not been provided by any of the papers cited above.

A word on notation. If V is a matrix, $\|V\| = \sqrt{tr(VV')}$ denotes its Frobenius norm. This reduces to the usual Euclidean norm in case V happens to be a vector. By a "vector" we mean a column vector. We do not make any notational distinction between a random vector and the value taken by it. The difference should be clear from the context. Unless mentioned otherwise, all limits are taken as $n \uparrow \infty$. The qualifier "with probability one" is abbreviated as "w.p.1".

2. The Estimator

Let $\{x_i, z_i\}_{i=1}^n$ be a random sample in $\mathbb{R}^s \times \mathbb{R}^d$. x is continuously distributed with Lebesgue density h, while z can be continuous, discrete, or mixed. Θ is a compact subset of $\mathbb{R}^p$ and $g(z,\theta) : \mathbb{R}^d \times \Theta \to \mathbb{R}^q$ is a vector of known functions. We consider the conditional moment restriction

$$\mathbb{E}\{g(z,\theta_0)|x\} = 0 \quad \text{w.p.1}, \tag{2.1}$$

where $\theta_0 \in int(\Theta)$ is the true parameter value. The goal is to efficiently estimate θ_0 under (2.1). This setup has numerous applications. See, for instance, Newey (1993). In particular, it may be used to model the linear or nonlinear conditional mean regression: Let $z = (x, y)$, where x is the vector of explanatory variables and y denotes the response variable. $g(z,\theta_0)$ is then simply the deviation of y from $\mathbb{E}(y|x)$; i.e. $g(z,\theta_0) = y - \mathbb{E}(y|x) = y - G(x,\theta_0)$ where G is known. More generally, we can apply this setup to separable models of the type $g(z,\theta_0) = \varepsilon$, where ε is a vector of unobserved errors. The nonlinear simultaneous equations model studied in Amemiya (1977) takes this form.

Notice that $g(z,\theta_0)$ is not correlated with any function of x in (2.1). Therefore, for a matrix of instrumental variables $v(x,\theta_0)$, (2.1) implies the unconditional moment restriction

$$\mathbb{E}\{v(x,\theta_0)g(z,\theta_0)\} = 0. \tag{2.2}$$

An interesting question is to find a v which yields an asymptotically efficient estimator of θ_0. Let[1] $D(x,\theta) = \mathbb{E}\{\frac{\partial g(z,\theta)}{\partial\theta}|x\}$ and $V(x,\theta) = \mathbb{E}\{g(z,\theta)g'(z,\theta)|x\}$. As shown in Chamberlain (1987), the asymptotic variance of any $n^{1/2}$-consistent regular estimator of θ_0 in (2.1) cannot be smaller than $I^{-1}(\theta_0)$, where $I(\theta_0) = \mathbb{E}\{D'(x,\theta_0)V^{-1}(x,\theta_0)D(x,\theta_0)\}$ denotes the minimal Fisher information for estimating θ_0 under (2.1). Using standard GMM theory, we can show that the lower bound $I^{-1}(\theta_0)$ is achieved by an optimal IV estimator which uses $v_*(x,\theta_0) = D'(x,\theta_0)V^{-1}(x,\theta_0)$ as the instruments in (2.2). But because θ_0 is unknown, as are usually the functional forms of D and V, an estimator using the "optimal instrument" v_* is infeasible. Newey (1993) proposed a feasible method of moments estimator which uses a preliminary estimator of θ_0 and estimates v_* nonparametrically. Under certain regularity conditions, Newey shows that his estimator is asymptotically efficient. However, in practice it is often difficult to find a well-behaved estimate of v_*. As a result, the feasible method of moments estimator can perform poorly.

In this paper we propose an alternative, yet asymptotically efficient, estimation technique which avoids estimating the optimal instruments. Our approach relies on the localized empirical likelihood.

[1]We denote the $q \times p$ Jacobian matrix of the partial derivatives of $g(z,\theta)$ with respect to θ as $\frac{\partial g(z,\theta)}{\partial\theta}$.

We use positive weights $w_{ij} = \frac{\mathcal{K}(\frac{x_i - x_j}{b_n})}{\sum_{j=1}^n \mathcal{K}(\frac{x_i - x_j}{b_n})} \stackrel{def}{=} \frac{\mathcal{K}_{ij}}{\sum_{j=1}^n \mathcal{K}_{ij}}$ to carry out the localization. For the sake of notational convenience, the dependence of w_{ij} and $\mathcal{K}_{ij}$ upon n is suppressed. The kernel function $\mathcal{K}$ is chosen to satisfy Assumption 3.3, and the bandwidth b_n is a null sequence of positive numbers such that $nb_n^s \uparrow \infty$[2]. In a b_n neighborhood of x_i, w_{ij} assigns smaller weights to those x_j's which are farther away from x_i.

Let p_{ij} be the probability mass placed at (x_i, z_j) by a discrete distribution that has support on $\{x_1, \ldots, x_n\} \times \{z_1, \ldots, z_n\}$. The reader can interpret p_{ij} as the conditional probability $\Pr\{z = z_j | x = x_i\}$. We start our estimation procedure by using the weights w_{ij} to obtain a "smoothed" log-likelihood $\sum_{i=1}^n \sum_{j=1}^n w_{ij} \log p_{ij}$. Next, for each $\theta \in \Theta$ we concentrate out the p_{ij}'s by solving

$$\tag{2.3} \begin{gathered} \max_{p_{ij}} \sum_{i=1}^n \sum_{j=1}^n w_{ij} \log p_{ij} \\ s.t. \quad p_{ij} \geq 0, \ \sum_{j=1}^n p_{ij} = 1, \ \sum_{j=1}^n g(z_j, \theta) p_{ij} = 0, \qquad \text{for } i, j = 1, \ldots, n. \end{gathered}$$

(2.3) can be conveniently solved by using Lagrange multipliers. The Lagrangian is[3]

$$\mathfrak{L}(\theta) = \sum_{i=1}^n \sum_{j=1}^n w_{ij} \log p_{ij} - \sum_{i=1}^n \mu_i (\sum_{j=1}^n p_{ij} - 1) - \sum_{i=1}^n \lambda_i' \sum_{j=1}^n g(z_j, \theta) p_{ij},$$

where $\mu_1, \ldots, \mu_n$ are the multipliers for the second set of constraints, and $\{\lambda_i \in \mathbb{R}^q : i = 1, \ldots, n\}$ the Lagrange multipliers for the third set of constraints. It is easily verified that the solution to (2.3) is

$$\tag{2.4} \hat{p}_{ij} = \frac{w_{ij}}{1 + \lambda_i' g(z_j, \theta)},$$

where, for each $\theta \in \Theta$, λ_i solves[4]

$$\tag{2.5} \sum_{j=1}^n \frac{w_{ij} g(z_j, \theta)}{1 + \lambda_i' g(z_j, \theta)} = 0, \qquad i = 1, \ldots, n.$$

Using (2.4), we define the smoothed empirical loglikelihood (SEL) at θ as

$$\text{SEL}(\theta) = \sum_{i=1}^n \sum_{j=1}^n \mathbb{T}_{i,n} w_{ij} \log \hat{p}_{ij} = \sum_{i=1}^n \sum_{j=1}^n \mathbb{T}_{i,n} w_{ij} \log \{ \frac{w_{ij}}{1 + \lambda_i' g(z_j, \theta)} \},$$

where λ_i solves (2.5), and $\mathbb{T}_{i,n}$ is a sequence of trimming functions which have been incorporated in the smoothed log-likelihood to deal with a technical problem. $\mathbb{T}_{i,n}$ will be defined shortly. Our "maximum smoothed empirical likelihood estimator" of θ_0 is defined as

$$\tag{2.6} \hat{\theta} = \underset{\theta \in \Theta}{\text{argmax}} \ \text{SEL}(\theta).$$

[2]Additional restrictions on the choice of b_n are described in Assumption 3.7.

[3]Since the objective function depends upon p_{ij} only through $\log p_{ij}$, the constraint $p_{ij} \geq 0$ does not bind.

[4]λ_i is shorthand for $\lambda(x_i, \theta)$. Its dependence upon θ is suppressed to reduce notation, and should not cause any confusion. However, when necessary, we explicitly write λ_i as $\lambda_i(\theta)$ to ensure that our arguments are unambiguous.

As noted above, the objective function $\text{SEL}(\theta)$ involves a trimming function. To see why trimming is necessary, let $\hat{h}(x_i) = \frac{1}{nb_n^s}\sum_{j=1}^n \mathcal{K}_{ij}$ denote the Nadaraya-Watson estimate of $h(x_i)$ and write $w_{ij} = \frac{\mathcal{K}_{ij}/(nb_n^s)}{\hat{h}(x_i)}$. The presence of the density estimate in the denominator means that the local log-empirical likelihood $\sum_{j=1}^n w_{ij}\log\hat{p}_{ij}$ may be ill-behaved for x's lying in the tails of h. This is the well-known "denominator problem" associated with kernel estimators. Different authors have used different approaches to deal with this problem. For instance, Robinson (1987) and Newey (1993) choose to avoid this problem altogether by using nearest neighbor estimators. However, since kernel estimators are mathematically and practically tractable, we retain them in this paper and deal with the denominator problem by trimming away small values of $\hat{h}(x_i)$. In this paper we use the indicator function $\mathbb{T}_{i,n} = \mathbb{I}\{\hat{h}(x_i) \geq b_n^\tau\}$ to do the trimming, where the trimming parameter $\tau \in (0,1)$. Lemma D.4 shows that if b_n and τ are chosen appropriately, $\mathbb{T}_{i,n} \xrightarrow{p} 1$ as $n \uparrow \infty$. Tripathi and Kitamura (2000) use a version of SEL which is trimmed over a fixed set to obtain a specification test for the validity of the conditional moment restriction $\mathbb{E}(z|x) = 0$.

Implementing our estimator is straightforward. From (2.5), it is easily seen that

$$\lambda_i = \underset{\gamma\in\mathbb{R}^q}{\operatorname{argmax}} \sum_{j=1}^n w_{ij}\log(1+\gamma' g(z_j,\theta)). \tag{2.7}$$

This is a well behaved optimization problem since the objective function is globally concave and can be solved by a simple Newton-Raphson numerical procedure. Once the λ_i's are calculated, $\hat{\theta}$ can be obtained by maximizing $\text{SEL}(\theta)$, which is equivalent to maximizing

$$-\sum_{i=1}^n\sum_{j=1}^n \mathbb{T}_{i,n} w_{ij}\log\{1+\lambda_i' g(z_j,\theta)\} = -\sum_{i=1}^n \mathbb{T}_{i,n}\max_{\gamma\in\mathbb{R}^q}\sum_{j=1}^n w_{ij}\log\{1+\gamma' g(z_j,\theta)\}$$

with respect to $\theta \in \Theta$. This "outer loop" minimization can be carried out using a numerical optimization procedure.

Finally, we comment upon a normalization-invariance property of $\hat{\theta}$. Let $A(x_i,\theta)$ be a $q \times q$ matrix which, for each $\theta \in \Theta$, is nonsingular w.p.1. The null set on which $A(x_i,\theta)$ is singular may depend upon θ. Obviously, the conditional mean restriction in (2.1) remains unaltered if $g(z,\theta_0)$ is replaced by $A(x,\theta_0)g(z,\theta_0)$. A nice feature of $\hat{\theta}$ is that it is invariant to such normalizations since the normalization factor $A(x_i,\theta)$ is simply absorbed into $\lambda_i \equiv \lambda(x_i,\theta)$ in (2.4). Note that the two-step estimators proposed in Robinson (1987) and Newey (1993) do not share this normalization-invariance property.

3. Large Sample Theory

In this section we present some asymptotic results for the maximum smoothed empirical likelihood estimator of θ_0 defined in (2.6). In addition to the previously defined symbols, the following notation is also used in the rest of the paper: $\mathbb{S}^a = \{\xi \in \mathbb{R}^a : \|\xi\| = 1\}$ is the unit sphere in $\mathbb{R}^a$, $x^{(i)}$ denotes the i^{th} component of the vector x, and $M^{(ij)}$ is the $(i,j)^{th}$ element of a matrix M. ∇_θ (the subscript indicates that differentiation is with respect to θ) is the gradient operator; i.e. $\nabla_\theta g(z,\theta) = \frac{\partial g'(z,\theta)}{\partial\theta}$, where $\frac{\partial g'(z,\theta)}{\partial\theta}$ denotes the transpose of $\frac{\partial g(z,\theta)}{\partial\theta}$. Obviously, $\nabla_\theta g(z,\theta)$ is a $p \times q$

matrix. If $f(\theta)$ is scalar valued, then the gradient $\nabla_\theta f(\theta)$ is a $p \times 1$ vector while the Hessian $\nabla_{\theta\theta} f(\theta)$ is a $p \times p$ matrix.

The following regularity conditions help us in doing asymptotic analysis.

Assumption 3.1. *For each $\theta \neq \theta_0$ there exists a set $\mathcal{X}_\theta \subseteq \mathbb{R}^s$ such that $\Pr\{x \in \mathcal{X}_\theta\} > 0$, and $\mathbb{E}\{g(z,\theta)|x\} \neq 0$ for every $x \in \mathcal{X}_\theta$.* □

Assumption 3.1 guarantees the identification of θ_0. It differs from the identification condition in Newey (1993) because here we provide a proof of the consistency of a fully iterative estimation procedure based on a global parameter search, while Newey considers an estimator obtained from one Newton-Raphson iteration using a preliminary consistent estimate.

Assumption 3.2. $\mathbb{E}\{\sup_{\theta\in\Theta} \|g(z,\theta)\|^m\} < \infty$ *for some* $m \geq 8$. □

$m = 8$ is used in the proof of Lemma B.6.

Assumption 3.3. *For $x = (x^{(1)}, \ldots, x^{(s)})$, let $\mathcal{K}(x) = \Pi_{i=1}^s \kappa(x^{(i)})$. Here $\kappa : \mathbb{R} \to \mathbb{R}$ is a continuously differentiable pdf with support $[-1, 1]$. κ is symmetric about the origin, and for some $a \in (0, 1)$ is bounded away from zero on $[-a, a]$.* □

$\mathcal{K}$ belongs to the class of second order product kernels. Since these kernels are employed to estimate probabilities, the use of kernels with order greater than two is ruled out. Furthermore, the nonnegativity of $\mathcal{K}$ is also explicitly used several times. See, for instance, the proof of Lemma B.1. Continuous differentiability of $\mathcal{K}$ allows us to use the uniform convergence rates for kernel estimators in Ai (1997). The requirement that $\mathcal{K}$ be bounded away from zero on a closed ball centered at the origin, allows us to use the consistency result of Devroye and Wagner (1980) in the proof of Lemma B.2.

Assumption 3.4. *Assume that*:

(*i*) $0 < h(x) \leq \sup_{x\in\mathbb{R}^s} h(x) < \infty$, $h \in \mathrm{C}^2(\mathbb{R}^s)$, *and* $\sup_{x\in\mathbb{R}^s} \|\nabla_{xx} h(x)\| < \infty$.

(*ii*) $\mathbb{E}\|x\|^{1+\varrho} < \infty$ *for some* $\varrho > 0$.

(*iii*) $\theta \mapsto g(z,\theta)$ *is continuous on* Θ *w.p.1, and* $\mathbb{E}\{\sup_{\theta\in\Theta} \|\frac{\partial g(z,\theta)}{\partial\theta}\|\} < \infty$.

(*iv*) $(\theta, x) \mapsto \|\nabla_{xx}\{\mathbb{E}[g^{(l)}(z,\theta)|x]h(x)\}\|$ *is uniformly bounded on* $\Theta \times \mathbb{R}^s$ *for* $1 \leq l \leq q$. □

Uniform boundedness of h and its second derivatives along with (*ii*) is used, for instance, in the proofs of Lemmas B.3 and D.8. (*iii*) and (*iv*) are useful when showing the consistency of $\hat{\theta}$.

Assumption 3.5. *There exists a closed ball $\mathcal{B}_0$ around θ_0 such that for $1 \leq i, r \leq q$ and $1 \leq j, k \leq p$*:

(*i*) $\theta \mapsto D(x,\theta)$ *and* $\theta \mapsto V(x,\theta)$ *are continuous on* $\mathcal{B}_0$ *w.p.1.*

(*ii*) $\inf_{(\xi,x,\theta)\in\mathbb{S}^q\times\mathbb{R}^s\times\mathcal{B}_0} \xi' V(x,\theta)\xi > 0$ *and* $\sup_{(\xi,x,\theta)\in\mathbb{S}^q\times\mathbb{R}^s\times\mathcal{B}_0} \xi' V(x,\theta)\xi < \infty$.

(*iii*) $\sup_{\theta\in\mathcal{B}_0} |\frac{\partial g^{(i)}(z,\theta)}{\partial\theta^{(j)}}| \leq d(z)$ *and* $\sup_{\theta\in\mathcal{B}_0} |\frac{\partial^2 g^{(i)}(z,\theta)}{\partial\theta^{(j)}\,\partial\theta^{(k)}}| \leq l(z)$ *hold w.p.1 for some real valued functions* $d(z)$ *and* $l(z)$ *such that* $\mathbb{E}d^\eta(z) < \infty$ *for some* $\eta \geq 4$, *and* $\mathbb{E}l(z) < \infty$.

(*iv*) $\sup_{(x,\theta)\in\mathbb{R}^s\times\mathcal{B}_0} \|\nabla_{xx}\{D^{(ij)}(x,\theta)h(x)\}\| < \infty$.

(*v*) $\sup_{(x,\theta)\in\mathbb{R}^s\times\mathcal{B}_0} \|\nabla_{xx}\{V^{(ir)}(x,\theta)h(x)\}\| < \infty$. □

(*i*), (*ii*), and (*iii*) imply $\theta \mapsto I(\theta)$ is continuous on $\mathcal{B}_0$. By (*ii*), $\sup_{(x,\theta)\in\mathbb{R}^s\times\mathcal{B}_0} \|V^{-1}(x,\theta)\| < \infty$ and $\sup_{(x,\theta)\in\mathbb{R}^s\times\mathcal{B}_0} \mathbb{E}\{\|g(z,\theta)\|^2|x\} < \infty$. These facts are used in the proofs. In (*iii*), existence of $d(z)$ ensures that $\mathbb{E}\|D(x,\theta_0)\|^\eta < \infty$. $\eta = 4$ is used in the proof of Theorem 3.2. (*iv*) and (*v*) are used, for instance, in the proofs of Lemma B.5 and Lemma B.6 respectively.

Assumption 3.6. *When solving* (2.5) *for* $\lambda_1, \ldots, \lambda_n$*, we only search over the set* $\{\gamma \in \mathbb{R}^q : \|\gamma\| \leq \bar{c}n^{-1/m}\}$ *for some* $\bar{c} > 0$. □

This is similar to Assumption 4.2(b) of Newey and Smith (2000). Since the λ_i's converge to zero under (2.1), when solving (2.5) for $\lambda_1, \ldots, \lambda_n$ it is reasonable to search for the solution in some neighborhood of the origin. Because $\Pr\{\max_{1\leq j\leq n} \sup_{\theta\in\Theta} \|g(z_j,\theta)\| = o(n^{1/m})\} = 1$ as $n \uparrow \infty$ by Lemma D.2, restricting the λ_i's to a $n^{-1/m}$-neighborhood of the origin ensures that

$$\Pr\{\max_{1\leq i,j\leq n} \sup_{\theta\in\Theta} |\lambda_i' g(z_j,\theta)| = o(1)\} = 1 \quad \text{as } n \uparrow \infty. \tag{3.1}$$

This, for instance, is used in the proof of Theorem 3.2. Note that we only need Assumption 3.6 to establish the asymptotic normality of $\hat{\theta}$. We prove consistency of $\hat{\theta}$ without using Assumption 3.6.

Finally, the following assumption collects the conditions on ϱ, τ, and b_n under which our consistency and asymptotic normality results hold.

Assumption 3.7. *Let* $\tau \in (0,1)$, $\varrho \geq 1.5/m + 1/4$, $b_n \downarrow 0$, *and* $\beta \in (0,1)$ *such that:* $n^{1-\beta} b_n^{(\frac{\eta+2}{\eta-2})\frac{s}{2}} \uparrow \infty$, $n^{1-2\beta-2/m} b_n^{2s+4\tau} \uparrow \infty$, $n^{\varrho} b_n^{2\tau} \uparrow \infty$, $n^{\varrho-1/\eta} b_n^{\tau} \uparrow \infty$, *and* $n^{\varrho-2/m} b_n^{\tau} \uparrow \infty$. □

$\tau < 1$ is required in the proof of Lemma D.5, and $\varrho \geq 1.5/m + 1/4$ is used in the proof of Theorem 3.2. $b_n \downarrow 0$ and $nb_n^s \uparrow \infty$, the latter following when $n^{1-\beta} b_n^{(\frac{\eta+2}{\eta-2})\frac{s}{2}} \uparrow \infty$, are standard conditions on the bandwidth to ensure consistency of kernel estimators. The parameter β appears because we are using uniform convergence rates for kernel estimators due to Ai (1997).

We are now ready to present our findings. The first result shows that $\hat{\theta}$ is consistent.

Theorem 3.1. *Let Assumptions 3.1–3.5 and 3.7 hold. Then* $\hat{\theta} \xrightarrow{p} \theta_0$.

Next comes asymptotic normality.

Theorem 3.2. *Let Assumptions 3.1–3.7 hold. Then* $n^{1/2}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathrm{N}(0, I^{-1}(\theta_0))$.

$I^{-1}(\theta_0)$ coincides with the efficiency bound in Chamberlain (1987) for estimating θ_0 under (2.1). Therefore, $\hat{\theta}$ is asymptotically efficient.

4. Hypothesis Testing

We now consider testing restrictions on θ_0. While it is straightforward to define an analog of the Wald test by using an estimate of $I^{-1}(\theta_0)$, obtaining good estimates of $I(\theta_0)$ can be difficult. Furthermore, explicit studentization destroys "implicit pivotalness", which is one of the attractive features of empirical likelihood. A more natural approach which fully exploits the pseudo-likelihood character of our methodology is to construct an analog of the conventional parametric likelihood ratio test. In the parametric likelihood framework, Wilks's theorem enables us to conduct asymptotic χ^2 inference based on the likelihood ratio test. We extend Wilks's theorem to models with conditional moment restrictions.

Suppose we want to test the parametric restriction $\mathrm{H}_0 : R(\theta_0) = 0$ against $\mathrm{H}_1 : R(\theta_0) \neq 0$, where $R(\theta_0)$ is a $r \times 1$ vector and $r \leq p$. The constrained version of $\hat{\theta}$ is

$$\hat{\theta}^R = \operatorname*{argmax}_{\theta\in\Theta} \mathrm{SEL}(\theta) \quad s.t. \quad R(\theta) = 0,$$

where $\mathrm{SEL}(\theta)$ was defined earlier. A SEL version of the likelihood ratio statistic for testing H_0 is then

$$\mathrm{LR}_n = 2\{\mathrm{SEL}(\hat{\theta}) - \mathrm{SEL}(\hat{\theta}^R)\}.$$

To get some intuition behind the limiting behavior of LR_n, consider testing the simple hypothesis $\mathrm{H}_0^* : \theta_0 = \bar{\theta}$ against $\mathrm{H}_1^* : \theta_0 \neq \bar{\theta}$, where $\bar{\theta}$ is known. The restricted estimator is now $\hat{\theta}^R = \bar{\theta}$, and LR_n reduces to $\mathrm{LR}_n = 2\{\mathrm{SEL}(\hat{\theta}) - \mathrm{SEL}(\bar{\theta})\}$. Taylor expand $\mathrm{SEL}(\bar{\theta})$ around $\hat{\theta}$ to get

$$\mathrm{SEL}(\bar{\theta}) = \mathrm{SEL}(\hat{\theta}) + (\bar{\theta} - \hat{\theta})'\nabla_\theta \mathrm{SEL}(\hat{\theta}) + \frac{1}{2}(\bar{\theta} - \hat{\theta})'\nabla_{\theta\theta}\mathrm{SEL}(\theta^*)(\bar{\theta} - \hat{\theta}),$$

for some θ^* between $\bar{\theta}$ and $\hat{\theta}$. But from (2.6) we know $\nabla_\theta \mathrm{SEL}(\hat{\theta}) = 0$, and Lemma C.1 shows that $\| - \nabla_{\theta\theta}\mathrm{SEL}(\theta^*)/n - I(\theta_0)\| = o_p(1)$. Therefore, by Theorem 3.2 it is straightforward to see that $2\{\mathrm{SEL}(\hat{\theta}) - \mathrm{SEL}(\bar{\theta})\} \xrightarrow{d} \chi_p^2$ under H_0^*. To handle the general case, we make the following assumption.

Assumption 4.1. *$R : \Theta \to \mathbb{R}^r$ is twice continuously differentiable and $\frac{\partial R(\theta_0)}{\partial \theta}$ has rank r.* □

The asymptotic distribution of LR_n is then given by the following result.

Theorem 4.1. *Let Assumptions 3.1–4.1 hold. Then $\mathrm{LR}_n \xrightarrow{d} \chi_r^2$ under H_0.*

We can also invert LR_n to construct asymptotically valid confidence intervals. For example, if one is interested in constructing a confidence interval for the j^{th} component of θ_0, treating the other components as nuisance parameters, an approximate $(1-\alpha)$ level confidence interval is given by

$$\{\theta^{(j)} : \min_{\theta^{(1)},\ldots,\theta^{(j-1)},\theta^{(j+1)},\ldots\theta^{(p)}} 2[\mathrm{SEL}(\hat{\theta}) - \mathrm{SEL}(\theta)] \leq u_\alpha\},$$

where u_α satisfies $P(\chi_1^2 \geq u_\alpha) = \alpha$. As in Qin and Lawless (1994) and Kitamura and Stutzer (1997), it is also possible to construct Lagrange Multiplier and Wald-type statistics, although these alternatives are less attractive because LR_n achieves pivotalness without requiring the estimation of variance.

It is straightforward to see that confidence intervals based on LR_n are invariant to nonsingular transformations of the moment conditions. They also automatically satisfy natural range restrictions. See a related discussion by Owen (1990, Section 3.2) for models with unconditional moment restrictions. Empirical likelihood has other nice theoretical properties such as Bartlett correctability and GNP-optimality at least in unconditional moment models. It is reasonable to expect that some of these features would carry over to the smoothed empirical likelihood approach considered here, although it is a technically challenging task to establish them rigorously.

Finally, it is also useful to note that even though $\mathrm{SEL}(\theta)$ was obtained on nonparametric considerations, it behaves very much like a parametric likelihood. This can be seen from Theorem 3.2, which shows that maximizing $\mathrm{SEL}(\theta)$ leads to an asymptotically efficient estimator of θ_0. Additional support is provided by Lemma C.1, which demonstrates that the "observed information" $-\nabla_{\theta\theta}\mathrm{SEL}(\hat{\theta})/n$ converges in probability to $I(\theta_0)$, the minimal Fisher information for estimating θ_0 in (2.1).

5. Monte Carlo Experiment

We now compare our estimator with some competitors using a Monte Carlo experiment. This experiment also provides some guidance regarding the choice of bandwidth for our estimator in practice. Our simulation design basically follows Cragg (1983). This design, also used in Newey (1993), is a linear model with heteroscedastic errors; namely,

$$y_i = \beta_1 + \beta_2 x_i + u_i, \quad u_i = \varepsilon_i \sqrt{0.1 + 0.2x_i + 0.3x_i^2}. \tag{5.1}$$

Here the true $\beta_1 = \beta_2 = 1$, $\ln(x_i) \sim \mathrm{N}(0,1)$, $\varepsilon_i \sim \mathrm{N}(0,1)$, and x_i and ε_i are independent. The number of replications is set to 500. Simulation results not reported here show that the performance of our estimator is relatively insensitive to the choice of the trimming parameter τ. Hence in this experiment we set $\mathbb{T}_{i,n} = 1$ for each i; i.e. we do not trim $\hat{h}$ when computing our estimator. Following Newey (1993), we also report estimates for β_1 and β_2 using ordinary least squares (OLS), (infeasible) generalized least squares (GLS), and feasible GLS (FGLS). Note that FGLS requires the knowledge of the functional form of the heteroscedasticity, while GLS requires perfect knowledge of the heteroscedasticity function.

The label "k-NN" denotes Newey's semiparametric efficient IV estimator where the heteroscedasticity function is estimated by nearest neighbor methods. For details about FGLS and "k-NN", the reader is referred to Newey (1993). The label "kernel" refers to an estimator similar to "k-NN," the only difference being that Nadaraya-Watson estimators are used in place of nearest neighbor estimators. Interestingly, "kernel" works favorably compared with "k-NN," as mentioned below. The final estimator we consider is the new estimator (2.6), denoted by "SEL".

In general, comparing semiparametric estimators is tricky since they depend on the choice of nonparametric techniques (e.g., nearest neighbor or kernel), as well as the choice of bandwidth parameters. Calculating "kernel" is therefore useful, because it enables us to compare a Newey type semiparametric estimator with our estimator using the same nonparametric regression methodology.

Newey's semiparametric IV estimator (with nearest neighbor or kernel) and our estimator depend on the choice of the number of nearest neighbors (denoted by k_n in the tables in Appendix E) or the bandwidth (b_n). The tables contain results with reasonable range of k_n's and b_n's. Also, the rows labeled "automatic" are obtained by choosing k and b_n by a cross-validation procedure suggested in Newey (1993).

Following Newey (1993), we use infeasible GLS as our baseline. "Ratio RMSE," for example, refers to the ratio of the RMSE of an estimator relative to that of GLS. For each estimator, the first (second) row corresponds to the estimate for β_1 (β_2). The results for OLS, FGLS, and "k-NN" in the tables match Newey (1993)'s simulation results with a reasonable degree of accuracy.

Tables 1 and 2 show results for $n = 50$. OLS is clearly inefficient, and FGLS works well, given the small sample size. The performance of "k-NN" and "kernel" is in between OLS and FGLS, although "kernel" works slightly better than "k-NN." SEL is as flexible as "k-NN" and "kernel" in terms of the treatment of heteroscedasticity, but its performance is better than these two. Notice that this good relative performance of SEL holds at each b_n over the range of bandwidths considered here. Naturally, SEL continues to work best among the three semiparametric estimators when cross-validation is used. For example, "Ratio RMSE" of SEL for β_2 is 1.22, whereas for "k-NN" and "kernel" it is 1.59 and 1.57, respectively. With $n = 100$ (Tables 3 and 4), all of the semiparametric estimators ("k-NN", "kernel", and SEL) behave well, though the performance of SEL is still considerably better than the other two. With $n = 200$ (Tables 5 and 6), SEL works remarkably well. After cross-validation, its RMSE and

MAE are only 9.5% larger than those for GLS. Recall that SEL achieves this excellent performance without using any knowledge of the optimal IV. It should also be noted that SEL is robust with respect to the choice of bandwidth. For instance, "Ratio RMSE" of SEL for β_2 is between 1.09 and 1.25 (i.e., an efficiency loss of 9% to 25%). The other estimators sometimes have large "Ratio RMSE" depending upon the bandwidth. In summary, our empirical likelihood-based estimator performs very well, at least within Cragg's simulation design considered here. Even though the performance of the estimators varies with the choice of bandwidth, SEL outperforms other estimators uniformly over the range of bandwidths used in our experiment.

6. Conclusion

In this paper we show how to extend the empirical likelihood methodology to estimate models with conditional moment restrictions. By using a localized version of empirical likelihood, we obtain a new normalization-invariant estimator that achieves the semiparametric efficiency bound automatically; i.e. without estimating the optimal instruments explicitly. The smoothed empirical likelihood approach also lends itself naturally to hypothesis testing. In particular, we propose a likelihood ratio type statistic for testing parametric restrictions. This statistic does not require the estimation of any variance term and we demonstrate that it achieves asymptotic pivotalness implicitly. Finally, we carry out a Monte Carlo experiment to examine the efficacy of our estimator in finite samples. Simulation results show that our estimator works remarkably well in practice when compared with some competing estimators. $\square$

Appendix A. Proofs of Main Results

Notation. Henceforth, the letter c denotes a generic constant which may vary from case to case. Furthermore, $B(\theta, \epsilon)$ denotes an open ball of radius ϵ centered at θ, $\hat{V}(x_i, \theta) = \sum_{j=1}^n w_{ij} g(z_j, \theta) g'(z_j, \theta)$, $\hat{\Omega}(x_i, \theta) = \frac{1}{nb_n^s} \sum_{j=1}^n \mathcal{K}_{ij} g(z_j, \theta) g'(z_j, \theta)$, $\Omega(x_i, \theta) = V(x_i, \theta) h(x_i)$, $\mathcal{K}_{\max} = \sup_{x \in [-1,1]^s} \mathcal{K}(x)$, $S_n = \{x \in \mathbb{R}^s : \|x\| \leq n\}$, $\hat{\mathbb{T}}_{i,n} = \mathbb{T}_{i,n} h(x_i)/\hat{h}(x_i)$, $\mathbb{I}_{i,n} = \mathbb{I}\{x_i \in S_n\}$, $\mathbb{I}_{i,n}^c = 1 - \mathbb{I}_{i,n}$, $g_*(z) = \sup_{\theta \in \Theta} \|g(z, \theta)\|$, and $\nabla_\theta g'(z, \theta) = \frac{\partial g(z,\theta)}{\partial \theta}$. The qualifier "with probability approaching one" is abbreviated as "w.p.a.1". $\square$

Proof of Theorem 3.1. Our consistency proof utilizes the approach developed in Kitamura (1997b) and Kitamura and Stutzer (1997). Recall that $\hat{\theta}$ maximizes the objective function

$$G_n(\theta) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n -\mathbb{T}_{i,n} w_{ij} \log(1 + \lambda_i'(\theta) g(z_j, \theta)).$$

For a constant $\tilde{c} \in (0, 1)$, define $C_n = \{z : \sup_{\theta \in \Theta} \|g(z, \theta)\| \leq \tilde{c} n^{1/m}\}$ and $g_n(z, \theta) = \mathbb{I}\{z \in C_n\} g(z, \theta)$. Define $u(x, \theta) = \mathbb{E}[g(z, \theta)|x]/\|\mathbb{E}[g(z, \theta)|x]\|$ with the convention that $0/0 = 0$. Let $q_n(x, z, \theta) = -\log(1 + n^{-1/m} u'(x, \theta) g_n(z, \theta))$, $f_\delta(x, z, \theta) = \sup_{\theta^* \in B(\theta, \delta)} u'(x, \theta^*) g(z, \theta^*)$, $f(x, z, \theta) = u'(x, \theta) g(z, \theta)$ and $H(\theta, \delta) = \mathbb{E}[f_\delta(x, z, \theta)]/3$. Note $\lim_{\delta \downarrow 0} f_\delta(x, z, \theta) \geq f(x, z, \theta)$ for all x, z and θ. This and the Monotone Convergence Theorem imply that

$$\lim_{\delta \downarrow 0} 3H(\theta, \delta) = \lim_{\delta \downarrow 0} \mathbb{E}[f_\delta(x, z, \theta)] = \mathbb{E}[\lim_{\delta \downarrow 0} f_\delta(x, z, \theta)] \geq \mathbb{E}[f(x, z, \theta)],$$

where the last term is bounded by

$$\mathbb{E}f(x,z,\theta) = \mathbb{E}[u'(x,\theta)g(z,\theta)] \geq \mathbb{E}[\{\mathbb{I}\{x \in \mathcal{X}_\theta\}\|\mathbb{E}[g(z,\theta)|x]\|].$$

But by Assumption (3.1), $\mathbb{E}[\{\mathbb{I}\{x \in \mathcal{X}_\theta\}\|\mathbb{E}[g(z,\theta)|x]\|] > 0$ at each $\theta \neq \theta_0$. Therefore, for sufficiently small δ, $H(\theta,\delta) > 0$ at each $\theta \neq \theta_0$. By the mean value theorem, for some $t \in (0,1)$,

$$q_n(x,z,\theta) = -n^{-1/m}u'(x,\theta)g(z,\theta) + R_n(t), \tag{A.1}$$

where $R_n(t) = n^{-1/m}u'(x,\theta)g(z,\theta)(1-\mathbb{I}\{z \in C_n\}) + \frac{n^{-2/m}\|u'(x,\theta)g_n(z,\theta)\|^2}{2(1-tn^{-1/m}u'(x,\theta)g_n(z,\theta))^2}$. By repeated applications of the Cauchy-Schwarz inequality,

$$|R_n(t)| \leq n^{-1/m}\sup_{\theta\in\Theta}\|g(z,\theta)\|(1-\mathbb{I}\{z \in C_n\}) + \frac{1}{2(1-\tilde{c})^2}n^{-2/m}\sup_{\theta\in\Theta}\|g(z,\theta)\|^2. \tag{A.2}$$

Therefore $\mathbb{E}|n^{1/m}R_n(t)|$ can be made arbitrarily small by choosing large enough n; in particular, we can find an integer $n(\theta,\delta)$ such that $\mathbb{E}|n^{1/m}R_n(t)| \leq H(\theta,\delta)$ for all $n \geq n(\theta,\delta)$. By (A.1), for sufficiently small δ,

$$n^{1/m}\mathbb{E}\sup_{\theta^*\in B(\theta,\delta)} q_n(x,z,\theta^*) \leq -3H(\theta,\delta) + \mathbb{E}|n^{1/m}R_n(t)| \leq -2H(\theta,\delta) < 0,$$

at each $\theta \neq \theta_0$ for all $n > n(\theta,\delta)$. By the compactness of Θ, we can find a finite number of open balls $B(\theta_k,\delta)$ that cover $\Theta_\delta \stackrel{def}{=} \Theta\backslash B(\theta_0,\delta)$ and also satisfy

$$n^{1/m}\mathbb{E}[\sup_{\theta^*\in B(\theta_k,\delta)} q_n(x,z,\theta^*)] \leq -2H_k(\delta), \quad n \geq n_k(\delta), \tag{A.3}$$

for $H_k(\delta) = H(\theta_k,\delta)$, $n_k(\delta) = n(\theta_k,\delta)$, and $k = 1,...,K$. Now define $q_n(x,\theta) = \mathbb{E}[q_n(x,z,\theta)|x]$. Then

$$\mathbb{E}[\sup_{\theta^*\in B(\theta_k,\delta)} q_n(x,\theta^*)] \leq \mathbb{E}[\mathbb{E}[\sup_{\theta^*\in B(\theta_k,\delta)} q_n(x,z,\theta^*)|x]] = \mathbb{E}[\sup_{\theta^*\in B(\theta_k,\delta)} q_n(x,z,\theta^*)],$$

which means that (A.3) continues to hold if we replace $q_n(x,z,\theta^*)$ with $q_n(x,\theta^*)$. With this result and a pointwise weak law of large numbers, there exists a large enough $\bar{n}_k = \bar{n}_k(\delta)$ such that

$$\Pr\{\frac{1}{n}\sum_{i=1}^n \sup_{\theta^*\in B(\theta_k,\delta)} q_n(x_i,\theta^*) > -n^{-1/m}H_k(\delta)\} < \delta/(2K), \quad k = 1,...,K,$$

for all $n > \bar{n}_k(\delta)$. These K inequalities imply that

$$\Pr\{\sup_{\theta^*\in\Theta_\delta}\frac{1}{n}\sum_{i=1}^n q_n(x_i,\theta^*) > -n^{-1/m}H(\delta)\} < \delta/2, \quad H(\delta) = \min_k H_k(\delta),$$

for all $n \geq \bar{n}(\delta) = \max_k \bar{n}_k(\delta)$. Applying Lemma D.4 to this we obtain

$$\Pr\{\sup_{\theta^*\in\Theta_\delta}\frac{1}{n}\sum_{i=1}^n \mathbb{T}_{i,n}q_n(x_i,\theta^*) > -n^{-1/m}H(\delta)\} < \delta/2$$

eventually. By Lemma D.2, $\max_{1\le j\le n}\sup_{\theta^*\in B(\theta,\delta)}\|g(z_j,\theta^*)\| = o(n^{1/m})$ w.p.1 as $n\uparrow\infty$. This justifies the use of g in place of g_n after the second equality below (almost surely) for sufficiently large n:

$$\begin{aligned}\sup_{\theta^*\in\Theta_\delta}\frac{1}{n}\sum_{i=1}^n \mathbb{T}_{i,n}q_n(x_i,\theta^*) &= \sup_{\theta^*\in\Theta_\delta}\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n \mathbb{T}_{i,n}w_{ij}q_n(x_i,z_j,\theta^*) + o_p(n^{-1/m})\\ &= \sup_{\theta^*\in\Theta_\delta}\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n -\mathbb{T}_{i,n}w_{ij}\log(1+n^{-1/m}u'(x_i,\theta^*)g(z_j,\theta^*)) + o_p(n^{-1/m})\\ &\ge \sup_{\theta^*\in\Theta_\delta}\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n -\mathbb{T}_{i,n}w_{ij}\log(1+\lambda_i'(\theta^*)g(z_j,\theta^*)) + o_p(n^{-1/m})\\ &= \sup_{\theta^*\in\Theta_\delta} G_n(\theta^*) + o_p(n^{-1/m}),\end{aligned}$$

where the first equality follows from Lemma B.8, and the inequality follows from the optimality of λ_i's (see (2.7)). We conclude that

$$\text{(A.4)}\qquad \Pr\{\sup_{\theta\in\Theta\backslash B(\theta_0,\delta)} G_n(\theta) > -n^{-1/m}H(\delta)\} < \delta/2 \quad \text{eventually.}$$

Next, we evaluate G_n at the true value θ_0. Note that $\max_{1\le i\le n}\mathbb{T}_{i,n}\|\lambda_i(\theta_0)\| = o_p(\sqrt{\frac{n^\beta}{nb_n^{s+2\tau}}}) + o_p(\frac{1}{n^{\varrho-1/m}})$ follows by (B.4). Use Lemma B.3 to obtain

$$\begin{aligned}G_n(\theta_0) &\ge -\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n \mathbb{T}_{i,n}w_{ij}\log(1+\lambda_i'(\theta_0)g(z_j,\theta_0)) \ge -\frac{1}{n}\sum_{i=1}^n \mathbb{T}_{i,n}\lambda_i'(\theta_0)\sum_{j=1}^n w_{ij}g(z_j,\theta_0)\\ &= \{o_p(\sqrt{\frac{n^\beta}{nb_n^{s+2\tau}}}) + o_p(\frac{1}{n^{\varrho-1/m}})\}\{o_p(\sqrt{\frac{n^\beta}{nb_n^{s+2\tau}}}) + o_p(\frac{1}{n^{\varrho-1/m}})\} \stackrel{def}{=} o_p(d_n^2).\end{aligned}$$

Therefore,

$$\text{(A.5)}\qquad \Pr\{G_n(\theta_0) < -d_n^2 H(\delta)\} < \delta/2 \quad \text{eventually.}$$

Under our conditions $n^{1/m}d_n^2\downarrow 0$. Thus by (A.4) and (A.5), for any $\delta>0$ there exists a positive integer $n_0(\delta)$ such that $\Pr\{\hat\theta\in B(\theta_0,\delta)\}\ge 1-\delta$ for all $n>n_0(\delta)$. The proof is complete. □

Proof of Theorem 3.2. The first order condition for (2.6) is $\nabla_\theta\mathrm{SEL}(\hat\theta)=0$. By a Taylor expansion,

$$\text{(A.6)}\qquad 0 = n^{-1/2}\nabla_\theta\mathrm{SEL}(\theta_0) + \frac{1}{n}\nabla_{\theta\theta}\mathrm{SEL}(\theta^*)\, n^{1/2}(\hat\theta-\theta_0)$$

for some θ^* between $\hat\theta$ and θ_0. From (C.1), $-\nabla_\theta\mathrm{SEL}(\theta_0) = \sum_{i=1}^n\sum_{j=1}^n \frac{\mathbb{T}_{i,n}w_{ij}[\nabla_\theta g(z_j,\theta_0)]\lambda_i(\theta_0)}{1+\lambda_i'(\theta_0)g(z_j,\theta_0)}$. Thus by Lemma B.1 we can write $-n^{-1/2}\nabla_\theta\mathrm{SEL}(\theta_0) = n^{-1/2}\hat A + n^{-1/2}\sum_{i=1}^n\sum_{j=1}^n\frac{\mathbb{T}_{i,n}w_{ij}\nabla_\theta g(z_j,\theta_0)r_i}{1+\lambda_i'(\theta_0)g(z_j,\theta_0)}$, where

$$\hat A \stackrel{def}{=} \sum_{i=1}^n \mathbb{T}_{i,n}\{\sum_{j=1}^n \frac{w_{ij}}{1+\lambda_i'(\theta_0)g(z_j,\theta_0)}\frac{\partial g'(z_j,\theta_0)}{\partial\theta}\}\hat V^{-1}(x_i,\theta_0)\{\sum_{j=1}^n w_{ij}g(z_j,\theta_0)\}.$$

Now we can use (3.1) to show that

$$\text{(A.7)}\qquad \max_{1\le i,j\le n}\sup_{\theta\in\Theta}\frac{1}{|1+\lambda_i' g(z_j,\theta)|} = O(1) \quad \text{holds w.p.1 as } n\uparrow\infty.$$

Thus by (A.7) and Assumption 3.5(iii)

$$\|\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\mathbb{T}_{i,n}w_{ij}\nabla_\theta g(z_j,\theta_0)r_i}{1+\lambda_i'(\theta_0)g(z_j,\theta_0)}\| = O(1)\max_{1\le i\le n}\mathbb{T}_{i,n}\|r_i\|\sum_{i=1}^{n}\sum_{j=1}^{n}d(z_j)w_{ij},$$

where the $O(1)$ term does not depend upon i, j, or $\theta\in\Theta$. Hence by Lemma B.1 and Lemma D.6

$$n^{-1/2}\|\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\mathbb{T}_{i,n}w_{ij}\nabla_\theta g(z_j,\theta_0)r_i}{1+\lambda_i'(\theta_0)g(z_j,\theta_0)}\| = o_p(\sqrt{\frac{n^{2\beta+2/m}}{nb_n^{2s+4\tau}}}) + o_p(\frac{1}{n^{2\varrho-3/m-1/2}}) = o_p(1),$$

since $\sqrt{\frac{n^{2\beta+2/m}}{nb_n^{2s+4\tau}}}\downarrow 0$ and $\varrho\ge 1.5/m+1/4$ under our conditions. It follows that

$$-n^{-1/2}\nabla_\theta\mathrm{SEL}(\theta_0) = n^{-1/2}\hat{A} + o_p(1). \tag{A.8}$$

Next, write $\hat{A} = A+\Delta$, where

$$A \stackrel{def}{=} \sum_{i=1}^{n}\mathbb{T}_{i,n}(\sum_{j=1}^{n}w_{ij}\frac{\partial g'(z_j,\theta_0)}{\partial\theta})\hat{V}^{-1}(x_i,\theta_0)\,(\sum_{j=1}^{n}w_{ij}g(z_j,\theta_0)), \quad \text{and} \tag{A.9}$$

$$\Delta \stackrel{def}{=} \sum_{i=1}^{n}\mathbb{T}_{i,n}(\sum_{j=1}^{n}w_{ij}\{\frac{\frac{\partial g'(z_j,\theta_0)}{\partial\theta}}{1+\lambda_i'(\theta_0)g(z_j,\theta_0)} - \frac{\partial g'(z_j,\theta_0)}{\partial\theta}\})\,\hat{V}^{-1}(x_i,\theta_0)(\sum_{j=1}^{n}w_{ij}g(z_j,\theta_0)).$$

Observe that $\|\Delta\|$ is majorized by

$$\sum_{i=1}^{n}\mathbb{T}_{i,n}\sum_{j=1}^{n}w_{ij}\|\frac{\frac{\partial g(z_j,\theta_0)}{\partial\theta}}{1+\lambda_i'(\theta_0)g(z_j,\theta_0)} - \frac{\partial g(z_j,\theta_0)}{\partial\theta}\|\max_{1\le i\le n}\mathbb{T}_{i,n}\|\hat{V}^{-1}(x_i,\theta_0)\|\max_{1\le i\le n}\mathbb{T}_{i,n}\|\sum_{j=1}^{n}w_{ij}g(z_j,\theta_0)\|.$$

Since $\sup_{x_i\in\mathbb{R}^s}\|V^{-1}(x_i,\theta_0)\|<\infty$ by Assumption 3.5(ii), $\max_{1\le i\le n}\|\hat{V}^{-1}(x_i,\theta_0)\| = O_p(1)$ follows by Lemma B.7. Hence by (A.7) and Assumption 3.5(iii)

$$\begin{aligned}\|\frac{\Delta}{\sqrt{n}}\| &= O_p(1)\max_{1\le i\le n}\mathbb{T}_{i,n}\|\sum_{j=1}^{n}w_{ij}g(z_j,\theta_0)\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbb{T}_{i,n}\sum_{j=1}^{n}w_{ij}\|\frac{\frac{\partial g(z_j,\theta_0)}{\partial\theta}}{1+\lambda_i'(\theta_0)g(z_j,\theta_0)} - \frac{\partial g(z_j,\theta_0)}{\partial\theta}\| \\ &= O_p(1)\max_{1\le i\le n}\mathbb{T}_{i,n}\|\sum_{j=1}^{n}w_{ij}g(z_j,\theta_0)\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbb{T}_{i,n}\|\lambda_i(\theta_0)\|\sum_{j=1}^{n}w_{ij}d(z_j)g_*(z_j) \\ &= O_p(\sqrt{n})\max_{1\le i\le n}\mathbb{T}_{i,n}\|\sum_{j=1}^{n}w_{ij}g(z_j,\theta_0)\|(\frac{1}{n}\sum_{i=1}^{n}\mathbb{T}_{i,n}\|\lambda_i(\theta_0)\|^2)^{1/2}(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}d^2(z_j)g_*^2(z_j)w_{ij})^{1/2},\end{aligned}$$

where the last equality follows by Cauchy-Schwarz and Jensen. Since $\eta\ge 4$, by Lemma D.6 it follows that $\sum_{i=1}^{n}\sum_{j=1}^{n}d^2(z_j)g_*^2(z_j)w_{ij} = O_p(n)$. Hence by Lemma B.3 and (B.4)

$$\|\frac{\Delta}{\sqrt{n}}\| = o_p(\sqrt{\frac{n^{2\beta}}{nb_n^{2s+4\tau}}}) + o_p(\frac{1}{n^{2\varrho-2/m-1/2}}) = o_p(1),$$

which implies that $n^{-1/2}\hat{A} = n^{-1/2}A + o_p(1)$. Thus (A.8) becomes

$$-n^{-1/2}\nabla_\theta\mathrm{SEL}(\theta_0) = n^{-1/2}A + o_p(1). \tag{A.10}$$

By (A.10), Lemma C.1, and the continuity of $\theta \mapsto I(\theta)$ on $\mathcal{B}_0$, (A.6) implies that

$$\begin{aligned}0 &= -n^{-1/2}A + o_p(1) + \{I(\theta_0) + o_p(1)\}n^{1/2}(\hat\theta - \theta_0)\\ &= -n^{-1/2}A + I(\theta_0)n^{1/2}(\hat\theta - \theta_0) + o_p(n^{1/2}\|\hat\theta - \theta_0\|) + o_p(1).\end{aligned}$$

Therefore,

$$n^{1/2}(\hat\theta - \theta_0) = -I^{-1}(\theta_0)\, n^{-1/2}A + o_p(1). \tag{A.11}$$

Since $n^{-1/2}A \xrightarrow{d} \mathrm{N}(0, I(\theta_0))$ by Lemma B.2, the desired result follows. □

Proof of Theorem 4.1. The basic idea behind this proof is outlined in Amemiya (1985, Section 4.5.1). Since $\frac{\partial R(\theta_0)}{\partial\theta}$ has rank r, it must contain a nonsingular $r \times r$ submatrix. Relabeling if necessary, we can assume without loss of generality that $[\, \frac{\partial R(\theta_0)}{\partial\theta^{(p-r+1)}} \cdots \frac{\partial R(\theta_0)}{\partial\theta^{(p)}}\,]_{r\times r}$ is the aforementioned submatrix. Define $\alpha = (\theta^{(1)}, \ldots, \theta^{(p-r)})$ and $\alpha_0 = (\theta_0^{(1)}, \ldots, \theta_0^{(p-r)})$. By the implicit function theorem, there exists a neighborhood $\mathcal{N}$ of θ_0, an open set $\mathcal{U} \subseteq \mathbb{R}^{p-r}$ containing α_0, and a twice continuously differentiable function $\phi : \mathcal{U} \to \mathbb{R}^r$, such that $\{\theta \in \mathcal{N} : R(\theta) = 0\} = \{(\alpha, \phi(\alpha)) : \alpha \in \mathcal{U}\}$. Hence if we let $\tilde R(\alpha) = [\,{\alpha \atop \phi(\alpha)}\,]$, then any $\theta \in \mathcal{N}$ can be expressed as $\theta = \tilde R(\alpha)$ for some $\alpha \in \mathcal{U}$. In particular, $\theta_0 = \tilde R(\alpha_0)$. Note that $\tilde R$ is twice continuously differentiable function from $\mathcal{U} \to \mathbb{R}^p$, and $\mathbb{D}_\alpha \tilde R(\alpha_0)$ has rank $p - r$. Letting

$$\hat\alpha = \operatorname*{argmax}_{\alpha\in\mathcal{U}} \mathrm{SEL}(\tilde R(\alpha)), \tag{A.12}$$

it follows that $\hat\theta^R = \tilde R(\hat\alpha)$. Because (A.12) is unconstrained, it can be handled in the same manner as (2.6). In particular, since

$$n^{1/2}(\hat\theta - \theta_0) = -I^{-1}(\theta_0)\frac{1}{\sqrt{n}}\sum_{t=1}^{n} v_*(x_t, \theta_0)g(z_t, \theta_0) + o_p(1) \tag{A.13}$$

follows from (A.11), (B.5), and (B.6), we can also show that

$$\begin{aligned}n^{1/2}(\hat\alpha - \alpha_0) = -[\mathbb{E}\{D'_\alpha(x, \tilde R(\alpha_0))\, V^{-1}(x, \tilde R(\alpha_0))\, D_\alpha(x, \tilde R(\alpha_0))\}]^{-1}&\\ \times \frac{1}{\sqrt{n}}\sum_{t=1}^{n} D'_\alpha(x_t, \tilde R(\alpha_0))\, V^{-1}(x_t, \tilde R(\alpha_0))\, g(z_t, \tilde R(\alpha_0)) + o_p(1),&\end{aligned}$$

where $D_\alpha(x, \tilde R(\alpha_0)) = \mathbb{E}\{\frac{\partial g(z,\tilde R(\alpha_0))}{\partial\alpha}|x\} = D(x, \theta_0)\,\frac{\partial\tilde R(\alpha_0)}{\partial\alpha}$; i.e.

$$n^{1/2}(\hat\alpha - \alpha_0) = -[\frac{\partial\tilde R'(\alpha_0)}{\partial\alpha}I(\theta_0)\frac{\partial\tilde R(\alpha_0)}{\partial\alpha}]^{-1}\frac{\partial\tilde R'(\alpha_0)}{\partial\alpha}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} v_*(x_t, \theta_0)g(z_t, \theta_0) + o_p(1). \tag{A.14}$$

By a Taylor expansion, $\mathrm{SEL}(\hat\theta) - \mathrm{SEL}(\theta_0) = -\frac{1}{2}(\hat\theta - \theta_0)'\nabla_{\theta\theta}\mathrm{SEL}(\theta^*)(\hat\theta - \theta_0)$ holds for some θ^* between $\hat\theta$ and θ_0. Similarly, using $\nabla_\alpha \mathrm{SEL}(\tilde R(\hat\alpha)) = 0$, $\mathrm{SEL}(\theta_0) - \mathrm{SEL}(\tilde R(\hat\alpha)) = \frac{1}{2}(\hat\alpha - \alpha_0)'\nabla_{\alpha\alpha}\mathrm{SEL}(\tilde R(\alpha^*))(\hat\alpha - \alpha_0)$ holds for some α^* between α_0 and $\hat\alpha$. Thus we get that

$$\begin{aligned}\mathrm{LR}_n = n^{1/2}(\hat\theta - \theta_0)'\{-n^{-1}\nabla_{\theta\theta}\mathrm{SEL}(\theta^*)\}n^{1/2}(\hat\theta - \theta_0)&\\ - n^{1/2}(\hat\alpha - \alpha_0)'\{-n^{-1}\nabla_{\alpha\alpha}\mathrm{SEL}(\tilde R(\alpha^*))\}n^{1/2}(\hat\alpha - \alpha_0).&\end{aligned} \tag{A.15}$$

Henceforth, let $\epsilon_n \stackrel{def}{=} \frac{1}{\sqrt{n}}\sum_{t=1}^{n} v_*(x_t,\theta_0)g(z_t,\theta_0)$. Using (A.13) and Lemma C.1, it is easy to see that

$$n^{1/2}(\hat{\theta}-\theta_0)'\{-n^{-1}\nabla_{\theta\theta}\text{SEL}(\theta^*)\}n^{1/2}(\hat{\theta}-\theta_0) = \epsilon_n' I^{-1}(\theta_0)\epsilon_n + o_p(1). \tag{A.16}$$

A little algebra reveals that

$$\nabla_{\alpha\alpha}\text{SEL}(\tilde{R}(\alpha)) = \frac{\partial \tilde{R}'(\alpha)}{\partial\alpha}\nabla_{\theta\theta}\text{SEL}(\tilde{R}(\alpha))\frac{\partial \tilde{R}(\alpha)}{\partial\alpha} + \sum_{k=1}^{p}\nabla_{\alpha\alpha}\tilde{R}^{(k)}(\alpha)\frac{\partial \text{SEL}(\tilde{R}(\alpha))}{\partial\theta^{(k)}}.$$

Thus we can use Lemmas C.1, C.6, and the twice continuously differentiability of $\tilde{R}$, to show that

$$-n^{-1}\nabla_{\alpha\alpha}\text{SEL}(\tilde{R}(\alpha^*)) = \frac{\partial \tilde{R}'(\alpha_0)}{\partial\alpha}I(\theta_0)\frac{\partial \tilde{R}(\alpha_0)}{\partial\alpha} + o_p(1).$$

Hence by (A.14),

$$\begin{aligned} \text{(A.17)}\quad & n^{1/2}(\hat{\alpha}-\alpha_0)'\{-n^{-1}\nabla_{\alpha\alpha}\text{SEL}(\tilde{R}(\alpha^*))\}n^{1/2}(\hat{\alpha}-\alpha_0) \\ &= \epsilon_n'\frac{\partial \tilde{R}(\alpha_0)}{\partial\alpha}[\frac{\partial \tilde{R}'(\alpha_0)}{\partial\alpha}I(\theta_0)\frac{\partial \tilde{R}(\alpha_0)}{\partial\alpha}]^{-1}\frac{\partial \tilde{R}'(\alpha_0)}{\partial\alpha}\epsilon_n + o_p(1). \end{aligned}$$

Using (A.16) and (A.17), (A.15) reduces to $\text{LR}_n = [I^{-1/2}(\theta_0)\epsilon_n]'\, M\, [I^{-1/2}(\theta_0)\epsilon_n] + o_p(1)$, where

$$M = I_{p\times p} - I^{1/2}(\theta_0)\frac{\partial \tilde{R}(\alpha_0)}{\partial\alpha}[\frac{\partial \tilde{R}'(\alpha_0)}{\partial\alpha}I(\theta_0)\frac{\partial \tilde{R}(\alpha_0)}{\partial\alpha}]^{-1}\frac{\partial \tilde{R}'(\alpha_0)}{\partial\alpha}I^{1/2}(\theta_0).$$

M is a symmetric idempotent matrix of rank r, and $I^{-1/2}(\theta_0)\epsilon_n \xrightarrow{d} \text{N}(0_{p\times 1}, I_{p\times p})$ by the CLT. Therefore, $\text{LR}_n \xrightarrow{d} \chi_r^2$ by the continuous mapping theorem. □

Appendix B. Auxiliary Results for Estimation

Lemma B.1. *Let Assumptions 3.2–3.5 hold. For some $\beta \in (0,1)$ and $b_n \downarrow 0$ let $n^{1-\beta-2/m}b_n^{s+2\tau} \uparrow \infty$, $n^{\varrho-2/m} \uparrow \infty$, and $n^{1-\beta}b_n^{(\frac{m+4}{m-4})\frac{s}{2}} \uparrow \infty$. Then $\mathbb{T}_{i,n}\lambda_i(\theta_0) = \mathbb{T}_{i,n}\hat{V}^{-1}(x_i,\theta_0)\sum_{j=1}^{n} w_{ij}g(z_j,\theta_0) + \mathbb{T}_{i,n}r_i$, where $\max_{1\le i\le n}\mathbb{T}_{i,n}\|r_i\| = o_p(\frac{n^{\beta+1/m}}{nb_n^{s+2\tau}}) + o_p(\frac{1}{n^{2\varrho-3/m}})$.*

Proof of Lemma B.1. Since $\lambda_i(\theta_0)$ solves (2.5),

$$\begin{aligned} 0 &= \sum_{j=1}^{n}\frac{w_{ij}g(z_j,\theta_0)}{1+\lambda_i'(\theta_0)g(z_j,\theta_0)} = \sum_{j=1}^{n}\frac{w_{ij}g(z_j,\theta_0)}{1+\lambda_i'(\theta_0)g(z_j,\theta_0)} \\ &= \sum_{j=1}^{n} w_{ij}g(z_j,\theta_0)\{1-\lambda_i'(\theta_0)g(z_j,\theta_0)+\frac{(\lambda_i'(\theta_0)g(z_j,\theta_0))^2}{1+\lambda_i'(\theta_0)g(z_j,\theta_0)}\} \\ &= \sum_{j=1}^{n} w_{ij}g(z_j,\theta_0) - \hat{V}(x_i,\theta_0)\lambda_i(\theta_0) + \sum_{j=1}^{n}\frac{w_{ij}g(z_j,\theta_0)(\lambda_i'(\theta_0)g(z_j,\theta_0))^2}{1+\lambda_i'(\theta_0)g(z_j,\theta_0)}. \end{aligned}$$

By Lemma B.6, $\max_{1\le i\le n}\mathbb{T}_{i,n}\|\hat{V}(x_i,\theta_0)-V(x_i,\theta_0)\| = o_p(1)$. Since $\inf_{x_i\in\mathbb{R}^s,\alpha\in\mathbb{S}^q}\alpha' V(x_i,\theta_0)\alpha > 0$ by Assumption 3.5(ii), $\inf_{x_i\in\mathbb{R}^s,\alpha\in\mathbb{S}^q}\alpha'\hat{V}(x_i,\theta_0)\alpha$ is also bounded away from zero w.p.a.1. Thus

$\mathbb{T}_{i,n}\hat{V}(x_i,\theta_0)$ is invertible w.p.a.1. Consequently,

$$\mathbb{T}_{i,n}\lambda_i(\theta_0) = \mathbb{T}_{i,n}\hat{V}^{-1}(x_i,\theta_0)\sum_{j=1}^{n} w_{ij}g(z_j,\theta_0) + \mathbb{T}_{i,n}\hat{V}^{-1}(x_i,\theta_0)r_{1,i}, \tag{B.1}$$

where $r_{1,i} = \sum_{j=1}^{n} \frac{w_{ij}g(z_j,\theta_0)(\lambda_i'(\theta_0)g(z_j,\theta_0))^2}{1+\lambda_i'(\theta_0)g(z_j,\theta_0)}$. (2.5) also shows that

$$\mathbb{T}_{i,n}\sum_{j=1}^{n}\frac{w_{ij}(\lambda_i'(\theta_0)g(z_j,\theta_0))^2}{1+\lambda_i'(\theta_0)g(z_j,\theta_0)} = \mathbb{T}_{i,n}\sum_{j=1}^{n} w_{ij}\lambda_i'(\theta_0)g(z_j,\theta_0). \tag{B.2}$$

Hence, as $1+\lambda_i'(\theta_0)g(z_j,\theta_0) \geq 0$ (because $\hat{p}_{ij} \geq 0$),

$$\mathbb{T}_{i,n}\|r_{1,i}\| \leq \max_{1\leq j\leq n}\|g(z_j,\theta_0)\mathbb{T}_{i,n}\|\sum_{j=1}^{n} w_{ij}\lambda_i'(\theta_0)g(z_j,\theta_0) = o(n^{1/m})\mathbb{T}_{i,n}\sum_{j=1}^{n} w_{ij}\lambda_i'(\theta_0)g(z_j,\theta_0),$$

where the equality follows from Lemma D.2, and the $o(n^{1/m})$ term does not depend upon i, j, or $\theta \in \Theta$. Thus by Lemma B.3

$$\mathbb{T}_{i,n}\|r_{1,i}\| = \mathbb{T}_{i,n}\|\lambda_i(\theta_0)\|o(n^{1/m})\{o_p(\sqrt{\frac{n^\beta}{nb_n^{s+2\tau}}}) + o_p(\frac{1}{n^{\varrho-1/m}})\}, \tag{B.3}$$

where the o_p terms do not depend upon i. Next, let $\lambda_i(\theta_0) = \rho_i\xi_i$, where $\rho_i \geq 0$ and $\xi_i \in \mathbb{S}^q$. Since

$$0 \leq 1+\lambda_i'(\theta_0)g(z_j,\theta_0) \leq 1+\rho_i\|g(z_j,\theta_0)\| \overset{\text{Lemma D.2}}{=} 1+\rho_i o(n^{1/m}),$$

(B.2) becomes $\frac{\mathbb{T}_{i,n}\rho_i}{1+\rho_i o(n^{1/m})} \leq \frac{\mathbb{T}_{i,n}\sum_{j=1}^{n} w_{ij}\xi_i' g(z_j,\theta_0)}{\xi_i'\hat{V}(x_i,\theta_0)\xi_i}$ Using Lemma B.6 and the fact that $\xi_i' V(x_i,\theta_0)\xi_i$ is bounded away from zero on $(x_i,\xi_i) \in \mathbb{R}^s \times \mathbb{S}^q$, it follows that

$$\max_{1\leq i\leq n}\frac{\mathbb{T}_{i,n}\rho_i}{1+\rho_i o(n^{1/m})} = o_p(\sqrt{\frac{n^\beta}{nb_n^{s+2\tau}}}) + o_p(\frac{1}{n^{\varrho-1/m}}).$$

But as $\sqrt{\frac{n^{\beta+2/m}}{nb_n^{s+2\tau}}} \downarrow 0$ and $\frac{1}{n^{\varrho-2/m}} \downarrow 0$ under our assumptions, we can solve for ρ_i to obtain

$$\max_{1\leq i\leq n}\mathbb{T}_{i,n}\rho_i = o_p(\sqrt{\frac{n^\beta}{nb_n^{s+2\tau}}}) + o_p(\frac{1}{n^{\varrho-1/m}}). \tag{B.4}$$

Therefore, by (B.3), $\max_{1\leq i\leq n}\mathbb{T}_{i,n}\|r_{1,i}\| = o_p(\frac{n^{\beta+1/m}}{nb_n^{s+2\tau}}) + o_p(\frac{1}{n^{2\varrho-3/m}})$. Since $\max_{1\leq i\leq n}\mathbb{T}_{i,n}\|\hat{V}^{-1}(x_i,\theta_0)\| = O_p(1)$ by Lemma B.7, (B.1) can be written as $\mathbb{T}_{i,n}\lambda_i(\theta_0) = \mathbb{T}_{i,n}\hat{V}^{-1}(x_i,\theta_0)\sum_{j=1}^{n} w_{ij}g(z_j,\theta_0) + \mathbb{T}_{i,n}r_{2,i}$, where $\max_{1\leq i\leq n}\mathbb{T}_{i,n}\|r_{2,i}\| = o_p(\frac{n^{\beta+1/m}}{nb_n^{s+2\tau}}) + o_p(\frac{1}{n^{2\varrho-3/m}})$. The desired result follows. □

Lemma B.2. *Let Assumptions 3.2–3.5 hold. Furthermore, for some $\beta \in (0,1)$ and $b_n \downarrow 0$ assume that* $\max\{\frac{n^\beta}{nb_n^{3s/2+2\tau}}, \frac{b_n^2}{b_n^\tau}, \frac{1}{n^{\varrho-1/\eta}b_n^\tau}, \frac{1}{n^{\varrho-2/m}b_n^\tau}, n\exp(-\frac{nb_n^{2(s+\tau)}}{8\mathcal{K}_{\max}^2})\} \downarrow 0$. *Then $n^{-1/2}A \xrightarrow{d} \mathrm{N}(0, I(\theta_0))$, where A is defined in* (A.9).

Proof of Lemma B.2. Since A is a $p\times 1$ vector, we use the Cramér-Wold device to prove asymptotic normality. Let $\zeta \in \mathbb{S}^p$ be arbitrary. Then

$$\begin{aligned}\zeta' A &= \sum_{i=1}^n \mathbb{T}_{i,n}\{\sum_{j=1}^n w_{ij}\zeta'\frac{\partial g'(z_j,\theta_0)}{\partial\theta}\}\hat{V}^{-1}(x_i,\theta_0)\{\sum_{j=1}^n w_{ij} g(z_j,\theta_0)\}\\ &= \sum_{i=1}^n\sum_{j=1}^n\sum_{t=1}^n \mathbb{T}_{i,n}\, w_{ij}\zeta'\frac{\partial g'(z_j,\theta_0)}{\partial\theta}\hat{V}^{-1}(x_i,\theta_0)w_{it}g(z_t,\theta_0)\\ &= \sum_{t=1}^n \zeta'\{\sum_{i=1}^n \mathbb{T}_{i,n}(\sum_{j=1}^n w_{ij}\frac{\partial g'(z_j,\theta_0)}{\partial\theta})\hat{V}^{-1}(x_i,\theta_0)w_{it}\}g(z_t,\theta_0) = \sum_{t=1}^n \zeta'\hat{v}_*(x_t,\theta_0)g(z_t,\theta_0),\end{aligned}$$

where the $p\times q$ matrix $\hat{v}_*(x_t,\theta_0) = \sum_{i=1}^n \mathbb{T}_{i,n}\{\sum_{j=1}^n w_{ij}\frac{\partial g'(z_j,\theta_0)}{\partial\theta}\}\hat{V}^{-1}(x_i,\theta_0)w_{it}$, and the third equality follows upon changing the order of summation. Recall that $v_*(x_t,\theta_0)$ denotes the matrix of optimal instruments. It is apparent that $\hat{v}_*(x_t,\theta_0)$ estimates $v_*(x_t,\theta_0)$. Now write

$$\zeta' A = \sum_{t=1}^n \zeta' v_*(x_t,\theta_0)g(z_t,\theta_0) + \sum_{t=1}^n\{\zeta'\hat{v}_*(x_t,\theta_0) - \zeta' v_*(x_t,\theta_0)\}g(z_t,\theta_0). \tag{B.5}$$

By the CLT, $n^{-1/2}\sum_{t=1}^n \zeta' v_*(x_t,\theta_0)g(z_t,\theta_0) \xrightarrow{d} \mathrm{N}(0,\zeta' I(\theta_0)\zeta)$. Hence we are done if we can show that

$$n^{-1/2}\sum_{t=1}^n\{\zeta'\hat{v}_*(x_t,\theta_0) - \zeta' v_*(x_t,\theta_0)\}g(z_t,\theta_0) = o_p(1). \tag{B.6}$$

To show (B.6), we proceed as follows. First, let $\tau_{n,1} \stackrel{def}{=} \max\{\sqrt{\frac{n^\beta}{nb_n^{3s/2+2\tau}}}, \frac{b_n^2}{b_n^\tau}, \frac{1}{n^{\varrho-1/\eta}b_n^\tau}, \frac{1}{n^{\varrho-2/m}b_n^\tau}\}$. By Lemma B.5, $\mathbb{T}_{i,n}\sum_{j=1}^n w_{ij}\zeta'\frac{\partial g'(z_j,\theta_0)}{\partial\theta} = \mathbb{T}_{i,n}\zeta' D'(x_i,\theta_0)\frac{h(x_i)}{\hat{h}(x_i)} + r'_{a,i}$, where $r_{a,i}$ is a $q\times 1$ vector such that $\max_{1\le i\le n}\|r_{a,i}\| = O_p(\tau_{n,1})$. By Lemma B.7, $\mathbb{T}_{i,n}\hat{V}^{-1}(x_i,\theta_0) = \mathbb{T}_{i,n}V^{-1}(x_i,\theta_0) + R_{b,i}$ where $R_{b,i}$ is a $q\times q$ matrix such that $\max_{1\le i\le n}\|R_{b,i}\| = O_p(\tau_{n,1})$. Hence

$$\begin{aligned}&\zeta'\hat{v}_*(x_t,\theta_0) = \sum_{i=1}^n\{\mathbb{T}_{i,n}\zeta' D'(x_i,\theta_0)\frac{h(x_i)}{\hat{h}(x_i)} + r'_{a,i}\}\times\{\mathbb{T}_{i,n}V^{-1}(x_i,\theta_0) + R_{b,i}\}w_{it}\\ &= \sum_{i=1}^n \hat{\mathbb{T}}_{i,n}\zeta' v_*(x_i,\theta_0)w_{it} + \sum_{i=1}^n \hat{\mathbb{T}}_{i,n}\zeta' D'(x_i,\theta_0)R_{b,i}w_{it} + \sum_{i=1}^n \mathbb{T}_{i,n}r'_{a,i}V^{-1}(x_i,\theta_0)w_{it} + \sum_{i=1}^n r'_{a,i}R_{b,i}w_{it}.\end{aligned}$$

Since $\sup_{x_i\in\mathbb{R}^s}\|V^{-1}(x_i,\theta_0)\| < \infty$ and $\sum_{i=1}^n w_{it} = 1$, $\max_{1\le t\le n}\|\sum_{i=1}^n \mathbb{T}_{i,n}r'_{a,i}V^{-1}(x_i,\theta_0)w_{it}\| = O_p(\tau_{n,1})$. Similarly, $\max_{1\le t\le n}\|\sum_{i=1}^n r_{a,i}R_{b,i}w_{it}\| = O_p(\tau_{n,1}^2)$. Therefore, it follows that

$$\max_{1\le t\le n}\|\zeta'\hat{v}_*(x_t,\theta_0) - \sum_{i=1}^n \hat{\mathbb{T}}_{i,n}\zeta' v_*(x_i,\theta_0)w_{it} - \sum_{i=1}^n \hat{\mathbb{T}}_{i,n}\zeta' D'(x_i,\theta_0)R_{b,i}w_{it}\| = O_p(\tau_{n,1}). \tag{B.7}$$

We are now ready to show (B.6). Let

$$\hat{\delta}_a'(x_t) = \zeta'\hat{v}_*(x_t,\theta_0) - \sum_{i=1}^n \hat{\mathbb{T}}_{i,n}\zeta' v_*(x_i,\theta_0)w_{it} - \sum_{i=1}^n \hat{\mathbb{T}}_{i,n}\zeta' D'(x_i,\theta_0)R_{b,i}w_{it},$$

$$\hat{\delta}_b'(x_t) = \sum_{i=1}^n \zeta' v_*(x_i,\theta_0)w_{it} - \zeta' v_*(x_t,\theta_0), \quad \hat{\delta}_c'(x_t) = \sum_{i=1}^n (\hat{\mathbb{T}}_{i,n}-1)\zeta' v_*(x_i,\theta_0)w_{it},$$

$$\hat{\delta}_d'(x_t) = \sum_{i=1}^n \zeta' D'(x_i,\theta_0)R_{b,i}w_{it}, \quad \text{and} \quad \hat{\delta}_e'(x_t) = \sum_{i=1}^n (\hat{\mathbb{T}}_{i,n}-1)\zeta' D'(x_i,\theta_0)R_{b,i}w_{it},$$

and observe that

$$\text{(B.8)} \qquad n^{-1/2}\sum_{t=1}^n \{\zeta'\hat{v}_*(x_t,\theta_0) - \zeta' v_*(x_t,\theta_0)\}g(z_t,\theta_0) = \sum_{k\in\{a,b,c,d,e\}} n^{-1/2}\sum_{t=1}^n \hat{\delta}_k'(x_t)g(z_t,\theta_0).$$

Define $\mathcal{U} = \{\xi \in \mathbb{R}^q \; s.t. \; \|\xi\| \leq 1\}$. Pick $\epsilon > 0$ and let $M_{1,\epsilon}$ denote a positive number that may depend upon ϵ. The appropriate $M_{1,\epsilon}$ will be determined later on. Since (B.7) shows that $\Pr\{\max_{1\leq t\leq n} \|\tau_{n,1}^{-1/2}\hat{\delta}_a(x_t)\| > 1\} = o(1)$, we have

$$\begin{aligned}
&\Pr\{|n^{-1/2}\sum_{t=1}^n \tau_{n,1}^{-1/2}\hat{\delta}_a'(x_t)g(z_t,\theta_0)| > M_{1,\epsilon}\} \\
&\qquad \leq \Pr\{|n^{-1/2}\sum_{t=1}^n \tau_{n,1}^{-1/2}\hat{\delta}_a'(x_t)g(z_t,\theta_0)| > M_{1,\epsilon}, \max_{1\leq t\leq n} \|\tau_n^{-1/2}\hat{\delta}_a(x_t)\| \leq 1\} + o(1) \\
&\qquad \leq \Pr\{\sup_{\xi\in\mathcal{U}} |n^{-1/2}\sum_{t=1}^n \xi' g(z_t,\theta_0)| > M_{1,\epsilon}\} + o(1),
\end{aligned}$$

where the last inequality follows because $\max_{1\leq t\leq n}\|\tau_{n,1}^{-1/2}\hat{\delta}_a(x_t)\| \leq 1$ implies that $\tau_{n,1}^{-1/2}\hat{\delta}_a(x_1), \ldots, \tau_{n,1}^{-1/2}\hat{\delta}_a(x_n)$ are elements of $\mathcal{U}$. In short, letting $a_n = n^{-1/2}\sum_{t=1}^n g(z_t,\theta_0)$, we have shown that $\Pr\{|n^{-1/2}\sum_{t=1}^n \tau_{n,1}^{-1/2}\hat{\delta}_a'(x_t)g(z_t,\theta_0)| > M_{1,\epsilon}\} \leq \Pr\{\sup_{\xi\in\mathcal{U}} |\xi' a_n| > M_{1,\epsilon}\} + o(1)$. From Cauchy-Schwarz, $|\xi' a_n| \leq \|\xi\|\cdot\|a_n\| \leq \|a_n\|$. Therefore, by Chebychev,

$$\text{(B.9)} \qquad \Pr\{|n^{-1/2}\sum_{t=1}^n \tau_{n,1}^{-1/2}\hat{\delta}_a'(x_t)g(z_t,\theta_0)| > M_{1,\epsilon}\} \leq \mathbb{E}\|a_n\|^2/M_{1,\epsilon}^2 + o(1).$$

But since $\mathbb{E}\|g(z,\theta_0)\|^2 < \infty$ and $\mathbb{E}\{g'(z_t,\theta_0)g(z_r,\theta_0)\} = 0$ for $t \neq r$,

$$\mathbb{E}\|a_n\|^2 = \frac{1}{n}\{\sum_{t=1}^n \mathbb{E}g'(z_t,\theta_0)g(z_t,\theta_0) + \sum_{t=1}^n \sum_{r=1,r\neq t}^n \mathbb{E}g'(z_t,\theta_0)g(z_r,\theta_0)\} = O(1).$$

So for large enough $M_{1,\epsilon}$, (B.9) reduces to $\Pr\{|n^{-1/2}\sum_{t=1}^n \tau_{n,1}^{-1/2}\hat{\delta}_a'(x_t)g(z_t,\theta_0)| > M_{1,\epsilon}\} \leq \epsilon + o(1)$. Since ϵ was arbitrary, this means that $n^{-1/2}\sum_{t=1}^n \tau_{n,1}^{-1/2}\hat{\delta}_a'(x_t)g(z_t,\theta_0) = O_p(1)$. Hence as $\tau_{n,1} \downarrow 0$,

$$\text{(B.10)} \qquad n^{-1/2}\sum_{t=1}^n \hat{\delta}_a'(x_t)g(z_t,\theta_0) = o_p(1).$$

Next, we show that $n^{-1/2}\sum_{t=1}^{n}\hat{\delta}_b'(x_t)g(z_t,\theta_0)=o_p(1)$. Since $\mathbb{E}\{\hat{\delta}_b'(x_t)g(z_t,\theta_0)\hat{\delta}_b'(x_r)g(z_r,\theta_0)\}=0$ for $t\neq r$, by Cauchy-Schwarz we have

$$\mathbb{E}\{n^{-1/2}\sum_{t=1}^{n}\hat{\delta}_b'(x_t)g(z_t,\theta_0)\}^2\leq\frac{1}{n}\sum_{t=1}^{n}\{\sqrt{\mathbb{E}\|\hat{\delta}_b(x_t)\|^4}\sqrt{\mathbb{E}\|g(z_t,\theta_0)\|^4}\}. \tag{B.11}$$

Devroye and Wagner (1980, Theorem 1, Page 232) show that if $f:\mathbb{R}^s\to\mathbb{R}$ is a Borel measurable function such that $\mathbb{E}|f(x)|^a<\infty$ for some $a\geq 1$ and the kernel $\mathcal{K}$ satisfies Assumption 3.3, then

$$\mathbb{E}|\sum_{i=1}^{n}f(x_i)w_{ij}-f(x_j)|^a\to 0\quad as\quad n\uparrow\infty. \tag{B.12}$$

By the c_r-inequality, $\|\hat{\delta}_b(x_t)\|^4\leq q\sum_{l=1}^{q}\{\sum_{i=1}^{n}[\zeta'v_*(x_i,\theta_0)]^{(l)}w_{it}-[\zeta'v_*(x_t,\theta_0)]^{(l)}\}^4$. Furthermore, $\mathbb{E}\{[\zeta'v_*(x_t,\theta_0)]^{(l)}\}^4\leq\mathbb{E}\|v_*(x_t,\theta_0)\|^4\leq c\,\mathbb{E}\|D(x_t,\theta_0)\|^4<\infty$ because $\sup_{x_i\in\mathbb{R}^s}\|V^{-1}(x_i,\theta_0)\|<\infty$ by Assumption 3.5(ii) and $\mathbb{E}\|D(x_t,\theta_0)\|^\eta<\infty$ (for $\eta\geq 4$) by Assumption 3.5(iii). Thus the conditions of Devroye and Wagner are satisfied, and we can use (B.12) to show $\mathbb{E}\|\hat{\delta}_b(x_t)\|^4=o(1)$. Since $\mathbb{E}\|g(z_t,\theta_0)\|^4<\infty$, (B.11) reduces to $\mathbb{E}\{n^{-1/2}\sum_{t=1}^{n}\hat{\delta}_b'(x_t)g(z_t,\theta_0)\}^2=o(1)$. Hence,

$$n^{-1/2}\sum_{t=1}^{n}\hat{\delta}_b'(x_t)g(z_t,\theta_0)=o_p(1). \tag{B.13}$$

Next, we show that $n^{-1/2}\sum_{t=1}^{n}\hat{\delta}_e'(x_t)g(z_t,\theta_0)=o_p(1)$. By Lemma D.5

$$\max_{1\leq i\leq n}|\hat{\mathbb{T}}_{i,n}-1|=O_p(\tau_{n,2}),\quad \tau_{n,2}\stackrel{def}{=}\max\{\sqrt{\frac{n^\beta}{nb_n^{s+2\tau}}},\frac{b_n^2}{b_n^\tau},\frac{1}{n^\varrho b_n^\tau},n\exp(-\frac{nb_n^{2(s+\tau)}}{8\mathcal{K}_{\max}^2})\}.$$

Hence $\max_{1\leq i\leq n}\|(\hat{\mathbb{T}}_{i,n}-1)R_{b,i}\|=O_p(\tau_{n,1}\tau_{n,2})$. As $\tau_{n,3}\stackrel{def}{=}\tau_{n,1}\tau_{n,2}\downarrow 0$ under our assumptions, it follows that $\Pr\{\max_{1\leq i\leq n}\|\tau_{n,3}^{-1/2}(\hat{\mathbb{T}}_{i,n}-1)R_{b,i}\|>1\}=o(1)$. Now define $\mathcal{V}=\{R\in\mathbb{R}^{q\times q}\ s.t.\ \|R\|\leq 1\}$ to be the set of all $q\times q$ matrices with norm bounded by unity. Pick $\epsilon>0$, and let $M_{2,\epsilon}$ denote a positive number that may depend upon ϵ. Since $\max_{1\leq i\leq n}\|\tau_{n,3}^{-1/2}(\hat{\mathbb{T}}_{i,n}-1)R_{b,i}\|\leq 1$ implies that $\tau_{n,1}^{-1/2}(\hat{\mathbb{T}}_{1,n}-1)R_{b,1}\in\mathcal{V}$, ..., $\tau_{n,1}^{-1/2}(\hat{\mathbb{T}}_{n,n}-1)R_{b,n}\in\mathcal{V}$, following the argument for $\hat{\delta}_a$ we have

$$\begin{aligned}&\Pr\{|n^{-1/2}\sum_{t=1}^{n}\tau_{n,3}^{-1/2}\hat{\delta}_e'(x_t)g(z_t,\theta_0)|>M_{2,\epsilon}\}\\&\qquad\leq\Pr\{|n^{-1/2}\sum_{t=1}^{n}\tau_{n,3}^{-1/2}\hat{\delta}_e'(x_t)g(z_t,\theta_0)|>M_{2,\epsilon},\max_{1\leq i\leq n}\|\tau_{n,3}^{-1/2}(\hat{\mathbb{T}}_{i,n}-1)R_{b,i}\|\leq 1\}+o(1)\\&\qquad\leq\Pr\{\sup_{R\in\mathcal{V}}|n^{-1/2}\sum_{t=1}^{n}\sum_{i=1}^{n}\zeta'D'(x_i,\theta_0)Rg(z_t,\theta_0)w_{it}|>M_{2,\epsilon}\}+o(1).\end{aligned}$$

For convenience, let $\hat{d}_t'=\sum_{i=1}^{n}\zeta'D'(x_i,\theta_0)w_{it}$. Since $|tr(AB)|\leq\|A\|\,\|B\|$ and $\|R\|\leq 1$,

$$|\sum_{t=1}^{n}\sum_{i=1}^{n}\zeta'D'(x_i,\theta_0)Rg(z_t,\theta_0)w_{it}|=|tr\,R\sum_{t=1}^{n}\sum_{i=1}^{n}g(z_t,\theta_0)\zeta'D'(x_i,\theta_0)w_{it}|\leq\|\sum_{t=1}^{n}g(z_t,\theta_0)\hat{d}_t'\|.$$

Hence by Chebychev, it follows that

$$\text{(B.14)} \qquad \Pr\{|n^{-1/2}\sum_{t=1}^{n}\tau_{n,3}^{-1/2}\hat{\delta}_e'(x_t)g(z_t,\theta_0)| > M_{2,\epsilon}\} \le \frac{\mathbb{E}\|\sum_{t=1}^{n} g(z_t,\theta_0)\hat{d}_t'\|^2}{nM_{2,\epsilon}^2} + o(1).$$

But as $\|\sum_{t=1}^{n} g(z_t,\theta_0)\hat{d}_t'\|^2 = tr\{\sum_{t=1}^{n} g(z_t,\theta_0)\hat{d}_t'\}\{\sum_{t=1}^{n}\hat{d}_t g'(z_t,\theta_0)\}$, we have

$$\|\sum_{t=1}^{n} g(z_t,\theta_0)\hat{d}_t'\|^2 = \sum_{t=1}^{n}\hat{d}_t'\hat{d}_t\, g'(z_t,\theta_0)g(z_t,\theta_0) + \sum_{t=1}^{n}\sum_{r=1,r\neq t}^{n}\hat{d}_t'\hat{d}_r\, g'(z_t,\theta_0)g(z_r,\theta_0).$$

Now $\mathbb{E}\{\hat{d}_t'\hat{d}_r\, g'(z_t,\theta_0)g(z_r,\theta_0)\} = 0$ when $r \neq t$, because $\mathbb{E}\{g(z_t,\theta_0)|x_t\} = 0$ and the observations are independent. Hence by Cauchy-Schwarz and the fact that $\mathbb{E}\|g(z_t,\theta_0)\|^4 < \infty$,

$$\mathbb{E}\|\sum_{t=1}^{n} g(z_t,\theta_0)\hat{d}_t'\|^2 = \sum_{t=1}^{n}\mathbb{E}\{\|\hat{d}_t\|^2\,\|g(z_t,\theta_0)\|^2\} \le c\sum_{t=1}^{n}\sqrt{\mathbb{E}\|\hat{d}_t\|^4}.$$

Since $\|\hat{d}_t\|^4 \le \sum_{i=1}^{n}\|D(x_i,\theta_0)\|^4 w_{it}$ by Jensen's inequality and $\mathbb{E}\|D(x_i,\theta_0)\|^4 < \infty$ by assumption, Lemma D.6 shows that $\mathbb{E}\|\hat{d}_t\|^4 < \infty$. Therefore, $\mathbb{E}\|\sum_{t=1}^{n} g(z_t,\theta_0)\hat{d}_t'\|^2 = O(n)$, and (B.14) reduces to

$$\Pr\{|n^{-1/2}\sum_{t=1}^{n}\tau_{n,3}^{-1/2}\hat{\delta}_e'(x_t)g(z_t,\theta_0)| > M_{2,\epsilon}\} \le \epsilon + o(1) \quad \text{for large enough } M_{2,\epsilon}.$$

Since ϵ was arbitrary and $\tau_{n,3} \downarrow 0$, this immediately implies that

$$\text{(B.15)} \qquad n^{-1/2}\sum_{t=1}^{n}\hat{\delta}_e'(x_t)g(z_t,\theta_0) = o_p(1).$$

Using the same argument we can also show that

$$\text{(B.16)} \qquad n^{-1/2}\sum_{t=1}^{n}\hat{\delta}_c'(x_t)g(z_t,\theta_0) = o_p(1) \quad \text{and} \quad n^{-1/2}\sum_{t=1}^{n}\hat{\delta}_d'(x_t)g(z_t,\theta_0) = o_p(1).$$

(B.8), (B.10), (B.13), (B.15), and (B.16) show that (B.6) holds. The desired result follows. □

Lemma B.3. *Let Assumptions 3.2–3.4 hold. Assume that* $b_n \downarrow 0$ *and* $n^{1-\beta}b_n^{(\frac{m+2}{m-2})s} \uparrow \infty$ *for some* $\beta \in (0,1)$. *Then* $\max_{1\le i\le n}\mathbb{T}_{i,n}\|\sum_{j=1}^{n} w_{ij}g(z_j,\theta_0)\| = o_p(\sqrt{\frac{n^\beta}{nb_n^{s+2\tau}}}) + o_p(\frac{1}{n^{\varrho-1/m}})$.

Proof of Lemma B.3. Decompose

$$\mathbb{T}_{i,n}\|\sum_{j=1}^{n} w_{ij}g(z_j,\theta_0)\| \le \max_{1\le i\le n}\mathbb{T}_{i,n}\|\sum_{j=1}^{n} w_{ij}g(z_j,\theta_0)\|\,\mathbb{I}_{i,n} + \max_{1\le i\le n}\mathbb{T}_{i,n}\|\sum_{j=1}^{n} w_{ij}g(z_j,\theta_0)\|\max_{1\le i\le n}\mathbb{I}_{i,n}^c.$$

By Lemma D.3 and Lemma D.7, $\max_{1\le i\le n}\mathbb{I}_{i,n}^c = o_p(\frac{1}{n^\varrho})$ and $\sup_{x_i\in\mathbb{R}^s}\|\sum_{j=1}^{n} w_{ij}g(z_j,\theta_0)\| \overset{w.p.1}{=} o(n^{1/m})$ as $n \uparrow \infty$. Therefore, $\max_{1\le i\le n}\mathbb{T}_{i,n}\|\sum_{j=1}^{n} w_{ij}g(z_j,\theta_0)\|\max_{1\le i\le n}\mathbb{I}_{i,n}^c = o_p(\frac{1}{n^{\varrho-1/m}})$. Next, pick any $\epsilon > 0$, $c_n \downarrow 0$, and observe that

$$\Pr\{\max_{1\le i\le n}\mathbb{T}_{i,n}\|\sum_{j=1}^{n} w_{ij}g(z_j,\theta_0)\|\,\mathbb{I}_{i,n} > \epsilon c_n\} \le \Pr\{\sup_{x_i\in S_n}\mathbb{T}_{i,n}\|\sum_{j=1}^{n} w_{ij}g(z_j,\theta_0)\| > \epsilon c_n\}.$$

Using the definition of $\mathbb{T}_{i,n}$, it follows that

$$\Pr\{\sup_{x_i \in S_n} \mathbb{T}_{i,n} \|\sum_{j=1}^n w_{ij} g(z_j,\theta_0)\| > \epsilon c_n\} \le \Pr\{\sup_{x_i\in S_n} \|\frac{1}{nb_n^s}\sum_{j=1}^n \mathcal{K}_{ij} g(z_j,\theta_0)\| > \epsilon c_n b_n^\tau\}.$$

Now let $1 \le l \le q$ and fix $x_i \in S_n$. Define $\varphi(x_i,x_j,z_j) = g^{(l)}(z_j,\theta_0)\mathcal{K}_{ij}/b_n^s$. Under Assumptions 3.2, 3.3, and 3.4, it can be easily verified that[5]:

(a) $b_n^s\,|\varphi(x_i,x_j,z_j)| \le c\|g(z_j,\theta_0)\|$, and $\mathbb{E}\|g(z_j,\theta_0)\|^m < \infty$ for $m > 2$;
(b) $b_n^{s+1}\,\|\frac{\partial\varphi(x_i,x_j,z_j)}{\partial x_i}\| \le c\,\|g(z_i,\theta_0)\|$, and $\mathbb{E}\|g(z_i,\theta_0)\| < \infty$;
(c) $\mathbb{E}\{b_n^{2s}\,\varphi^2(x_i,x_j,z_j)\} \le c\,b_n^s$.

Thus the sufficient conditions in Ai (1997, Lemma B.1, Page 955) are satisfied, and

$$\sup_{x_i\in S_n} |\frac{1}{nb_n^s}\sum_{j=1}^n \mathcal{K}_{ij} g^{(l)}(z_j,\theta_0)| = o_p(\sqrt{\frac{n^\beta}{nb_n^s}})$$

holds if $n^{1-\beta} b_n^{(\frac{m+2}{m-2})s} \uparrow \infty$ for some $\beta \in (0,1)$. Hence $\Pr\{\sup_{x_i\in S_n} \|\frac{1}{nb_n^s}\sum_{j=1}^n \mathcal{K}_{ij} g(z_j,\theta_0)\| > \epsilon c_n b_n^\tau\} \le \epsilon$ if $c_n = \sqrt{\frac{n^\beta}{nb_n^{s+2\tau}}}$. This shows that $\max_{1\le i\le n} \mathbb{T}_{i,n}\,\|\sum_{j=1}^n w_{ij} g(z_j,\theta_0)\|\,\mathbb{I}_{i,n} = o_p(\sqrt{\frac{n^\beta}{nb_n^{s+2\tau}}})$. The desired result follows. □

Lemma B.4. *Let Assumptions 3.3 and 3.4 hold. Assume that* $b_n \downarrow 0$ *and* $n^{1-\beta} b_n^s \uparrow \infty$ *for some* $\beta \in (0,1)$. *Then* $\max_{1\le i\le n} |\hat{h}(x_i) - h(x_i)| = o_p(\sqrt{\frac{n^\beta}{nb_n^s}}) + O(b_n^2) + o_p(\frac{1}{n^\varrho})$.

Proof of Lemma B.4. Observe that

$$\begin{aligned}\max_{1\le i\le n} |\hat{h}(x_i) - h(x_i)| &\le \max_{1\le i\le n} |\hat{h}(x_i) - h(x_i)|\,\mathbb{I}_{i,n} + \max_{1\le i\le n} |\hat{h}(x_i) - h(x_i)|\,\mathbb{I}_{i,n}^c \\ &\le \sup_{x_i\in S_n} |\hat{h}(x_i) - h(x_i)| + \sup_{x_i\in\mathbb{R}^s} |\hat{h}(x_i) - h(x_i)| \max_{1\le i\le n} \mathbb{I}_{i,n}^c.\end{aligned}$$

Fix $x_i \in S_n$ and define $\varphi(x_i,x_j) = \mathcal{K}_{ij}/b_n^s$. Under Assumptions 3.3 and 3.4, it is easily verified that: (a) $b_n^s\,|\varphi(x_i,x_j)| \le c$; (b) $b_n^{s+1}\|\frac{\partial\varphi(x_i,x_j)}{\partial x_i}\| \le c$; and (c) $\mathbb{E}\{b_n^{2s}\,\varphi^2(x_i,x_j)\} \le c\,b_n^s$. Thus the sufficient conditions in Ai (1997, Lemma B.1, Page 955) are satisfied, and $\sup_{x_i\in S_n} |\hat{h}(x_i) - \mathbb{E}\hat{h}(x_i)| = o_p(\sqrt{\frac{n^\beta}{nb_n^s}})$ provided $n^{1-\beta} b_n^s \uparrow \infty$ for some $\beta \in (0,1)$. Since $\sup_{x_i\in\mathbb{R}^s} \|\nabla_{xx} h(x_i)\| < \infty$ by assumption, we also have $\sup_{x_i\in\mathbb{R}^s} |\mathbb{E}\hat{h}(x_i) - h(x_i)| = O(b_n^2)$. Hence $\sup_{x_i\in S_n} |\hat{h}(x_i) - h(x_i)| = o_p(\sqrt{\frac{n^\beta}{nb_n^s}}) + O(b_n^2)$, provided $n^{1-\beta} b_n^s \uparrow \infty$ for some $\beta \in (0,1)$. From Prakasa Rao (1983, Page 185) we know that $\sup_{x_i\in\mathbb{R}^s} |\hat{h}(x_i) - h(x_i)| \xrightarrow{a.s.} 0$ if $\frac{\log n}{nb_n^s} \downarrow 0$, while Lemma D.3 shows $\max_{1\le i\le n} \mathbb{I}_{i,n}^c = o_p(\frac{1}{n^\varrho})$. Therefore, $\sup_{x_i\in\mathbb{R}^s} |\hat{h}(x_i) - h(x_i)| \max_{1\le i\le n} \mathbb{I}_{i,n}^c = o_p(\frac{1}{n^\varrho})$ provided $\frac{\log n}{nb_n^s} \downarrow 0$. The desired result follows. □

Lemma B.5. *Let Assumptions 3.2–3.5 hold. Let* $b_n \downarrow 0$ *and* $n^{1-\beta} b_n^{(\frac{\eta+2}{\eta-2})\frac{s}{2}} \uparrow \infty$ *for some* $\beta \in (0,1)$. *Then* $\max_{1\le i\le n}\sup_{\theta\in\mathcal{B}_0} \mathbb{T}_{i,n}\|\sum_{j=1}^n \frac{\partial g(z_j,\theta)}{\partial\theta} w_{ij} - D(x_i,\theta)\frac{h(x_i)}{\hat{h}(x_i)}\| = o_p(\sqrt{\frac{n^\beta}{nb_n^{3s/2+2\tau}}}) + O(\frac{b_n^2}{b_n^\tau}) + o_p(\frac{1}{n^{\varrho-1/\eta} b_n^\tau})$.

[5](a) and (b) are obvious. To show (c), notice that since $\sup_{x_j\in\mathbb{R}^s} \mathbb{E}\{\|g(z_j,\theta_0)\|^2|x_j\} < \infty$ by Assumption 3.5(ii), we have $\mathbb{E}\{b_n^{2s}\varphi^2(x_i,x_j,z_j)\} \le c\mathbb{E}\{\mathbb{E}[\|g(z_j,\theta_0)\|^2|x_j]\mathcal{K}_{ij}^2\} \le c\mathbb{E}\mathcal{K}_{ij}^2 \le cb_n^s$; i.e., (c) follows.

Proof of Lemma B.5. Observe that $\mathbb{T}_{i,n}\|\sum_{j=1}^n \frac{\partial g(z_j,\theta)}{\partial\theta}w_{ij} - D(x_i,\theta)\frac{h(x_i)}{\hat{h}(x_i)}\| \leq (1)$, where

$$(1) = \max_{1\leq i\leq n}\sup_{\theta\in\mathcal{B}_0}\frac{\mathbb{T}_{i,n}}{b_n^\tau}\|\frac{1}{nb_n^s}\sum_{j=1}^n\frac{\partial g(z_j,\theta)}{\partial\theta}\mathcal{K}_{ij} - D(x_i,\theta)\,h(x_i)\|.$$

Write $(1)\leq(1)_A+(1)_B$, where

$$(1)_A = \max_{1\leq i\leq n}\sup_{\theta\in\mathcal{B}_0}\frac{\mathbb{T}_{i,n}}{b_n^\tau}\|\frac{1}{nb_n^s}\sum_{j=1}^n\frac{\partial g(z_j,\theta)}{\partial\theta}\mathcal{K}_{ij} - D(x_i,\theta)\,h(x_i)\|\mathbb{I}_{i,n},$$

$$(1)_B = \max_{1\leq i\leq n}\sup_{\theta\in\mathcal{B}_0}\frac{\mathbb{T}_{i,n}}{b_n^\tau}\|\frac{1}{nb_n^s}\sum_{j=1}^n\frac{\partial g(z_j,\theta)}{\partial\theta}\mathcal{K}_{ij} - D(x_i,\theta)\,h(x_i)\|\mathbb{I}_{i,n}^c.$$

Let us examine $(1)_B$ first. Define $(1)_{B1} = \|\frac{1}{nb_n^s}\sum_{j=1}^n\frac{\partial g(z_j,\theta)}{\partial\theta}\mathcal{K}_{ij} - D(x_i,\theta)\,h(x_i)\|$, and observe that $(1)_B\leq\max_{1\leq i\leq n}\sup_{\theta\in\mathcal{B}_0}\frac{(1)_{B1}}{b_n^\tau}\cdot\max_{1\leq i\leq n}\mathbb{I}_{i,n}^c$. But since $\sup_{x_i\in\mathbb{R}^s}h(x_i)<\infty$,

$$\max_{1\leq i\leq n}\sup_{\theta\in\mathcal{B}_0}(1)_{B1}\leq c\{\sup_{x_i\in\mathbb{R}^s}\frac{1}{nb_n^s}\sum_{j=1}^n\sup_{\theta\in\mathcal{B}_0}\|\frac{\partial g(z_j,\theta)}{\partial\theta}\|\,\mathcal{K}_{ij} + \max_{1\leq i\leq n}\sup_{\theta\in\mathcal{B}_0}\|D(x_i,\theta)\|\}.$$

By Lemma D.8, $\sup_{x_i\in\mathbb{R}^s}\frac{1}{nb_n^s}\sum_{j=1}^n\sup_{\theta\in\mathcal{B}_0}\|\frac{\partial g(z_j,\theta)}{\partial\theta}\|\,\mathcal{K}_{ij} = o(n^{1/\eta})$ holds w.p.1 as $n\uparrow\infty$. Moreover, since $\mathbb{E}\{\sup_{\theta\in\mathcal{B}_0}\|D(x_i,\theta)\|^\eta\}<\infty$ by Assumption 3.5(iii), as in Lemma D.2 we can show that $\max_{1\leq i\leq n}\sup_{\theta\in\mathcal{B}_0}\|D(x_i,\theta)\| = o(n^{1/\eta})$ holds w.p.1 as $n\uparrow\infty$. Hence $(1)_B = o(\frac{n^{1/\eta}}{b_n^\tau})\max_{1\leq i\leq n}\mathbb{I}_{i,n}^c = o_p(\frac{1}{n^{\varrho-1/\eta}b_n^\tau})$ by Lemma D.3. Next, use the triangle inequality to write $(1)_A\leq\frac{(1)_{A1}+(1)_{A2}}{b_n^\tau}$, where

$$(1)_{A1} = \sup_{(\theta,x_i)\in\mathcal{B}_0\times S_n}\|\frac{1}{nb_n^s}\sum_{j=1}^n\frac{\partial g(z_j,\theta)}{\partial\theta}\mathcal{K}_{ij} - \mathbb{E}\{\frac{1}{nb_n^s}\sum_{j=1}^n\frac{\partial g(z_j,\theta)}{\partial\theta}\mathcal{K}_{ij}\}\|,$$

$$(1)_{A2} = \sup_{(\theta,x_i)\in\mathcal{B}_0\times S_n}\|\mathbb{E}\{\frac{1}{nb_n^s}\sum_{j=1}^n\frac{\partial g(z_j,\theta)}{\partial\theta}\mathcal{K}_{ij}\} - D(x_i,\theta)\,h(x_i)\|.$$

Now under Assumption 3.5(iv), it is straightforward to show that

$$\sup_{(\theta,x_i)\in\mathcal{B}_0\times\mathbb{R}^s}\|\mathbb{E}\{\frac{1}{nb_n^s}\sum_{j=1}^n\frac{\partial g(z_j,\theta)}{\partial\theta}\mathcal{K}_{ij}\} - D(x_i,\theta)\,h(x_i)\| = O(b_n^2).$$

As $S_n\subset\mathbb{R}^s$, this yields $(1)_{A2} = O(b_n^2)$. Let $1\leq l\leq p$, $1\leq r\leq q$, and $\frac{\partial g^{(lr)}(z_j,\theta)}{\partial\theta}$ denote the $(l,r)^{th}$ element of the $q\times p$ Jacobian matrix $\frac{\partial g(z_j,\theta)}{\partial\theta}$. To find the rate at which $(1)_{A1}$ goes to zero in probability, it suffices to determine the rate for

$$\sup_{(\theta,x_i)\in\mathcal{B}_0\times S_n}|\frac{1}{nb_n^s}\sum_{j=1}^n\frac{\partial g^{(lr)}(z_j,\theta)}{\partial\theta}\mathcal{K}_{ij} - \mathbb{E}\{\frac{1}{nb_n^s}\sum_{j=1}^n\frac{\partial g^{(lr)}(z_j,\theta)}{\partial\theta}\mathcal{K}_{ij}\}|.$$

To do so, we use a result of Ai (1997) on the uniform consistency of kernel estimators over compact but expanding sets. Fix $(\theta,x_i)\in\mathcal{B}_0\times S_n$ and define $\varphi(\theta,x_i,x_j,z_j) = \frac{\partial g^{(lr)}(z_j,\theta)}{\partial\theta}\mathcal{K}_{ij}/b_n^s$. Under Assumptions 3.3, 3.4, and 3.5, it can be easily shown that[6]:

[6](a) and (b) are straightforward. (c) follows by Cauchy-Schwarz and the fact that $\eta\geq 4$.

(a) $b_n^s\,|\varphi(\theta,x_i,x_j,z_j)| \le c\,d(z_j)$, and $\mathbb{E}d^\eta(z_j)<\infty$ for $\eta>2$;
(b) $b_n^{s+1}\,\|\frac{\partial\varphi(\theta,x_i,x_j,z_j)}{\partial\binom{\theta}{x_i}}\| \le c\,\{d(z_j)+b_n l(z_j)\}$, and the RHS has finite expectation;
(c) $\mathbb{E}\{b_n^{2s}\,\varphi^2(\theta,x_i,x_j,z_j)\} \le c\,b_n^{s/2}$.

Thus the sufficient conditions in Ai (1997, Lemma B.1, Page 955) are satisfied, and

$$\sup_{(\theta,x_i)\in\mathcal{B}_0\times S_n} |\frac{1}{nb_n^s}\sum_{j=1}^n \frac{\partial g^{(lr)}(z_j,\theta)}{\partial\theta}\,\mathcal{K}_{ij} - \mathbb{E}\{\frac{1}{nb_n^s}\sum_{j=1}^n \frac{\partial g^{(lr)}(z_j,\theta)}{\partial\theta}\,\mathcal{K}_{ij}\}| = o_p(\sqrt{\frac{n^\beta}{nb_n^{3s/2}}})$$

provided $n^{1-\beta}b_n^{(\frac{\eta+2}{\eta-2})\frac{s}{2}}\uparrow\infty$ for some $\beta\in(0,1)$. This implies $(1)_{A1}=o_p(\sqrt{\frac{n^\beta}{nb_n^{3s/2}}})$. Combining the results for $(1)_{A1}$ and $(1)_{A2}$, we have $(1)_A = o_p(\sqrt{\frac{n^\beta}{nb_n^{3s/2+2\tau}}}) + O(\frac{b_n^2}{b_n^\tau})$. Hence using the result for $(1)_B$, $(1) = o_p(\sqrt{\frac{n^\beta}{nb_n^{3s/2+2\tau}}}) + O(\frac{b_n^2}{b_n^\tau}) + o_p(\frac{1}{n^{\varrho-1/\eta}b_n^\tau})$. The desired result follows. □

Lemma B.6. *Let Assumptions 3.2–3.5 hold. If* $b_n\downarrow 0$ *and* $\min\{n^{1-\beta}b_n^{(\frac{m+4}{m-4})\frac{s}{2}}, n^{1-\beta}b_n^s\}\uparrow\infty$ *for some* $\beta\in(0,1)$, *then* $\max_{1\le i\le n}\sup_{\theta\in\mathcal{B}_0}\mathbb{T}_{i,n}\|\hat V(x_i,\theta)-V(x_i,\theta)\| = o_p(\sqrt{\frac{n^\beta}{nb_n^{3s/2+2\tau}}}) + O(\frac{b_n^2}{b_n^\tau}) + o_p(\frac{1}{n^{\varrho-2/m}b_n^\tau})$.

Proof of Lemma B.6. By the triangle inequality

$$\max_{1\le i\le n}\sup_{\theta\in\mathcal{B}_0}\mathbb{T}_{i,n}\|\hat V(x_i,\theta)-V(x_i,\theta)\| \le (I)+(II), \quad \text{where}$$

$(I) = \max_{1\le i\le n}\sup_{\theta\in\mathcal{B}_0}\frac{\mathbb{T}_{i,n}}{b_n^\tau}\|\hat\Omega(x_i,\theta)-\Omega(x_i,\theta)\|$, $(II) = \max_{1\le i\le n}\sup_{\theta\in\mathcal{B}_0}\frac{\mathbb{T}_{i,n}}{b_n^\tau}\|V(x_i,\theta)\|\,|\hat h(x_i)-h(x_i)|$. Write $(I)\le(I)_A+(I)_B$, where

$$(I)_A = \max_{1\le i\le n}\frac{\mathbb{T}_{i,n}}{b_n^\tau}\,\|\hat\Omega(x_i,\theta)-\Omega(x_i,\theta)\|\,\mathbb{I}_{i,n}, \quad (I)_B = \max_{1\le i\le n}\frac{\mathbb{T}_{i,n}}{b_n^\tau}\,\|\hat\Omega(x_i,\theta)-\Omega(x_i,\theta)\|\,\mathbb{I}_{i,n}^c.$$

Because $\sup_{(x_i,\theta)\in\mathbb{R}^s\times\mathcal{B}_0}\|V(x_i,\theta)\|<\infty$ and $\sup_{x_i\in\mathbb{R}^s} h(x_i)<\infty$,

$$\sup_{(x_i,\theta)\in\mathbb{R}^s\times\mathcal{B}_0}\|\hat\Omega(x_i,\theta)-\Omega(x_i,\theta)\| \le \sup_{x_i\in\mathbb{R}^s}\frac{1}{nb_n^s}\sum_{j=1}^n\sup_{\theta\in\Theta}\|g(z_j,\theta)\|^2\,\mathcal{K}_{ij}+c.$$

Since $\mathbb{E}\{\sup_{\theta\in\Theta}\|g(z,\theta)\|^2\}^{m/2}<\infty$, from Lemma D.8 we know that if $\frac{\log n}{nb_n^s}\downarrow 0$, then

$$\sup_{x_i\in\mathbb{R}^s}\frac{1}{nb_n^s}\sum_{j=1}^n\|g(z_j,\theta)\|^2\,\mathcal{K}_{ij} \overset{w.p.1}{=} o(n^{2/m}) \quad \text{as } n\uparrow\infty.$$

Hence using Lemma D.3, it follows that $(I)_B = o_p(\frac{n^{2/m}}{n^\varrho b_n^\tau})$ if $\frac{\log n}{nb_n^s}\downarrow 0$. Next write $(I)_A\le\frac{(I)_{A1}+(I)_{A2}}{b_n^\tau}$, where $(I)_{A1}=\sup_{(x_i,\theta)\in S_n\times\mathcal{B}_0}\|\hat\Omega(x_i,\theta)-\mathbb{E}\hat\Omega(x_i,\theta)\|$, and $(I)_{A2}=\sup_{(x_i,\theta)\in S_n\times\mathcal{B}_0}\|\mathbb{E}\hat\Omega(x_i,\theta)-\Omega(x_i,\theta)\|$. Fix $x_i\in S_n$, and for $1\le l,r\le q$ define $\psi(\theta,x_i,x_j,z_j)=g^{(l)}(z_j,\theta)g^{(r)}(z_j,\theta)\mathcal{K}_{ij}/b_n^s$. Under Assumptions 3.2–3.5, it is easy to verify that:

(a) $b_n^s|\psi(\theta,x_i,x_j,z_j)|\le c\sup_{\theta\in\Theta}\|g(z_j,\theta)\|^2$, $\mathbb{E}\{\sup_{\theta\in\Theta}\|g(z_j,\theta)\|^2\}^{m/2}<\infty$ and $\frac{m}{2}>2$;
(b) $b_n^{s+1}\|\frac{\partial\psi(\theta,x_i,x_j,z_j)}{\partial\binom{\theta}{x_i}}\|\le c\{\sup_{\theta\in\Theta}\|g(z_j,\theta)\|^2+b_n\sup_{\theta\in\Theta}\|g(z_j,\theta)\|\sup_{\theta\in\mathcal{B}_0}\|\frac{\partial g(z_j,\theta)}{\partial\theta}\|\}$, and the right hand side has finite expectation;
(c) $\mathbb{E}\{b_n^{2s}\,\psi^2(\theta,x_i,x_j,z_j)\}<c\,b_n^{s/2}$ if $\mathbb{E}\{\sup_{\theta\in\Theta}\|g(z_j,\theta)\|^8\}<\infty$.

Thus the sufficient conditions in Ai (1997, Lemma B.1, Page 955) are satisfied, and

$$\sup_{(x_i,\theta)\in S_n\times\mathcal{B}_0} |\hat{\Omega}^{(lr)}(x_i,\theta) - \mathbb{E}\{\hat{\Omega}^{(lr)}(x_i,\theta)\}| = o_p(\sqrt{\frac{n^\beta}{nb_n^{3s/2}}})$$

if $n^{1-\beta}b_n^{(\frac{m+4}{m-4})\frac{s}{2}} \uparrow \infty$ for some $\beta \in (0,1)$. Under Assumption 3.5(iii), it is straightforward to show

$$\sup_{(x_i,\theta)\in\mathbb{R}^s\times\mathcal{B}_0} \|\mathbb{E}\hat{\Omega}^{(lr)}(x_i,\theta) - \Omega^{(lr)}(x_i,\theta)\| = O(b_n^2),$$

and it follows that $(I)_A = o_p(\sqrt{\frac{n^\beta}{nb_n^{3s/2+2\tau}}}) + O(\frac{b_n^2}{b_n^\tau})$. Combined with the result for $(I)_B$, we have

$$(I) = o_p(\sqrt{\frac{n^\beta}{nb_n^{3s/2+2\tau}}}) + O(\frac{b_n^2}{b_n^\tau}) + o_p(\frac{n^{2/m}}{n^\varrho b_n^\tau}). \tag{B.17}$$

Finally, since $\sup_{(x_i,\theta)\in\mathbb{R}^s\times\mathcal{B}_0}\|V(x_i,\theta)\| < \infty$ by Assumption 3.5(iv), by Lemma B.4

$$(II) = o_p(\sqrt{\frac{n^\beta}{nb_n^{s+2\tau}}}) + O(\frac{b_n^2}{b_n^\tau}) + o_p(\frac{1}{n^\varrho b_n^\tau}) \tag{B.18}$$

if $\frac{\log n}{nb_n^s} \downarrow 0$ and $n^{1-\beta}b_n^s \uparrow \infty$ for some $\beta \in (0,1)$. The desired result follows by (B.17) and (B.18). □

Lemma B.7. $\max_{1\le i\le n}\mathbb{T}_{i,n}\|\hat{V}^{-1}(x_i,\theta_0)-V^{-1}(x_i,\theta_0)\| = o_p(\sqrt{\frac{n^\beta}{nb_n^{3s/2+2\tau}}})+O(\frac{b_n^2}{b_n^\tau})+o_p(\frac{1}{n^{\varrho-2/m}b_n^\tau})$ *under conditions of Lemma B.6.*

Proof of Lemma B.7. For convenience, let $o_p(\sqrt{\frac{n^\beta}{nb_n^{3s/2+2\tau}}}) + O(\frac{b_n^2}{b_n^\tau}) + o_p(\frac{1}{n^{\varrho-2/m}b_n^\tau}) \overset{def}{=} O_p(a_n)$. By Lemma B.6, $\max_{1\le i\le n}\sup_{\alpha\in\mathbb{S}^q}\mathbb{T}_{i,n}|\alpha'\hat{V}(x_i,\theta_0)\alpha-\alpha'V(x_i,\theta_0)\alpha| = O_p(a_n)$. Also, $(\alpha,x_i)\mapsto\alpha'V(x_i,\theta_0)\alpha$ is bounded away from zero on $\mathbb{S}^q\times\mathbb{R}^s$ by Assumption 3.5(ii). Hence by Lemma D.1,

$$\max_{1\le i\le n}\sup_{\alpha\in\mathbb{S}^q}\mathbb{T}_{i,n}|\frac{1}{\alpha'\hat{V}(x_i,\theta_0)\alpha} - \frac{1}{\alpha'V(x_i,\theta_0)\alpha}| = O_p(a_n).$$

Thus for any $\xi\in\mathbb{S}^q$, $\max_{1\le i\le n}\sup_{\alpha\in\mathbb{S}^q}\mathbb{T}_{i,n}|\frac{(\alpha'\xi)^2}{\alpha'\hat{V}(x_i,\theta_0)\alpha} - \frac{(\alpha'\xi)^2}{\alpha'V(x_i,\theta_0)\alpha}| = O_p(a_n)$. Therefore,

$$\max_{1\le i\le n}\mathbb{T}_{i,n}|\sup_{\alpha\in\mathbb{S}^q}\frac{(\alpha'\xi)^2}{\alpha'\hat{V}(x_i,\theta_0)\alpha} - \sup_{\alpha\in\mathbb{S}^q}\frac{(\alpha'\xi)^2}{\alpha'V(x_i,\theta_0)\alpha}| = O_p(a_n).$$

Since $\hat{V}(x_i,\theta_0)$ is invertible w.p.a.1, $\max_{1\le i\le n}\mathbb{T}_{i,n}|\xi'\hat{V}^{-1}(x_i,\theta_0)\xi - \xi'V^{-1}(x_i,\theta_0)\xi| = O_p(a_n)$. The desired result follows as $\xi\in\mathbb{S}^q$ was arbitrary. □

Lemma B.8. *Let Assumptions 3.2–3.4 hold. Furthermore, for some* $\beta\in(0,1)$ *and* $b_n\downarrow 0$ *assume that* $\max\{\frac{n^\beta}{nb_n^{3s/2+2\tau}}, \frac{b_n^2}{b_n^\tau}, \frac{1}{n^\rho b_n^\tau}\}\downarrow 0$*. Then recalling the notation defined in proof of Theorem 3.1,* $\sup_{\theta\in\Theta_\delta}|\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n\mathbb{T}_{i,n}w_{ij}q_n(x_i,z_j,\theta) - \frac{1}{n}\sum_{i=1}^n\mathbb{T}_{i,n}q_n(x_i,\theta)| = o_p(n^{-1/m})$.

Proof of Lemma B.8. By (A.1) and the fact that $\|u(x_i,\theta)\| \le 1$,

$$n^{1/m} \sup_{\theta\in\Theta_\delta} |\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n \mathbb{T}_{i,n} w_{ij} q_n(x_i,z_j,\theta) - \frac{1}{n}\sum_{i=1}^n \mathbb{T}_{i,n} q_n(x_i,\theta)|$$

$$\le \sup_{\theta\in\Theta_\delta} \frac{1}{n}\sum_{i=1}^n \mathbb{T}_{i,n} \|\sum_{j=1}^n w_{ij} g(z_j,\theta) - \mathbb{E}\{g(z_i,\theta)|x_i\}\|$$

$$+ n^{1/m} \sup_{\theta\in\Theta_\delta} |\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n \mathbb{T}_{i,n} w_{ij} R_n(t) - \frac{1}{n}\sum_{i=1}^n \mathbb{T}_{i,n}\mathbb{E}\{R_n(t)|x_i\}|.$$

Using (A.2) and the fact that $\max_{1\le j\le n}\mathbb{I}\{z_j\notin C_n\} = o(1)$, it is straightforward to show

$$n^{1/m} \sup_{\theta\in\Theta_\delta} |\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n \mathbb{T}_{i,n} w_{ij} R_n(t) - \frac{1}{n}\sum_{i=1}^n \mathbb{T}_{i,n}\mathbb{E}\{R_n(t)|x_i\}| = o_p(1).$$

Letting $(A) = \sup_{\theta\in\Theta_\delta} \frac{1}{n}\sum_{i=1}^n \mathbb{T}_{i,n}\|\sum_{j=1}^n w_{ij} g(z_j,\theta) - \mathbb{E}\{g(z_i,\theta)|x_i\}\|$, it follows that

$$n^{1/m} \sup_{\theta\in\Theta_\delta} |\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n \mathbb{T}_{i,n} w_{ij} q_n(x_i,z_j,\theta) - \frac{1}{n}\sum_{i=1}^n \mathbb{T}_{i,n} q_n(x_i,\theta)| \le (A) + o_p(1). \tag{B.19}$$

By the triangle inequality $(A) \le (A_1) + (A_2)$, where

$$(A_1) = \frac{1}{b_n^\tau}\sup_{\theta\in\Theta_\delta} \frac{1}{n}\sum_{i=1}^n \mathbb{T}_{i,n}\|\frac{1}{nb_n^s}\sum_{j=1}^n \mathcal{K}_{ij} g(z_j,\theta) - \mathbb{E}\{g(z_i,\theta)|x_i\}h(x_i)\|,$$

$$(A_2) = \frac{1}{b_n^\tau}\sup_{\theta\in\Theta_\delta} \frac{1}{n}\sum_{i=1}^n \mathbb{T}_{i,n}\|\mathbb{E}\{g(z_i,\theta)|x_i\}\|\,|\hat{h}(x_i) - h(x_i)|$$

$$\le \frac{1}{b_n^\tau}\max_{1\le i\le n}|\hat{h}(x_i) - h(x_i)|\,\frac{1}{n}\sum_{i=1}^n \mathbb{E}\{\sup_{\theta\in\Theta_\delta}\|g(z_i,\theta)\|\,|x_i\}.$$

But $\frac{1}{n}\sum_{i=1}^n \mathbb{E}\{\sup_{\theta\in\Theta_\delta}\|g(z_i,\theta)\|\,|x_i\} = O_p(1)$. Thus by Lemma B.4,

$$(A_2) = o_p(\sqrt{\frac{n^\beta}{nb_n^{s+2\tau}}}) + O(\frac{b_n^2}{b_n^\tau}) + o_p(\frac{1}{n^\rho b_n^\tau}) = o_p(1)$$

under our conditions. Now to (A_1). By the triangle inequality $(A_1) \le (A_{1a}) + (A_{1b})$, where

$$(A_{1a}) = \frac{1}{b_n^\tau}\sup_{\theta\in\Theta_\delta} \frac{1}{n}\sum_{i=1}^n \mathbb{T}_{i,n}\|\frac{1}{nb_n^s}\sum_{j=1}^n \mathcal{K}_{ij} g(z_j,\theta) - \mathbb{E}\{g(z_i,\theta)|x_i\}h(x_i)\|\mathbb{I}_{i,n},$$

$$(A_{1b}) = \frac{1}{b_n^\tau}\sup_{\theta\in\Theta_\delta} \frac{1}{n}\sum_{i=1}^n \mathbb{T}_{i,n}\|\frac{1}{nb_n^s}\sum_{j=1}^n \mathcal{K}_{ij} g(z_j,\theta) - \mathbb{E}\{g(z_i,\theta)|x_i\}h(x_i)\|\mathbb{I}_{i,n}^c.$$

Using Lemmas D.3 and D.9, it follows that

$$(A_{1b}) \leq \frac{c}{b_n^\tau} \max_{1\leq i\leq n} \mathbb{I}_{i,n}^c \{\frac{1}{n}\sum_{i=1}^n \frac{1}{nb_n^s}\sum_{j=1}^n \mathcal{K}_{ij} \sup_{\theta\in\Theta_\delta} \|g(z_j,\theta)\| + \frac{1}{n}\sum_{i=1}^n \mathbb{E}[\sup_{\theta\in\Theta_\delta} \|g(z_i,\theta)\| \, |x_i]\}$$

$$= o_p(\frac{1}{n^\rho b_n^\tau})\{O_p(1) + O_p(1)\} = o_p(1),$$

because $\frac{1}{n^\rho b_n^\tau} \downarrow 0$ by assumption. To handle (A_{1a}), note that

$$(A_{1a}) \leq \frac{1}{b_n^\tau} \sup_{(\theta,x_i)\in\Theta_\delta\times S_n} \|\frac{1}{nb_n^s}\sum_{j=1}^n \mathcal{K}_{ij} g(z_j,\theta) - \mathbb{E}\{\frac{1}{nb_n^s}\sum_{j=1}^n \mathcal{K}_{ij} g(z_j,\theta)\}\|$$

$$+ \frac{1}{b_n^\tau} \sup_{(\theta,x_i)\in\Theta_\delta\times S_n} \|\mathbb{E}\{\frac{1}{nb_n^s}\sum_{j=1}^n \mathcal{K}_{ij} g(z_j,\theta)\} - \mathbb{E}\{g(z_i,\theta)|x_i\}h(x_i)\|.$$

Under Assumption 3.4(iv), it is straightforward to show that

$$\sup_{(\theta,x_i)\in\Theta\times\mathbb{R}^s} \|\mathbb{E}\{\frac{1}{nb_n^s}\sum_{j=1}^n \mathcal{K}_{ij} g(z_j,\theta)\} - \mathbb{E}\{g(z_i,\theta)|x_i\}h(x_i)\| = O(b_n^2).$$

As $b_n^{2-\tau} \downarrow 0$ by assumption, it follows that

$$(A_{1a}) \leq \frac{1}{b_n^\tau} \sup_{(\theta,x_i)\in\Theta_\delta\times S_n} \|\frac{1}{nb_n^s}\sum_{j=1}^n \mathcal{K}_{ij} g(z_j,\theta) - \mathbb{E}\{\sum_{j=1}^n \mathcal{K}_{ij} g(z_j,\theta)\}\| + o(1).$$

Now fix $(\theta, x_i) \in \Theta_\delta \times S_n$ and define $\psi(\theta, x_i, x_j, z_j) = g^{(l)}(z_j,\theta)\mathcal{K}_{ij}/b_n^s$ for $1 \leq l \leq q$. Under Assumptions 3.2, 3.3, and 3.4, it is straightforward to verify that:

(a) $b_n^s|\psi(\theta, x_i, x_j, z_j)| \leq c \sup_{\theta\in\Theta} \|g(z_j,\theta)\|$, $\mathbb{E}\{\sup_{\theta\in\Theta} \|g(z_j,\theta)\|^m\} < \infty$ and $m > 2$;
(b) $b_n^{s+1}\|\frac{\partial\psi(\theta,x_i,x_j,z_j)}{\partial(\begin{smallmatrix}\theta\\ x_i\end{smallmatrix})}\| \leq c\{\sup_{\theta\in\Theta} \|g(z_j,\theta)\| + b_n \sup_{\theta\in\Theta} \|\frac{\partial g(z_j,\theta)}{\partial\theta}\|\}$, and RHS has finite expectation;
(c) $\mathbb{E}\{b_n^{2s}\,\psi^2(\theta, x_i, x_j, z_j)\} < c\, b_n^{s/2}$.

Therefore, the sufficient conditions in Ai (1997, Lemma B.1, Page 955) are satisfied, and

$$\sup_{(\theta,x_i)\in\Theta_\delta\times S_n} \|\frac{1}{nb_n^s}\sum_{j=1}^n \mathcal{K}_{ij} g(z_j,\theta) - \mathbb{E}\{\frac{1}{nb_n^s}\sum_{j=1}^n \mathcal{K}_{ij} g(z_j,\theta)\}\| = o_p(\sqrt{\frac{n^\beta}{nb_n^{3s/2}}})$$

provided $n^{1-\beta} b_n^{(\frac{m+2}{m-2})\frac{s}{2}} \uparrow \infty$ for some $\beta \in (0, 1)$. But since $\sqrt{\frac{n^\beta}{nb_n^{3s/2+2\tau}}} \downarrow 0$ under our conditions,

$$(A_{1a}) = o_p(\sqrt{\frac{n^\beta}{nb_n^{3s/2+2\tau}}}) + o(1) = o_p(1).$$

Together with the result for (A_{1b}), this implies that $(A_1) = o_p(1)$. Hence $(A) \leq (A_1) + (A_2) = o_p(1)$, and the desired result follows from (B.19). □

Appendix C. Auxiliary Results for Hypothesis Testing

Lemma C.1. *Let Assumptions 3.2–3.7 hold. Then* $\sup_{\theta\in\mathcal{B}_0}\|-\frac{1}{n}\nabla_{\theta\theta}\mathrm{SEL}(\theta)-I(\theta)\|=o_p(1)$.

Proof of Lemma C.1. Observe that

$$\mathrm{SEL}(\theta)=\sum_{i=1}^n\sum_{j=1}^n\mathbb{T}_{i,n}w_{ij}\log\{\frac{w_{ij}}{n}\}-\sum_{i=1}^n\sum_{j=1}^n\mathbb{T}_{i,n}w_{ij}\log\{1+\lambda_i'(\theta)g(z_j,\theta)\},$$

where $\lambda_i(\theta)$ solves (2.5). Since $\sum_{j=1}^n\frac{w_{ij}g(z_j,\theta)}{1+\lambda_i'g(z_j,\theta)}=0$ for all $\theta\in\Theta$,

$$-\nabla_\theta\mathrm{SEL}(\theta)=\sum_{i=1}^n\sum_{j=1}^n\frac{\mathbb{T}_{i,n}w_{ij}[\nabla_\theta g(z_j,\theta)]\lambda_i(\theta)}{1+\lambda_i'(\theta)g(z_j,\theta)}. \tag{C.1}$$

Hence we can write $-\nabla_{\theta\theta}\mathrm{SEL}(\theta)=T_1(\theta)+T_2(\theta)+T_3(\theta)$, where

$$T_1(\theta)=-\sum_{i=1}^n\sum_{j=1}^n\frac{\mathbb{T}_{i,n}w_{ij}[\nabla_\theta\{\lambda_i'(\theta)g(z_j,\theta)\}]\lambda_i'(\theta)\nabla_\theta g'(z_j,\theta)}{[1+\lambda_i'(\theta)g(z_j,\theta)]^2},$$

$$T_2(\theta)=\sum_{i=1}^n\sum_{j=1}^n\frac{\mathbb{T}_{i,n}w_{ij}[\nabla_\theta\lambda_i(\theta)]\nabla_\theta g'(z_j,\theta)}{1+\lambda_i'(\theta)g(z_j,\theta)},\ T_3(\theta)=\sum_{i=1}^n\sum_{j=1}^n\frac{\mathbb{T}_{i,n}w_{ij}}{1+\lambda_i'(\theta)g(z_j,\theta)}\sum_{k=1}^q[\nabla_{\theta\theta}g^{(k)}(z_j,\theta)]\lambda_i^{(k)}(\theta).$$

The desired result follows by Lemma C.2, Lemma C.3, and Lemma C.4. □

Lemma C.2. *Let Assumptions 3.2–3.7 hold. Then* $\sup_{\theta\in\mathcal{B}_0}\|\frac{T_1(\theta)}{n}\|=o_p(1)$.

Proof of Lemma C.2. Since $\nabla_\theta\{\lambda_i'(\theta)g(z_j,\theta)\}=[\nabla_\theta g(z_j,\theta)]\lambda_i(\theta)+[\nabla_\theta\lambda_i(\theta)]g(z_j,\theta)$, we can write $T_1(\theta)=T_{1,a}(\theta)+T_{1,b}(\theta)$, where

$$T_{1,a}(\theta)=-\sum_{i=1}^n\sum_{j=1}^n\frac{\mathbb{T}_{i,n}w_{ij}}{[1+\lambda_i'(\theta)g(z_j,\theta)]^2}[\nabla_\theta g(z_j,\theta)]\lambda_i(\theta)\lambda_i'(\theta)\nabla_\theta g'(z_j,\theta),$$

$$T_{1,b}(\theta)=-\sum_{i=1}^n\sum_{j=1}^n\frac{\mathbb{T}_{i,n}w_{ij}}{[1+\lambda_i'(\theta)g(z_j,\theta)]^2}[\nabla_\theta\lambda_i(\theta)]g(z_j,\theta)\lambda_i'(\theta)\nabla_\theta g'(z_j,\theta).$$

By Assumptions 3.5(ii) and 3.6, $\sup_{\theta\in\mathcal{B}_0}\|T_{1,a}(\theta)\|\le o(1)\sum_{i=1}^n\sum_{j=1}^n w_{ij}d^2(z_j)$, where the $o(1)$ term does not depend upon i, j, or $\theta\in\Theta$. Hence $\sup_{\theta\in\mathcal{B}_0}\|T_{1,a}(\theta)/n\|=o_p(1)$ follows by Lemma D.6. Similarly, by $\sup_{\theta\in\mathcal{B}_0}\|T_{1,b}(\theta)\|\le o(1)\sup_{\theta\in\mathcal{B}_0}\sum_{i=1}^n\mathbb{T}_{i,n}\|\nabla_\theta\lambda_i(\theta)\|\sum_{j=1}^n w_{ij}d(z_j)$ and the Cauchy-Schwarz and Jensen inequalities,

$$\sup_{\theta\in\mathcal{B}_0}\|\frac{T_{1,b}(\theta)}{n}\|=o(1)\{\sup_{\theta\in\mathcal{B}_0}\frac{1}{n}\sum_{i=1}^n\mathbb{T}_{i,n}\|\nabla_\theta\lambda_i(\theta)\|^2\}^{1/2}\{\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n w_{ij}d^2(z_j)\}^{1/2}=o_p(1)$$

from (C.2) and Lemma D.6. The desired result follows. □

Lemma C.3. *Let Assumptions 3.2–3.7 hold. Then* $\sup_{\theta\in\mathcal{B}_0}\|\frac{T_2(\theta)}{n}-I(\theta)\|=o_p(1)$.

Proof of Lemma C.3. By (C.5)

$$\frac{T_2(\theta)}{n}=\frac{1}{n}\sum_{i=1}^n\hat{\mathbb{T}}_{i,n}[\nabla_\theta\lambda_i(\theta)]D(x_i,\theta)+\frac{1}{n}\sum_{i=1}^n\hat{\mathbb{T}}_{i,n}[\nabla_\theta\lambda_i(\theta)]\mathbb{E}\{d(z_i)|x_i\}R_{2,i}(\theta)+\frac{1}{n}\sum_{i=1}^n\mathbb{T}_{i,n}[\nabla_\theta\lambda_i(\theta)]R_{3,i}(\theta),$$

where $\max_{1\le i\le n}\sup_{\theta\in\mathcal{B}_0}\|R_{2,i}(\theta)\| = o_p(1)$ and $\max_{1\le i\le n}\sup_{\theta\in\mathcal{B}_0}\|R_{3,i}(\theta)\| = o_p(1)$. Under our assumptions, $\max_{1\le i\le n}\hat{\mathbb{T}}_{i,n} = O_p(1)$ and $\sup_{(x_i,\theta)\in\mathbb{R}^s\times\mathcal{B}_0}\|V^{-1}(x_i,\theta)\| < \infty$. So by Lemma C.5,

$$\text{(C.2)}\qquad \sup_{\theta\in\mathcal{B}_0}\frac{1}{n}\sum_{i=1}^n \mathbb{T}_{i,n}\|\nabla_\theta\lambda_i(\theta)\|^2 = O_p(1) \quad\text{and}\quad \sup_{\theta\in\mathcal{B}_0}\frac{1}{n}\sum_{i=1}^n \mathbb{T}_{i,n}\|\nabla_\theta\lambda_i(\theta)\| = O_p(1).$$

Use (C.2) and Cauchy-Schwarz to obtain $\sup_{\theta\in\mathcal{B}_0}\|\frac{T_2(\theta)}{n} - \frac{1}{n}\sum_{i=1}^n \hat{\mathbb{T}}_{i,n}[\nabla_\theta\lambda_i(\theta)]D(x_i,\theta)\| = o_p(1)$. Applying the Cauchy-Schwarz and Jensen inequalities to Lemma C.5 once again,

$$\sup_{\theta\in\mathcal{B}_0}\|\frac{1}{n}\sum_{i=1}^n \hat{\mathbb{T}}_{i,n}[\nabla_\theta\lambda_i(\theta)]D(x_i,\theta) - \frac{1}{n}\sum_{i=1}^n \hat{\mathbb{T}}_{i,n}^2 D'(x_i,\theta)V^{-1}(x_i,\theta)D(x_i,\theta)\| = o_p(1).$$

Therefore, as $\max_{1\le i\le n}|\hat{\mathbb{T}}_{i,n}^2 - 1| = o_p(1)$ by Lemma D.5,

$$\sup_{\theta\in\mathcal{B}_0}\|\frac{T_2(\theta)}{n} - \frac{1}{n}\sum_{i=1}^n D'(x_i,\theta)V^{-1}(x_i,\theta)D(x_i,\theta)\| = o_p(1).$$

Since $\sup_{\theta\in\mathcal{B}_0}\|\frac{1}{n}\sum_{i=1}^n D'(x_i,\theta)V^{-1}(x_i,\theta)D(x_i,\theta) - I(\theta)\| = o_p(1)$ by a uniform WLLN[7], the desired result follows. □

Lemma C.4. *Let Assumptions 3.3, 3.5, and 3.6 hold. Then* $\sup_{\theta\in\mathcal{B}_0}\|\frac{T_3(\theta)}{n}\| = o_p(1)$.

Proof of Lemma C.4. By Assumptions 3.5(*iii*) and 3.6, $\sup_{\theta\in\mathcal{B}_0}\|T_3(\theta)\| \le o(1)\sum_{i=1}^n\sum_{j=1}^n w_{ij}l(z_j)$, where the $o(1)$ term does not depend upon i, j, or $\theta\in\Theta$. The desired result follows by Lemma D.6. □

Lemma C.5. *Let Assumptions 3.2–3.7 hold. Then for each i and $\theta\in\mathcal{B}_0$ we can write*

$$\mathbb{T}_{i,n}\nabla_\theta\lambda_i'(\theta) = \hat{\mathbb{T}}_{i,n}V^{-1}(x_i,\theta)D(x_i,\theta) + \hat{\mathbb{T}}_{i,n}M_{1,i}(\theta)D(x_i,\theta) + \hat{\mathbb{T}}_{i,n}\mathbb{E}\{d(z_i)|x_i\}M_{2,i}(\theta) + M_{3,i}(\theta)\sum_{j=1}^n d(z_j)w_{ij} + M_{4,i}(\theta),$$

where $M_{1,i}$ is a $q\times q$ matrix such that $\max_{1\le i\le n}\sup_{\theta\in\mathcal{B}_0}\|M_{1,i}(\theta)\| = o_p(1)$, *and $M_{2,i}$, $M_{3,i}$, $M_{4,i}$ are $q\times p$ matrices such that* $\max_{1\le i\le n}\sup_{\theta\in\mathcal{B}_0}\|M_{k,i}(\theta)\| = o_p(1)$ *for $k = 2,3,4$.*

Proof of Lemma C.5. From (2.5), we know that $\lambda_i(\theta)$ solves $\sum_{j=1}^n \frac{w_{ij}g(z_j,\theta)}{1+\lambda_i'(\theta)g(z_j,\theta)} = 0$ for all $\theta\in\Theta$. Differentiating this identity with respect to θ and rearranging,

$$\text{(C.3)}\qquad \sum_{j=1}^n \frac{w_{ij}g(z_j,\theta)g'(z_j,\theta)}{[1+\lambda_i'(\theta)g(z_j,\theta)]^2}\nabla_\theta\lambda_i'(\theta) = \sum_{j=1}^n \frac{w_{ij}\nabla_\theta g'(z_j,\theta)}{1+\lambda_i'(\theta)g(z_j,\theta)} - \sum_{j=1}^n \frac{w_{ij}g(z_j,\theta)\lambda_i'(\theta)\nabla_\theta g'(z_j,\theta)}{[1+\lambda_i'(\theta)g(z_j,\theta)]^2}.$$

Let us simplify (C.3). First, by Assumption 3.6

$$\|\sum_{j=1}^n \frac{w_{ij}g(z_j,\theta)g'(z_j,\theta)}{[1+\lambda_i'(\theta)g(z_j,\theta)]^2} - V(x_i,\theta)\| \le O(1)\|\hat{V}(x_i,\theta) - V(x_i,\theta)\| + o(1)\|V(x_i,\theta)\|,$$

[7] See, for example, Newey and McFadden (1994, Lemma 2.4, Page 2129).

where the $O(1)$ and $o(1)$ terms do not depend upon i, j, or $\theta \in \Theta$. Since $\sup_{(x_i,\theta)\in\mathbb{R}^s\times\mathcal{B}_0} \|V(x_i,\theta)\| < \infty$ by Assumption 3.5(ii), Lemma B.6 shows that

$$\max_{1\le i\le n}\sup_{\theta\in\mathcal{B}_0} \mathbb{T}_{i,n}\|\sum_{j=1}^n \frac{w_{ij}g(z_j,\theta)g'(z_j,\theta)}{[1+\lambda_i'(\theta)g(z_j,\theta)]^2} - V(x_i,\theta)\| = o_p(1).$$

Therefore, by Assumption 3.5(ii), we can write

$$\mathbb{T}_{i,n}\{\sum_{j=1}^n \frac{w_{ij}g(z_j,\theta)g'(z_j,\theta)}{[1+\lambda_i'(\theta)g(z_j,\theta)]^2}\}^{-1} = \mathbb{T}_{i,n}V^{-1}(x_i,\theta) + R_{1,i}(\theta), \tag{C.4}$$

where $R_{1,i}$ is a $q\times q$ matrix such that $\max_{1\le i\le n}\sup_{\theta\in\mathcal{B}_0}\|R_{1,i}(\theta)\| = o_p(1)$. Next, by Assumption 3.6

$$\|\sum_{j=1}^n \frac{w_{ij}\nabla_\theta g'(z_j,\theta)}{1+\lambda_i'(\theta)g(z_j,\theta)} - D(x_i,\theta)\frac{h(x_i)}{\hat{h}(x_i)}\| \le O(1)\|\sum_{j=1}^n w_{ij}\nabla_\theta g'(z_j,\theta) - D(x_i,\theta)\frac{h(x_i)}{\hat{h}(x_i)}\| \\ + o(1)\|D(x_i,\theta)\|\frac{h(x_i)}{\hat{h}(x_i)},$$

where the $O(1)$ and $o(1)$ terms do not depend upon i, j, or $\theta \in \Theta$. As $\|D(x_i,\theta)\| \le \mathbb{E}\{d(z_i)|x_i\}$ by Assumption 3.5(iii), we have

$$\mathbb{T}_{i,n}\|\sum_{j=1}^n \frac{w_{ij}\nabla_\theta g'(z_j,\theta)}{1+\lambda_i'(\theta)g(z_j,\theta)} - D(x_i,\theta)\frac{h(x_i)}{\hat{h}(x_i)}\| = O(1)\max_{1\le i\le n}\sup_{\theta\in\mathcal{B}_0}\mathbb{T}_{i,n}\|\sum_{j=1}^n w_{ij}\nabla_\theta g'(z_j,\theta) - D(x_i,\theta)\frac{h(x_i)}{\hat{h}(x_i)}\| \\ + o(1)\hat{\mathbb{T}}_{i,n}\mathbb{E}\{d(z_i)|x_i\}.$$

Hence using Lemma B.5, we can write

$$\mathbb{T}_{i,n}\sum_{j=1}^n \frac{w_{ij}\nabla_\theta g'(z_j,\theta)}{1+\lambda_i'(\theta)g(z_j,\theta)} = \hat{\mathbb{T}}_{i,n}D(x_i,\theta) + \hat{\mathbb{T}}_{i,n}\mathbb{E}\{d(z_i)|x_i\}R_{2,i}(\theta) + R_{3,i}(\theta), \tag{C.5}$$

where $R_{2,i}$ and $R_{3,i}$ are $q\times p$ matrices such that we have $\max_{1\le i\le n}\sup_{\theta\in\mathcal{B}_0}\|R_{2,i}(\theta)\| = o_p(1)$ and $\max_{1\le i\le n}\sup_{\theta\in\mathcal{B}_0}\|R_{3,i}(\theta)\| = o_p(1)$. Finally, by Assumption 3.5(iii) and 3.6

$$\|\sum_{j=1}^n \frac{w_{ij}g(z_j,\theta)\lambda_i'(\theta)\nabla_\theta g'(z_j,\theta)}{[1+\lambda_i'(\theta)g(z_j,\theta)]^2}\| \le o(1)\sum_{j=1}^n d(z_j)w_{ij},$$

where the $o(1)$ term does not depend upon i, j, or $\theta \in \Theta$. Hence we can write

$$\sum_{j=1}^n \frac{w_{ij}g(z_j,\theta)\lambda_i'(\theta)\nabla_\theta g'(z_j,\theta)}{[1+\lambda_i'(\theta)g(z_j,\theta)]^2} = R_{4,i}(\theta)\sum_{j=1}^n d(z_j)w_{ij}, \tag{C.6}$$

where $R_{4,i}$ is a $q\times p$ matrix such that $\max_{1\le i\le n}\sup_{\theta\in\mathcal{B}_0}\|R_{4,i}(\theta)\| = o_p(1)$. By (C.4), (C.5), and (C.6), (C.3) can be written as

$$\mathbb{T}_{i,n}\nabla_\theta\lambda_i'(\theta) = \{\mathbb{T}_{i,n}V^{-1}(x_i,\theta) + R_{1,i}(\theta)\}\{\hat{\mathbb{T}}_{i,n}D(x_i,\theta) + \hat{\mathbb{T}}_{i,n}\mathbb{E}[d(z_i)|x_i]R_{2,i}(\theta) + R_{3,i}(\theta) \\ + R_{4,i}(\theta)\sum_{j=1}^n d(z_j)w_{ij}\}.$$

The desired result follows by $\sup_{(x_i,\theta)\in\mathbb{R}^s\times\mathcal{B}_0}\|V(x_i,\theta)\|<\infty$ and the properties of $R_{1,i},\ldots,R_{4,i}$. □

Lemma C.6. *Let Assumptions 3.3, 3.5, and 3.6 hold. Then* $\sup_{\theta\in\mathcal{B}_0}\|\frac{\nabla_\theta \mathrm{SEL}(\theta)}{n}\|=o_p(1)$.

Proof of Lemma C.6. Using (C.1), Assumption 3.6, and Assumption 3.5(*iii*), it is easily seen that $\sup_{\theta\in\mathcal{B}_0}\|\nabla_\theta\mathrm{SEL}(\theta)\|\le o(1)\sum_{i=1}^n\sum_{j=1}^n w_{ij}d(z_j)$, where the $o(1)$ term does not depend upon i, j, or $\theta\in\Theta$. Hence the desired result follows by Lemma D.6. □

Appendix D. Other Useful Results

Lemma D.1. *Let a_n and b_n be sequences of positive numbers such that $a_n,b_n\downarrow 0$. r_n is a sequence of functions such that $\sup_x|r_n(x)-r(x)|=O_p(a_n)$ and $\sup_x|r(x)|<\infty$. s_n is a sequence of functions such that $\sup_x|s_n(x)-s(x)|=O_p(b_n)$ and $\inf_x|s(x)|>0$. Then $\sup_x|\frac{r_n(x)}{s_n(x)}-\frac{r(x)}{s(x)}|=O_p(\max\{a_n,b_n\})$.*

Proof of Lemma D.1. See Tripathi and Kitamura (2000, Lemma C.1). □

Lemma D.2. *If $\mathbb{E}\{\sup_{\theta\in\Theta}\|g(z,\theta)\|^m\}<\infty$, then $\Pr\{\max_{1\le j\le n}\sup_{\theta\in\Theta}\|g(z_j,\theta)\|=o(n^{1/m})\}=1$ as $n\uparrow\infty$.*

Proof of Lemma D.2. Our proof is based on the idea described in Owen (1990, Lemma 3). Let $\epsilon>0$. Since $\sum_{n=1}^{\infty}\Pr\{[\sup_{\theta\in\Theta}\|g(z_1,\theta)\|]^m/\epsilon^m\ge n\}\le\mathbb{E}[\sup_{\theta\in\Theta}\|g(z_1,\theta)\|]^m/\epsilon^m$, it follows that $\sum_{n=1}^{\infty}\Pr\{[\sup_{\theta\in\Theta}\|g(z_1,\theta)\|]^m/\epsilon^m\ge n\}<\infty$. But since the random vectors $z_1,\ldots,z_n$ are identically distributed, we have $\sum_{n=1}^{\infty}\Pr\{[\sup_{\theta\in\Theta}\|g(z_n,\theta)\|]^m/\epsilon^m\ge n\}<\infty$. Therefore, by Borel-Cantelli the event $\{[\sup_{\theta\in\Theta}\|g(z_n,\theta)\|]^m/\epsilon^m\ge n\}$ happens infinitely often w.p.0. Equivalently, the event $\{\sup_{\theta\in\Theta}\|g(z_n,\theta)\|/\epsilon<n^{1/m}\}$ happens for all but finitely many n w.p.1. Since $n^{1/m}$ eventually exceeds the largest element in the finite collection of $\sup_{\theta\in\Theta}\|g(z_k,\theta)\|/\epsilon$'s that exceed $k^{1/m}$, we get that $\Pr\{\max_{1\le j\le n}\sup_{\theta\in\Theta}\|g(z_j,\theta)\|<n^{1/m}\epsilon\}=1$ for large enough n. The desired result follows because ϵ can be chosen arbitrarily small. □

Lemma D.3. *Let $x_1,\ldots,x_n$ be identically distributed random vectors such that $\mathbb{E}\|x_1\|^{1+\delta}<\infty$ for some $\delta\ge 0$ and define $\mathbb{I}_{i,n}^c=\mathbb{I}\{\|x_i\|>n\}$. Then $\max_{1\le i\le n}\mathbb{I}_{i,n}^c=o_p(\frac{1}{n^\delta})$.*

Proof of Lemma D.3. Since $\mathbb{E}\|x_i\|^{1+\delta}<\infty$ implies that $\mathbb{E}\{\|x_i\|^{1+\delta}\,\mathbb{I}(\|x_i\|>n)\}=o(1)$ as $n\uparrow\infty$, we have $n^{1+\delta}\Pr\{\|x_i\|>n\}<\mathbb{E}\{\|x_i\|^{1+\delta}\,\mathbb{I}(\|x_i\|>n)\}=o(1)$. Thus $\Pr\{\|x_i\|>n\}=o(n^{-(1+\delta)})$ for each i because $x_1,\ldots,x_n$ are identically distributed. Therefore, using the fact that $\max_{1\le i\le n}\mathbb{I}_{i,n}^c\le\sum_{i=1}^n\mathbb{I}_{i,n}^c$, $\mathbb{E}\{\max_{1\le i\le n}\mathbb{I}_{i,n}^c\}\le\sum_{i=1}^n\Pr\{\|x_i\|>n\}=o(n^{-\delta})$. The desired result follows. □

Lemma D.4. *Let Assumptions 3.3 and 3.4 hold. If $b_n\downarrow 0$ and $nb_n^s\uparrow\infty$, then $\max_{1\le i\le n}|\mathbb{T}_{i,n}-1|=O_p(n\exp\{-\frac{nb_n^{2(s+\tau)}}{8\mathcal{K}_{\max}^2}\})$.*

Proof of Lemma D.4. Pick $\epsilon>0$ and let M_ϵ denote a positive number which may depend upon ϵ. We will see later how to choose M_ϵ. Because $\max_{1\le i\le n}|\mathbb{T}_{i,n}-1|\le\sum_{i=1}^n|\mathbb{T}_{i,n}-1|$,

$$\Pr\{\max_{1\le i\le n}|\mathbb{T}_{i,n}-1|>M_\epsilon\}\le\frac{1}{M_\epsilon}\sum_{i=1}^n\Pr\{\hat{h}(x_i)<b_n^\tau\} \tag{D.1}$$

follows by Chebychev's inequality. Since $\hat{h} \geq 0$,

$$\begin{aligned}\Pr\{\hat{h}(x_i) < b_n^\tau\} &\leq \Pr\{|\hat{h}(x_i) - \mathbb{E}[\hat{h}(x_i)|x_i]| > b_n^\tau\} + \Pr\{|\mathbb{E}[\hat{h}(x_i)|x_i]| < 2b_n^\tau\} \\ &\leq \Pr\{|\hat{h}(x_i) - \mathbb{E}[\hat{h}(x_i)|x_i]| > b_n^\tau\} + \Pr\{\sup_{x_i \in \mathbb{R}^s} |\mathbb{E}[\hat{h}(x_i)|x_i]| < 2b_n^\tau\}.\end{aligned}$$

Under Assumption 3.4 we can show $\sup_{x_i \in \mathbb{R}^s} |\mathbb{E}[\hat{h}(x_i)|x_i] - h(x_i)| \leq c(b_n^2 + \frac{1}{n} + \frac{1}{nb_n^s})$, which implies $\sup_{x_i \in \mathbb{R}^s} h(x_i) \leq c(b_n^2 + \frac{1}{n} + \frac{1}{nb_n^s}) + \sup_{x_i \in \mathbb{R}^s} \mathbb{E}[\hat{h}(x_i)|x_i]$. Since $\sup_{x_i \in \mathbb{R}^s} h(x_i) > 0$, $b_n \downarrow 0$, and $nb_n^s \uparrow \infty$,

$$\Pr\{\sup_{x_i \in \mathbb{R}^s} \mathbb{E}[\hat{h}(x_i)|x_i] < 2b_n^\tau\} \leq \Pr\{\sup_{x_i \in \mathbb{R}^s} h(x_i) \leq c(b_n^2 + \frac{1}{n} + \frac{1}{nb_n^s}) + 2b_n^\tau\} = 0$$

for large enough n. Hence when n is large enough, (D.1) reduces to

$$\Pr\{\max_{1 \leq i \leq n} |\mathbb{T}_{i,n} - 1| > M_\epsilon\} \leq \frac{1}{M_\epsilon} \sum_{i=1}^n \Pr\{|\hat{h}(x_i) - \mathbb{E}[\hat{h}(x_i)|x_i]| > b_n^\tau\}. \tag{D.2}$$

Now let $y_{ij} = \mathcal{K}_{ij} - \mathbb{E}\{\mathcal{K}_{ij}|x_i\}$ and note that: (i) conditional on x_i the random variables $y_{i1}, \ldots, y_{in}$ are mutually independent, (ii) $\mathbb{E}\{y_{ij}|x_i\} = 0$, and (iii) $|y_{ij}| \leq 2\mathcal{K}_{\max}$ for $1 \leq i, j \leq n$. Therefore,

$$\Pr\{|\hat{h}(x_i) - \mathbb{E}[\hat{h}(x_i)|x_i]| > b_n^\tau\} = \Pr\{|\sum_{j=1}^n y_{ij}| > nb_n^{s+\tau}\} \leq 2\exp\{-\frac{nb_n^{2(s+\tau)}}{8\mathcal{K}_{\max}^2}\},$$

where the last inequality follows from an application of Hoeffding's inequality[8]. Using this result, for large enough n, (D.2) becomes $\Pr\{\max_{1 \leq i \leq n} |\mathbb{T}_{i,n} - 1| > M_\epsilon\} \leq 2n\exp\{-\frac{nb_n^{2(s+\tau)}}{8\mathcal{K}_{\max}^2}\}M_\epsilon^{-1}$. Thus for large enough n, $\Pr\{\max_{1 \leq i \leq n} |\mathbb{T}_{i,n} - 1| > n\exp\{-\frac{nb_n^{2(s+\tau)}}{8\mathcal{K}_{\max}^2}\}\frac{2}{\epsilon}\} \leq \epsilon$ follows on replacing M_ϵ by $n\exp\{-\frac{nb_n^{2(s+\tau)}}{8\mathcal{K}_{\max}^2}\}\frac{2}{\epsilon}$. Since ϵ was arbitrary, the desired result follows. □

Lemma D.5. *Let Assumptions 3.3 and 3.4 hold. Assume that $b_n \downarrow 0$ and $n^{1-\beta}b_n^s \uparrow \infty$ for some $\beta \in (0,1)$. Then for $\hat{\mathbb{T}}_{i,n} \stackrel{def}{=} \mathbb{T}_{i,n}h(x_i)/\hat{h}(x_i)$,*

$$\max_{1 \leq i \leq n} |\hat{\mathbb{T}}_{i,n} - 1| = o_p(\sqrt{\frac{n^\beta}{nb_n^{s+2\tau}}}) + O(\frac{b_n^2}{b_n^\tau}) + o(\frac{1}{n^\varrho b_n^\tau}) + O_p(n\exp\{-\frac{nb_n^{2(s+\tau)}}{8\mathcal{K}_{\max}^2}\})$$

$$\max_{1 \leq i \leq n} |\hat{\mathbb{T}}_{i,n}^2 - 1| = o_p(\sqrt{\frac{n^\beta}{nb_n^{s+4\tau}}}) + O_p(\frac{b_n^2}{b_n^{2\tau}}) + o_p(\frac{1}{n^\varrho b_n^{2\tau}}) + O_p(n\exp\{-\frac{nb_n^{2(s+\tau)}}{8\mathcal{K}_{\max}^2}\}).$$

Proof of Lemma D.5. Since

$$|\hat{\mathbb{T}}_{i,n} - 1| = |\frac{\mathbb{T}_{i,n}\{h(x_i) - \hat{h}(x_i)\}}{\hat{h}(x_i)} + \mathbb{T}_{i,n} - 1| \leq \frac{1}{b_n^\tau}|\hat{h}(x_i) - h(x_i)| + |\mathbb{T}_{i,n} - 1|,$$

the first result follows by Lemma B.4 and Lemma D.4. Similarly,

$$|\hat{\mathbb{T}}_{i,n}^2 - 1| \leq \frac{1}{b_n^{2\tau}}|\hat{h}^2(x_i) - h^2(x_i)| + |\mathbb{T}_{i,n} - 1| = \frac{1}{b_n^{2\tau}}|\hat{h}(x_i) - h(x_i)|\,|\hat{h}(x_i) + h(x_i)| + |\mathbb{T}_{i,n} - 1|,$$

[8] The usual statement of Hoeffding's inequality (see, for instance, Pollard (1984, Appendix B)) requires that the summands be mutually independent with zero mean. However, it is easy to verify that Hoeffding's inequality also holds when the summands are conditionally independent with zero conditional mean. See, for example, Devroye, Györfi, and Lugosi (1996, Section 9.1).

and the second result also follows by Lemma B.4 and Lemma D.4. □

Lemma D.6. *Let $f(z)$ be a real valued function such that $\mathbb{E}|f(z)| < \infty$, and let Assumption 3.3 hold. Then $\mathbb{E}\{\sum_{j=1}^n |f(z_j)|w_{ij}\} \leq c\mathbb{E}|f(z_1)|$, where the constant c only depends upon the kernel.*

Proof of Lemma D.6. Follows directly from Devroye and Wagner (1980, Lemma 2, Page 233). □

Lemma D.7. *Let $f(z)$ be a real valued function such that $\mathbb{E}|f(z)|^a < \infty$ for $a > 0$, and let Assumption 3.3 hold. Then $\Pr\{\sup_{x_i \in \mathbb{R}^s} |\sum_{j=1}^n f(z_j)w_{ij}| = o(n^{1/a})\} = 1$ as $n \uparrow \infty$.*

Proof of Lemma D.7. Observe that $|\sum_{j=1}^n f(z_j)w_{ij}| \leq \max_{1\leq j\leq n} |f(z_j)|$. The desired result now follows by Lemma D.2. □

Lemma D.8. *Let $f(z)$ be a real valued function such that $\mathbb{E}|f(z)|^a < \infty$ for $a > 0$, and let Assumptions 3.3–3.4 hold. If $b_n \downarrow 0$ and $\frac{\log n}{nb_n^s} \downarrow 0$, then $\Pr\{\sup_{x_i \in \mathbb{R}^s} |\frac{1}{nb_n^s}\sum_{j=1}^n f(z_j)\mathcal{K}_{ij}| = o(n^{1/a})\} = 1$ as $n \uparrow \infty$.*

Proof of Lemma D.8. By the triangle inequality,

$$\sup_{x_i \in \mathbb{R}^s} \frac{1}{nb_n^s}\sum_{j=1}^n |f(z_j)|\mathcal{K}_{ij} \leq \max_{1\leq j\leq n} |f(z_j)|\{\sup_{x_i \in \mathbb{R}^s} |\hat{h}(x_i) - h(x_i)| + \sup_{x_i \in \mathbb{R}^s} h(x_i)\}.$$

But under Assumptions 3.3 and 3.4, we can use the strong uniform consistency of $\hat{h}(x_i)$ (see, for instance, Prakasa Rao (1983, Page 185)) to show $\sup_{x_i \in \mathbb{R}^s} |\hat{h}(x_i) - h(x_i)| \xrightarrow{a.s.} 0$ if $\frac{\log n}{nb_n^s} \downarrow 0$. Furthermore, $h(x_i)$ is uniformly bounded on $\mathbb{R}^s$ by assumption, and from Lemma D.2 we know that $\max_{1\leq j\leq n} |f(z_j)| = o(n^{1/a})$ w.p.1 as $n \uparrow \infty$. Therefore, $\sup_{x_i \in \mathbb{R}^s} \frac{1}{nb_n^s}\sum_{j=1}^n |f(z_j)|\mathcal{K}_{ij} = o(n^{1/a})$ holds w.p.1 provided $\frac{\log n}{nb_n^s} \downarrow 0$. The desired result follows. □

Lemma D.9. *Let $f(z)$ be a real valued function such that $\mathbb{E}|f(z)|^2 < \infty$, and let Assumptions 3.3 and 3.4 hold. Then $\mathbb{E}\{\frac{1}{nb_n^s}\sum_{j=1}^n |f(z_j)|\mathcal{K}_{ij}\} < \infty$.*

Proof of Lemma D.9. Since $\frac{1}{nb_n^s}\sum_{j=1}^n |f(z_j)|\mathcal{K}_{ij} = \sum_{j=1}^n |f(z_j)|w_{ij}\,\hat{h}(x_i)$,

$$\frac{1}{nb_n^s}\sum_{j=1}^n |f(z_j)|\mathcal{K}_{ij} \leq \frac{1}{2}\{[\sum_{j=1}^n |f(z_j)|w_{ij}]^2 + \hat{h}^2(x_i)\} \overset{\text{Jensen}}{\leq} \frac{1}{2}\{\sum_{j=1}^n |f(z_j)|^2 w_{ij} + \hat{h}^2(x_i)\}.$$

It is easily shown that $\mathbb{E}\hat{h}^2(x_i) < \infty$. Hence the desired result follows by Lemma D.6. □

Appendix E. Simulation Results

Table 1. $n = 50$

Estimator	Bandwidth	Std. Dev.	Bias	RMSE	Ratio RMSE	MAE	Ratio MAE
OLS		0.4183	0.0154	0.4186	3.1837	0.2551	2.7950
		0.3208	-0.0070	0.3209	2.1750	0.1991	2.0429
FGLS		0.1668	0.0054	0.1669	1.2696	0.1033	1.1316
		0.2078	0.0033	0.2078	1.4085	0.1284	1.3177
GLS		0.1314	0.0054	0.1315	1.0000	0.0913	1.0000
		0.1475	0.0022	0.1475	1.0000	0.0975	1.0000
k-NN	automatic	0.2217	0.0112	0.2220	1.6884	0.1337	1.4648
		0.2347	-0.0008	0.2347	1.5910	0.1494	1.5329
	$k_n = 3$	0.2545	0.0137	0.2549	1.9389	0.1566	1.7159
		0.2311	-0.0049	0.2312	1.5669	0.1444	1.4816
	$k_n = 6$	0.2172	0.0121	0.2175	1.6543	0.1345	1.4738
		0.2190	-0.0012	0.2190	1.4842	0.1360	1.3957
	$k_n = 9$	0.2097	0.0100	0.2100	1.5970	0.1253	1.3733
		0.2212	0.0007	0.2212	1.4994	0.1331	1.3653
	$k_n = 12$	0.2111	0.0103	0.2114	1.6077	0.1231	1.3487
		0.2248	0.0003	0.2248	1.5236	0.1335	1.3701
	$k_n = 15$	0.2147	0.0119	0.2151	1.6358	0.1262	1.3829
		0.2306	-0.0012	0.2307	1.5633	0.1400	1.4370
	$k_n = 18$	0.2196	0.0132	0.2200	1.6736	0.1297	1.4211
		0.2354	-0.0025	0.2354	1.5957	0.1463	1.5008
	$k_n = 21$	0.2231	0.0132	0.2235	1.6998	0.1296	1.4200
		0.2394	-0.0029	0.2394	1.6228	0.1506	1.5457
	$k_n = 24$	0.2281	0.0117	0.2284	1.7373	0.1329	1.4567
		0.2440	-0.0028	0.2440	1.6537	0.1562	1.6025

Table 2. $n = 50$ (continued)

Estimator	Bandwidth	Std. Dev.	Bias	RMSE	Ratio RMSE	MAE	Ratio MAE
Kernel	automatic	0.2093	0.0117	0.2096	1.5944	0.1354	1.4836
		0.2314	-0.0033	0.2314	1.5685	0.1501	1.5402
	$b_n = 0.1524$	0.2296	0.0087	0.2298	1.7476	0.1463	1.6030
		0.2329	0.0012	0.2329	1.5785	0.1465	1.5033
	$b_n = 0.3049$	0.1994	0.0099	0.1997	1.5189	0.1271	1.3927
		0.2176	-0.0006	0.2176	1.4749	0.1392	1.4286
	$b_n = 0.4573$	0.1890	0.0101	0.1893	1.4395	0.1181	1.2936
		0.2120	-0.0019	0.2121	1.4372	0.1385	1.4212
	$b_n = 0.6097$	0.1869	0.0102	0.1872	1.4237	0.1205	1.3202
		0.2118	-0.0027	0.2118	1.4357	0.1384	1.4200
	$b_n = 0.7622$	0.1915	0.0105	0.1918	1.4588	0.1239	1.3573
		0.2165	-0.0031	0.2165	1.4676	0.1408	1.4443
	$b_n = 0.9146$	0.2016	0.0108	0.2019	1.5358	0.1295	1.4188
		0.2249	-0.0035	0.2250	1.5247	0.1467	1.5055
	$b_n = 1.0670$	0.2157	0.0113	0.2160	1.6425	0.1369	1.5003
		0.2351	-0.0039	0.2351	1.5936	0.1524	1.5637
	$b_n = 1.2195$	0.2319	0.0117	0.2322	1.7664	0.1405	1.5399
		0.2454	-0.0043	0.2454	1.6635	0.1620	1.6620
SEL	automatic	0.1777	0.0078	0.1778	1.3525	0.1100	1.2047
		0.1803	0.0001	0.1803	1.2218	0.1090	1.1186
	$b_n = 0.1524$	0.1957	0.0073	0.1959	1.4898	0.1285	1.4083
		0.1996	0.0004	0.1996	1.3526	0.1297	1.3311
	$b_n = 0.3049$	0.1874	0.0073	0.1876	1.4266	0.1209	1.3241
		0.1909	0.0006	0.1909	1.2938	0.1197	1.2279
	$b_n = 0.4573$	0.1803	0.0061	0.1804	1.3724	0.1121	1.2287
		0.1834	0.0016	0.1834	1.2428	0.1136	1.1657
	$b_n = 0.6097$	0.1769	0.0049	0.1770	1.3463	0.1114	1.2200
		0.1809	0.0028	0.1809	1.2260	0.1107	1.1363
	$b_n = 0.7622$	0.1710	0.0065	0.1711	1.3015	0.1100	1.2056
		0.1754	0.0014	0.1754	1.1886	0.1123	1.1522
	$b_n = 0.9146$	0.1715	0.0076	0.1717	1.3058	0.1095	1.2000
		0.1750	0.0004	0.1750	1.1863	0.1109	1.1384
	$b_n = 1.0670$	0.1747	0.0082	0.1749	1.3302	0.1099	1.2046
		0.1766	-0.0002	0.1766	1.1969	0.1100	1.1285
	$b_n = 1.2195$	0.1797	0.0088	0.1799	1.3681	0.1110	1.2166
		0.1796	-0.0007	0.1796	1.2170	0.1128	1.1574

Table 3. $n = 100$

Estimator	Bandwidth	Std. Dev.	Bias	RMSE	Ratio RMSE	MAE	Ratio MAE
OLS		0.3735	0.0075	0.3735	4.0987	0.2084	3.4215
		0.2693	-0.0039	0.2693	2.5979	0.1587	2.2635
FGLS		0.1192	0.0099	0.1196	1.3121	0.0751	1.2328
		0.1471	-0.0057	0.1472	1.4201	0.0955	1.3627
GLS		0.0909	0.0064	0.0911	1.0000	0.0609	1.0000
		0.1036	-0.0020	0.1037	1.0000	0.0701	1.0000
k-NN	automatic	0.1568	0.0084	0.1571	1.7235	0.0814	1.3360
		0.1690	-0.0028	0.1690	1.6307	0.1026	1.4628
	$k_n = 3$	0.2131	0.0110	0.2134	2.3416	0.1252	2.0549
		0.1785	-0.0089	0.1787	1.7243	0.1238	1.7663
	$k_n = 6$	0.1577	0.0082	0.1579	1.7326	0.0960	1.5766
		0.1529	-0.0053	0.1530	1.4763	0.1043	1.4884
	$k_n = 10$	0.1480	0.0077	0.1482	1.6266	0.0867	1.4229
		0.1527	-0.0032	0.1527	1.4734	0.1004	1.4325
	$k_n = 13$	0.1475	0.0079	0.1477	1.6211	0.0802	1.3170
		0.1557	-0.0024	0.1557	1.5022	0.0997	1.4221
	$k_n = 16$	0.1489	0.0067	0.1490	1.6353	0.0770	1.2641
		0.1605	-0.0015	0.1605	1.5486	0.1026	1.4630
	$k_n = 20$	0.1538	0.0068	0.1540	1.6893	0.0813	1.3341
		0.1662	-0.0017	0.1662	1.6031	0.1032	1.4715
	$k_n = 23$	0.1563	0.0066	0.1564	1.7160	0.0811	1.3321
		0.1700	-0.0018	0.1700	1.6399	0.1023	1.4596
	$k_n = 26$	0.1596	0.0070	0.1597	1.7529	0.0823	1.3513
		0.1733	-0.0024	0.1733	1.6719	0.1025	1.4623

Table 4. $n = 100$ (continued)

Estimator	Bandwidth	Std. Dev.	Bias	RMSE	Ratio RMSE	MAE	Ratio MAE
Kernel	automatic	0.1413	0.0099	0.1417	1.5543	0.0831	1.3638
		0.1607	-0.0057	0.1608	1.5508	0.0993	1.4160
	$b_n = 0.1327$	0.1640	0.0122	0.1645	1.8047	0.1002	1.6458
		0.1673	-0.0061	0.1674	1.6151	0.1139	1.6246
	$b_n = 0.2654$	0.1373	0.0096	0.1376	1.5102	0.0825	1.3549
		0.1499	-0.0041	0.1500	1.4470	0.1005	1.4332
	$b_n = 0.3981$	0.1239	0.0091	0.1243	1.3636	0.0755	1.2403
		0.1413	-0.0044	0.1413	1.3634	0.0912	1.3013
	$b_n = 0.5308$	0.1209	0.0089	0.1212	1.3298	0.0727	1.1941
		0.1392	-0.0046	0.1393	1.3434	0.0900	1.2833
	$b_n = 0.6635$	0.1225	0.0089	0.1228	1.3477	0.0741	1.2163
		0.1416	-0.0048	0.1417	1.3668	0.0915	1.3058
	$b_n = 0.7962$	0.1281	0.0090	0.1284	1.4090	0.0763	1.2520
		0.1479	-0.0049	0.1480	1.4277	0.0932	1.3289
	$b_n = 0.9289$	0.1371	0.0091	0.1374	1.5072	0.0799	1.3117
		0.1568	-0.0049	0.1568	1.5131	0.0965	1.3759
	$b_n = 1.0616$	0.1487	0.0091	0.1489	1.6342	0.0851	1.3972
		0.1668	-0.0048	0.1669	1.6099	0.1050	1.4979
SEL	automatic	0.1145	0.0086	0.1148	1.2596	0.0743	1.2202
		0.1206	-0.0035	0.1207	1.1643	0.0769	1.0962
	$b_n = 0.1327$	0.1343	0.0106	0.1347	1.4777	0.0875	1.4358
		0.1384	-0.0017	0.1384	1.3355	0.0909	1.2962
	$b_n = 0.2654$	0.1232	0.0094	0.1235	1.3554	0.0757	1.2433
		0.1289	-0.0007	0.1289	1.2431	0.0831	1.1859
	$b_n = 0.3981$	0.1172	0.0083	0.1175	1.2893	0.0726	1.1912
		0.1242	-0.0006	0.1242	1.1981	0.0771	1.1004
	$b_n = 0.5308$	0.1142	0.0084	0.1145	1.2561	0.0709	1.1643
		0.1217	-0.0016	0.1217	1.1739	0.0768	1.0956
	$b_n = 0.6635$	0.1122	0.0083	0.1125	1.2343	0.0698	1.1461
		0.1197	-0.0025	0.1198	1.1553	0.0753	1.0746
	$b_n = 0.7962$	0.1119	0.0084	0.1122	1.2316	0.0703	1.1534
		0.1192	-0.0031	0.1192	1.1504	0.0743	1.0601
	$b_n = 0.9289$	0.1131	0.0086	0.1135	1.2450	0.0730	1.1991
		0.1197	-0.0036	0.1198	1.1554	0.0740	1.0559
	$b_n = 1.0616$	0.1156	0.0088	0.1159	1.2718	0.0754	1.2376
		0.1212	-0.0038	0.1213	1.1697	0.0776	1.1074

Table 5. $n = 200$

Estimator	Bandwidth	Std. Dev.	Bias	RMSE	Ratio RMSE	MAE	Ratio MAE
OLS		0.3124	-0.0019	0.3124	4.6385	0.1676	3.4328
		0.2216	0.0003	0.2216	2.9158	0.1325	2.5419
FGLS		0.0977	0.0053	0.0979	1.4529	0.0563	1.1544
		0.1019	-0.0028	0.1019	1.3413	0.0661	1.2678
GLS		0.0673	0.0030	0.0674	1.0000	0.0488	1.0000
		0.0760	-0.0011	0.0760	1.0000	0.0521	1.0000
k-NN	automatic	0.1006	0.0037	0.1006	1.4941	0.0594	1.2163
		0.1125	-0.0016	0.1125	1.4805	0.0708	1.3579
	$k_n = 4$	0.1414	0.0035	0.1415	2.1006	0.0931	1.9077
		0.1238	-0.0012	0.1238	1.6285	0.0827	1.5867
	$k_n = 8$	0.1026	0.0033	0.1027	1.5245	0.0632	1.2948
		0.1045	-0.0013	0.1045	1.3752	0.0695	1.3334
	$k_n = 12$	0.0959	0.0042	0.0960	1.4252	0.0583	1.1934
		0.1026	-0.0025	0.1026	1.3501	0.0664	1.2734
	$k_n = 16$	0.0963	0.0047	0.0964	1.4313	0.0584	1.1957
		0.1046	-0.0022	0.1046	1.3763	0.0655	1.2567
	$k_n = 20$	0.0969	0.0039	0.0970	1.4404	0.0565	1.1568
		0.1071	-0.0020	0.1072	1.4101	0.0677	1.2981
	$k_n = 24$	0.0980	0.0040	0.0981	1.4566	0.0582	1.1928
		0.1096	-0.0020	0.1097	1.4428	0.0703	1.3477
	$k_n = 28$	0.1001	0.0037	0.1001	1.4869	0.0598	1.2256
		0.1126	-0.0016	0.1126	1.4822	0.0715	1.3712
	$k_n = 32$	0.1017	0.0033	0.1017	1.5100	0.0603	1.2363
		0.1153	-0.0015	0.1153	1.5177	0.0731	1.4014

TABLE 6. $n = 200$ (continued)

Estimator	Bandwidth	Std. Dev.	Bias	RMSE	Ratio RMSE	MAE	Ratio MAE
Kernel	automatic	0.0899	0.0048	0.0900	1.3363	0.0627	1.2841
		0.1043	-0.0033	0.1043	1.3729	0.0682	1.3082
	$b_n = 0.1155$	0.1122	0.0053	0.1123	1.6672	0.0639	1.3084
		0.1169	-0.0020	0.1169	1.5381	0.0771	1.4781
	$b_n = 0.2310$	0.0918	0.0042	0.0919	1.3641	0.0544	1.1144
		0.1009	-0.0021	0.1009	1.3279	0.0663	1.2719
	$b_n = 0.3466$	0.0834	0.0042	0.0836	1.2404	0.0530	1.0856
		0.0941	-0.0027	0.0942	1.2389	0.0632	1.2122
	$b_n = 0.4621$	0.0808	0.0048	0.0810	1.2019	0.0520	1.0657
		0.0920	-0.0033	0.0920	1.2111	0.0623	1.1944
	$b_n = 0.5776$	0.0803	0.0048	0.0805	1.1945	0.0534	1.0930
		0.0922	-0.0035	0.0923	1.2144	0.0629	1.2072
	$b_n = 0.6931$	0.0818	0.0048	0.0819	1.2163	0.0564	1.1561
		0.0947	-0.0034	0.0947	1.2464	0.0627	1.2027
	$b_n = 0.8087$	0.0854	0.0048	0.0855	1.2699	0.0592	1.2133
		0.0992	-0.0034	0.0992	1.3057	0.0648	1.2426
	$b_n = 0.9242$	0.0910	0.0048	0.0911	1.3525	0.0626	1.2818
		0.1053	-0.0034	0.1054	1.3862	0.0687	1.3169
SEL	automatic	0.0777	0.0041	0.0778	1.1546	0.0528	1.0824
		0.0831	-0.0022	0.0831	1.0941	0.0571	1.0947
	$b_n = 0.1155$	0.0910	0.0039	0.0910	1.3516	0.0600	1.2291
		0.0944	-0.0009	0.0944	1.2426	0.0634	1.2165
	$b_n = 0.2310$	0.0868	0.0034	0.0868	1.2894	0.0566	1.1589
		0.0897	-0.0010	0.0897	1.1797	0.0578	1.1077
	$b_n = 0.3466$	0.0827	0.0034	0.0827	1.2283	0.0554	1.1349
		0.0869	-0.0012	0.0869	1.1435	0.0560	1.0742
	$b_n = 0.4621$	0.0800	0.0038	0.0801	1.1898	0.0518	1.0621
		0.0847	-0.0016	0.0847	1.1147	0.0573	1.0984
	$b_n = 0.5776$	0.0781	0.0040	0.0782	1.1608	0.0506	1.0359
		0.0834	-0.0018	0.0835	1.0982	0.0569	1.0917
	$b_n = 0.6931$	0.0772	0.0040	0.0773	1.1472	0.0498	1.0205
		0.0829	-0.0020	0.0829	1.0906	0.0571	1.0942
	$b_n = 0.8087$	0.0771	0.0041	0.0772	1.1458	0.0512	1.0488
		0.0828	-0.0021	0.0828	1.0893	0.0581	1.1137
	$b_n = 0.9242$	0.0778	0.0041	0.0779	1.1561	0.0539	1.1047
		0.0831	-0.0022	0.0831	1.0940	0.0575	1.1035

References

Ai, C. (1997): "A Semiparametric Maximum Likelihood Estimator," *Econometrica*, 65, 933–963.

Amemiya, T. (1974): "The Non-Linear Least Square Estimator," *Journal of Econometrics*, 2, 105–110.

——— (1977): "The Maximum Likelihood and the Nonlinear Three Stage Least Squares Estimator in the General Nonlinear Simultaneous Equation Model.," *Econometrica*, 45, 955–968.

——— (1985): *Advanced Econometrics*. Harvard University Press.

Beran, R. (1988): "Prepivoting Test Statistics: A Bootstrap View of Asymptotic Refinements," *Journal of the American Statistical Association*, 83, 687–697.

Brown, B. W., and W. K. Newey (1998): "Efficient Bootstrapping for Semiparametric Models," Manuscript.

Chamberlain, G. (1987): "Asymptotic efficiency in estimation with conditional moment restrictions," *Journal of Econometrics*, 34, 305–334.

Chen, S. X. (1996): "Empirical Likelihood Confidence Intervals for Nonparametric Density Estimation," *Biometrika*, 83, 329–341.

Cragg, J. (1983): "More efficient estimation in the presence of heteroscedasticity of unknown form," *Econometrica*, 49, 751–764.

Devroye, L., L. Györfi, and G. Lugosi (1996): *A Probabilistic Theory of Pattern Recognition*. Springer.

Devroye, L. P., and T. J. Wagner (1980): "Distribution-Free Consistency Results in Nonparametric Discrimination and Regression Function Estimation," *The Annals of Statistics*, 8, 231–239.

DiCiccio, T., P. Hall, and J. Romano (1991): "Empirical Likelihood is Bartlett-Correctable," *Annals of Statisics*, 19, 1053–1061.

Donald, S. G., G. W. Imbens, and W. K. Newey (2001): "Empirical Likelihood Estimation and Consistent Tests with Conditional Moment Restrictions," Manuscript.

Fisher, N. I., P. Hall, B.-Y. Jing, and A. T. Wood (1996): "Improved Pivotal Methods for Constructing Confidence Regions with Directional Data," *Journal of the American Statistical Association*, 91, 1062–1070.

Hall, P. (1990): "Pseudo-Likelihood Theory for Empirical Likelihood," *Annals of Statisics*, 18, 121–140.

Hall, P., and B. LaScala (1990): "Methodology and Algorithms for Empirical Likelihood," *International Statistical Review*, 58, 109–127.

Hansen, L. P. (1982): "Large Sample Properties of Generalized Methods of Moments Estimators," *Econometrica*, 50, 1029–1054.

Hoeffding, W. (1963): "Asymptotically Optimal Tests for Multinomial Distributions," *Annals of Mathematical Statistics*, 36, 369–408.

Imbens, G. W. (1993): "A New Approach to Generalized Method of Moments Estimators," Manuscript. Department of Economics, Harvard University.

Imbens, G. W., R. H. Spady, and P. Johnson (1998): "Information Theoretic Approaches to Inference in Moment Condition Models," *Econometrica*, 66, 333–357.

Kitamura, Y. (1997a): "Empirical Likelihood and the Bootstrap for Time Series Regressions," Working Paper, Department of Economics, University of Minnesota.

——— (1997b): "Empirical likelihood methods with weakly dependent processes," *Annals of Statistics*, 25(5), 2084–2102.

——— (2000): "Asymptotic Optimality of Emptrical Likelihood for Testing Moment Restrictions," Forthcoming in *Econometrica*.

Kitamura, Y., and M. Stutzer (1997): "An Information Theoretic Alternative to Generalized Method of Moments Estimation," *Econometrica*, 65(4), 861–874.

LeBlanc, M., and J. Crowley (1995): "Semiparametric Regression Functionals," *Journal of the American Statistical Association*, 90(429), 95–105.

Newey, W. K. (1990): "Efficient Instrumental Variables Estimation of Nonlinear Models," *Econometrica*, 58, 809–837.

——— (1993): "Efficient Estimation of Models with Conditional Moment Restrictions," in *Handbook of Statistics, vol. 11*, ed. by G. Maddala, C. Rao, and H. Vinod. Elsevier Science B.V., pp. 2111–2245.

Newey, W. K., and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics, vol. IV*, ed. by R. Engle, and D. McFadden. Elsevier Science B.V., pp. 2111–2245.

Newey, W. K., and R. J. Smith (2000): "Asymptotic Bias and Equivalence of GMM and GEL," Manuscript.

Owen, A. (1988): "Empirical Likelihood Ratio Confidence Intervals for a Single Functional," *Biometrika*, 75(2), 237–249.

——— (1990): "Empirical Likelihood Ratio Confidence Regions," *The Annals of Statistics*, 18(1), 90–120.

——— (1991): "Empirical Likelihood for Linear Models," *The Annals of Statistics*, 19(4), 1725–1747.

——— (1995): "Empirical Likelihood," *Encyclopedia of Statistical Sciences*, Forthcoming.

Pollard, D. (1984): *Convergence of Stochastic Processes*. Springer-Verlag.

Prakasa Rao, B. (1983): *Nonparametric Functional Estimation*. Academic Press.

Qin, J., and J. Lawless (1994): "Empirical Likelihood and General Estimating Equations," *Annals of Statistics*, 22, 300–325.

Robinson, P. M. (1987): "Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form," *Econometrica*, 55, 875–891.

Tripathi, G., and Y. Kitamura (2000): "On Testing Conditional Moment Restrictions: The Canonical Case," Manuscript. Department of Economics, University of Wisconsin-Madison.

Zhang, J., and I. Gijbels (2001): "Sieve Empirical Likelihood and Extensions of the Generalized Least Squares," Manuscript.

Department of Economics, University of Wisconsin, Madison, WI-53706.
E-mail address: ykitamur@ssc.wisc.edu

Department of Economics, University of Wisconsin, Madison, WI-53706.
E-mail address: gtripath@ssc.wisc.edu

Department of Economics, Dongguk University, Seoul, Korea, 100-715.
E-mail address: htahn@uoscc.uos.ac.kr