



LOCAL GEL METHODS FOR CONDITIONAL MOMENT RESTRICTIONS

Richard J. Smith

THE INSTITUTE FOR FISCAL STUDIES
DEPARTMENT OF ECONOMICS, UCL
cemmap working paper CWP15/05

LOCAL GEL METHODS FOR CONDITIONAL MOMENT RESTRICTIONS*

Richard J. Smith
cemmap
U.C.L and I.F.S.
and
Faculty of Economics
University of Cambridge

This Revision: November 2005

Abstract

The principal purpose of this paper is to adapt to the conditional moment context the GEL unconditional moment methods described in Smith (1997, 2001) and Newey and Smith (2004). In particular we develop GEL estimators which achieve the semi-parametric efficiency lower bound. The requisite GEL criteria are constructed by local smoothing and parallel the local semiparametric efficient EL method formulated by Kitamura, Tripathi and Ahn (2004) for conditional moment restrictions. A particular advantage of these efficient local methods is the avoidance of the necessity of providing explicit estimators for the Jacobian and conditional variance matrices. The class of local GEL estimators admits a number of alternative first order equivalent estimators such as local EL, local ET and local CUE as in the unconditional moment restrictions case. The paper also provides a local GEL criterion function test statistic for parametric restrictions.

JEL Classification: C12, C13, C14, C20, C30.

Keywords: Conditional Moment Restrictions, Local Generalized Empirical Likelihood, GMM, Semi-Parametric Efficiency.

*This is a revision of an Invited Paper entitled “Conditional Generalized Empirical Likelihood Estimation” presented at the International Conference “Contributions to Econometric Theory” held in memory of Michael Magdalinos at the Athens University of Economics and Business, 20-21 November 2003. An earlier version of this paper also comprised part of the Invited Address “Local GEL Inference for Conditional Moment Restrictions Models” presented at the Econometric Society European Meetings, Madrid, 20-24 August, 2004. Financial support for this research from a 2002 Leverhulme Major Research Fellowship is gratefully acknowledged.

1 Introduction

Simulation evidence increasingly indicates that for many models specified by unconditional moment restrictions the generalized method of moments (GMM) estimator, Hansen (1982), may be substantially biased in finite samples, especially so when there are large numbers of moment conditions. See, for example, Altonji and Segal (1996), Imbens and Spady (2001), Judge and Mittelhammer (2001), Ramalho (2001) and Newey, Ramalho and Smith (2005). Newey and Smith (2004), henceforth NS, provides theoretical underpinning for these findings. Alternative estimators which are first order asymptotically equivalent to GMM include empirical likelihood (EL), [Owen (1988), Qin and Lawless (1994), and Imbens (1997)], the continuous updating estimator (CUE), [Hansen, Heaton, and Yaron (1996)], and exponential tilting (ET), [Kitamura and Stutzer (1997) and Imbens, Spady and Johnson (1998)]. See also Owen (2001). NS show that these estimators and those from the Cressie and Read (1984) power divergence family of discrepancies are members of a class of generalized empirical likelihood (GEL) estimators and have a common structure; see Brown and Newey (1992, 2002) and Smith (1997, 2001). Correspondingly NS also demonstrate that GEL and GMM estimators are asymptotically equivalent and thus possess the same first order asymptotic properties. For the unconditional context, NS describe the higher order efficiency of bias-corrected EL. Also see Kitamura (2001).

The principal aim of this paper is adapt the GEL method to the conditional moment context and, thereby, to describe GEL estimators which achieve the semi-parametric efficiency lower bound. In an important recent paper, Kitamura, Tripathi and Ahn (2004), henceforth KTA, develops a semi-parametric efficient estimation method based on EL for models specified by conditional moment restrictions. Like KTA for EL we employ a kernel weighted version of GEL. The resultant GEL criterion may be regarded as a form of local GEL. We thus term the resultant estimators local GEL estimators. We show that local GEL estimators are asymptotically first order equivalent to the local EL estimator proposed by KTA. Consequently local GEL estimators achieve the semi-parametric effi-

ciency lower bound; see Chamberlain (1987). The class of local GEL estimators includes local versions of EL as in KTA, the ET estimator and the CUE which is related to the estimator suggested by Bonnal and Renault (2003). Because of their one-step nature a particular advantage of these efficient local methods is the avoidance of the necessity of providing explicit nonparametric estimators for the conditional Jacobian and variance matrices which may require large numbers of observations to be good approximants. See, for example, Robinson (1987) and Newey (1990, 1993) for semi-parametric approaches based on explicit conditional Jacobian and variance matrix estimation. An alternative approach to the local EL and GEL methods suggested here is that in Donald, Imbens and Newey (2001) which employs a sequence of unconditional moment restrictions based, for example, on spline or series approximants, within the standard unconditional GEL set-up as discussed in NS. The first order conditions arising from this sequence of restrictions approximate those based on semi-parametric efficient conditional moment restrictions from which, therefore, a semi-parametric efficient estimator also results. Their method has the computational virtue of avoiding estimation of nuisance parameter vectors whose number increases directly with sample size although the number of unconditional moment restrictions is required to increase with sample size but at a slower rate. It also incurs the expense of not producing an estimator for the conditional distribution of the data.

A reformulation of the first order conditions defining the local GEL estimator facilitates an intuition for the semi-parametric efficiency of the local GEL estimator. The structure of these conditions conforms to those describing a semi-parametric efficient GMM estimator, that is, they implicitly incorporate consistent estimators of the conditional Jacobian matrix and conditional variance matrix of the associated conditional moment restrictions.

A test for parametric restrictions may be based on the local GEL criterion function. Unlike asymptotically equivalent Wald or Lagrange multiplier statistics but similar to the fully parametric likelihood ratio statistic this form of statistic does not require an estimator for the asymptotic variance matrix of the local GEL estimator which may be

problematic in small samples.

The outline of the paper then is as follows. In Section 2 the conditional moment restrictions model is described. Section 3 details the local GEL method, obtains local EL, ET, CUE and Cressie-Read type discrepancy estimators as special cases and provides some interpretations for local GEL estimators. Various regularity conditions are given and the consistency, asymptotic normality and semi-parametric efficiency of the local GEL estimator stated in section 4. Section 5 discusses the local GEL criterion function statistic for parametric restrictions. Proofs of the results are given in Appendix A with certain subsidiary results and proofs in Appendix B.

2 The Model

Let (x_i, z_i) , $(i = 1, \dots, n)$, be i.i.d. observations on the s - and d -dimensional data vectors x and z . As in KTA, we assume x to be continuously distributed whereas z may be discrete, mixed or continuous, although the analysis may be straightforwardly adapted for x discrete or mixed, see KTA, section 3. Also, let β be a $p \times 1$ parameter vector which is of inferential interest and $u(z, \beta)$ be a q -vector of known functions of the data observation z and β . The parameter vector β is assumed to lie in the compact parameter space \mathcal{B} .

The model is completed by the true parameter value $\beta_0 \in \text{int}(\mathcal{B})$ which satisfies the conditional moment restriction

$$E[u(z, \beta_0)|x] = 0 \text{ w.p.1,} \tag{2.1}$$

where $E[\cdot|x]$ denotes expectation taken with respect to the conditional distribution of z given x . In many applications, the conditional moment indicator $u(z, \beta)$ would be a vector of residuals.

From (2.1), by the law of iterated expectations, any measurable function of the conditioning vector x is uncorrelated with $u(z, \beta_0)$. Therefore, we may construct a $m \times q$ matrix of instruments, $v(x, \beta_0)$ say, with $m \geq p$, and formulate the unconditional moment

restrictions

$$E[v(x, \beta_0)u(z, \beta_0)] = 0 \quad (2.2)$$

from (2.1), where $E[\cdot]$ denotes expectation taken with respect to the joint unconditional distribution of x and z . Under appropriate regularity conditions, see *inter alia* Newey and McFadden (1994) and NS, GMM or GEL estimation using $v(x, \beta)u(z, \beta)$ as the vector of (unconditional) moment indicators will deliver consistent estimators for β_0 . In general, neither unconditional GMM nor GEL estimation will achieve the semi-parametric efficiency bound because the instrumental variables $v(x, \beta_0)$ are inefficient. Chamberlain (1987) demonstrated that the semi-parametric efficiency lower bound for any $n^{1/2}$ -consistent regular estimator of β_0 under (2.1) is given by \mathcal{I}^{-1} where $\mathcal{I} \equiv \mathcal{I}(\beta_0)$ and $\mathcal{I}(\beta) \equiv E[D(x, \beta)'V(x, \beta)^{-1}D(x, \beta)]$ with the conditional Jacobian matrix $D(x, \beta) \equiv E[\partial u(z, \beta)/\partial \beta' | x]$ and conditional second moment matrix $V(x, \beta) \equiv E[u(z, \beta)u(z, \beta)' | x]$. An optimal GMM or GEL estimator based on the unconditional moment restrictions (2.2), therefore, would require the infeasible matrix of instrumental variables $v_*(x, \beta) \equiv D(x, \beta)'V(x, \beta)^{-1}$.

Like KTA, this paper develops estimators for β_0 which achieve the semi-parametric efficiency bound \mathcal{I}^{-1} but which avoid explicit estimation of the conditional Jacobian and conditional variance matrices, $D(x, \beta_0)$ and $V(x, \beta_0)$.

3 Estimators

The principal concern of this paper then is estimators which achieve the semi-parametric efficiency bound \mathcal{I}^{-1} under (2.1). We consider a local version of the GEL criterion suggested in Smith (1997, 2001) and more recently reconsidered in Newey and Smith (2004); see also Brown and Newey (1992, 2002). In particular, we are interested in the first order large sample properties of the estimator for β_0 which results from optimising a local GEL criterion. We term the resultant estimator a local GEL estimator for β_0 .

Let $u_i(\beta) \equiv u(z_i, \beta)$, ($i = 1, \dots, n$). Also let $\rho(v)$ be a function of a scalar v that is concave on its domain, an open interval \mathcal{V} containing zero. Define the positive weights

$w_{ij} \equiv \mathcal{K}_{ij} / \sum_{k=1}^n \mathcal{K}_{ik}$ where $\mathcal{K}_{ij} \equiv \mathcal{K}(\frac{x_i - x_j}{b_n})$, $\mathcal{K}(\cdot)$ is a symmetric positive kernel and b_n a bandwidth parameter, the properties of which will be described later. Note that $\sum_{j=1}^n w_{ij} = 1$. We consider a recentred local GEL criterion, cf. NS, given by

$$\hat{P}(\beta, \lambda) = \sum_{i=1}^n T_{i,n} \sum_{j=1}^n w_{ij} [\rho(\lambda'_i u_j(\beta)) - \rho(0)] / n, \quad (3.1)$$

where $\lambda = (\lambda'_1, \dots, \lambda'_n)'$. The sequence of trimming functions $T_{i,n}$ is required to bound the denominator of the weights w_{ij} away from zero and are defined as in KTA; that is, $T_{i,n} \equiv I\{\hat{h}(x_i) \geq b_n^\tau\}$ for some $\tau \in (0, 1)$ where $\hat{h}(x_i) \equiv \sum_{j=1}^n \mathcal{K}(\frac{x_i - x_j}{b_n}) / nb_n^s$ is the standard kernel estimator for the density $h(\cdot)$ of x at $x = x_i$ and $I\{\cdot\}$ is an indicator function. The local GEL criterion $\hat{P}(\beta, \lambda)$ (3.1) employs the Nadaraya-Watson estimator $\sum_{j=1}^n w_{ij} \rho(\lambda'_i u_j(\beta))$ of the conditional expectation of $\rho(\lambda'_i u_i(\beta))$ given x_i , i.e. $E[\rho(\lambda'_i u_i(\beta)) | x_i]$, ($i = 1, \dots, n$). Hence, we may consider $\hat{P}(\beta, \lambda)$ to be an estimator of the centred average conditional expectation $\sum_{i=1}^n (E[\rho(\lambda'_i u_i(\beta)) | x_i] / n - \rho(0)) / n$.

Let $\Lambda_n = \{\lambda \in \mathcal{R}^q : \|\lambda\| \leq Cn^{-1/m}\}$ for some positive integer m and finite constant $C > 0$. The local GEL estimator then is the solution to a saddle point problem

$$\hat{\beta} = \arg \inf_{\beta \in \mathcal{B}} \sum_{i=1}^n T_{i,n} \sup_{\lambda_i \in \Lambda_n} \hat{P}_i(\beta, \lambda_i) / n, \quad (3.2)$$

where \mathcal{B} denotes the parameter space and $\hat{P}_i(\beta, \lambda_i) \equiv \sum_{j=1}^n w_{ij} [\rho(\lambda'_i u_j(\beta)) - \rho(0)]$, ($i = 1, \dots, n$). Note that the recentring term $\rho(0) \sum_{i=1}^n T_{i,n} / n$ ensures that $\sup_{\lambda_i \in \Lambda_n} \hat{P}_i(\beta, \lambda_i) \geq 0$ as $\hat{P}_i(\beta, 0) = 0$ which in turn ensures that $\sum_{i=1}^n T_{i,n} \sup_{\lambda_i \in \Lambda_n} \hat{P}_i(\hat{\beta}, \lambda_i)$ is a suitable candidate statistic for hypothesis testing.

It will be convenient to impose a normalization on $\rho(v)$ as in NS. Let $\rho_j(v) \equiv \partial^j \rho(v) / \partial v^j$ and $\rho_j \equiv \rho_j(0)$. We normalize so that $\rho_1 = \rho_2 = -1$.¹

Specialisation of the function $\rho(\cdot)$ provides a number of interesting cases. The local empirical likelihood (EL) estimator suggested by KTA results when $\rho(v) = \log(1 - v)$ and $\mathcal{V} = (-\infty, 1)$; cf. Imbens (1997), Qin and Lawless (1994), NS and Smith (1997). A local exponential tilting (ET) estimator is obtained with $\rho(v) = -\exp(v)$, cf. Imbens, Spady and Johnson (1998), Kitamura and Stutzer (1997), NS and Smith (1997).

¹We will require $\rho_1 \neq 0$ and $\rho_2 < 0$. This normalization can always be imposed by replacing $\rho(v)$ by $[-\rho_2 / \rho_1^2] \rho([\rho_1 / \rho_2]v)$, which leaves the estimator of β_0 unaffected.

Let $\hat{u}_i(\beta) \equiv \sum_{j=1}^n w_{ij}u_j(\beta)$ and $\hat{V}(x_i, \beta) \equiv \sum_{j=1}^n w_{ij}u_j(\beta)u_j(\beta)'$, the Nadaraya-Watson estimators of $E[u_i(\beta)|x_i]$ and $E[u_i(\beta)u_i(\beta)'|x_i]$ respectively. A local version of the continuous updating estimator (CUE) of Hansen, Heaton, and Yaron (1996), cf. Bonnal and Renault (2003) and Smith (2003), is readily seen to be a local GEL estimator when $\rho(v)$ is quadratic; cf. NS, Theorem 2.1, which demonstrates an analogous result for unconditional moment restrictions. The local CUE is constructed as²

$$\hat{\beta}_{CUE} = \arg \min_{\beta \in \mathcal{B}} \sum_{i=1}^n \mathbf{T}_{i,n} \hat{u}_i(\beta)' \hat{V}(x_i, \beta)^{-1} \hat{u}_i(\beta). \quad (3.3)$$

Theorem 3.1 *If $\rho(v)$ is quadratic, then $\hat{\beta} = \hat{\beta}_{CUE}$.*

In contradistinction to the local CUE $\hat{\beta}_{CUE}$ which simultaneously minimizes the objective function over β in $\hat{V}(x_i, \beta)$, a local GMM estimator is given by

$$\hat{\beta}_{GMM} = \arg \min_{\beta \in \mathcal{B}} \sum_{i=1}^n \mathbf{T}_{i,n} \hat{u}_i(\beta)' \hat{V}(x_i, \tilde{\beta})^{-1} \hat{u}_i(\beta), \quad (3.4)$$

where $\tilde{\beta}$ denotes an initial consistent estimator for β_0 ; see, for example, Newey (1990, 1993).

In a similar fashion to NS, we may describe alternative estimators related to the family of discrepancy measures given by Cressie and Read (1984). Recall from NS, Theorem 2.2, that the equivalent unconditional GEL criterion to the Cressie-Read discrepancy criterion is given by $\rho(v) = -(1 + \gamma v)^{(\gamma+1)/\gamma}/(\gamma + 1)$, with EL, ET and CUE as special cases obtained by setting $\gamma = -1$, $\gamma = 0$ and $\gamma = 1$ respectively. A local Cressie-Read discrepancy criterion is therefore given by

$$\hat{P}(\beta, \lambda) = - \sum_{i=1}^n \mathbf{T}_{i,n} \left[\sum_{j=1}^n w_{ij} (1 + \gamma \lambda'_i u_j(\beta))^{(\gamma+1)/\gamma} / (\gamma + 1) - 1 \right] / n;$$

cf. Bonnal and Renault (2003) and Smith (2003).

²An alternative local CUE more in the spirit of Hansen, Heaton, and Yaron (1996) would minimize $\sum_{i=1}^n \mathbf{T}_{i,n} \hat{u}_i(\beta)' [\sum_{j=1}^n w_{ij} u_j(\beta) u_j(\beta)' - \hat{u}_i(\beta) \hat{u}_i(\beta)']^{-1} \hat{u}_i(\beta)$; see Bonnal and Renault (2003) and Smith (2003). In contrast to the unconditional moment case, see NS, fn.1, the resultant CUE does not coincide with $\hat{\beta}_{CUE}$.

3.1 Empirical Probabilities

We may also define empirical conditional probabilities for the observations for each member of the GEL class. Let $\hat{u}_i \equiv u_i(\hat{\beta})$ where $\hat{\beta}$ denotes a GEL estimator. Also let $\hat{\lambda}_i(\beta) \equiv \arg \sup_{\lambda_i \in \Lambda_n} \sum_{j=1}^n w_{ij} \rho(\lambda_i' u_j(\beta))$ and $\hat{\lambda}_i \equiv \hat{\lambda}_i(\hat{\beta})$, ($i = 1, \dots, n$). For a given function $\rho(v)$, the empirical conditional probabilities are defined by

$$\hat{\pi}_{ij} \equiv w_{ij} \rho_1(\hat{\lambda}_i' \hat{u}_j) / \sum_{k=1}^n w_{ik} \rho_1(\hat{\lambda}_i' \hat{u}_k), (j = 1, \dots, n), \quad (3.5)$$

The empirical probabilities $\hat{\pi}_{ij}$, ($j = 1, \dots, n; i = 1, \dots, n$), sum to one by construction over $j = 1, \dots, n$, satisfy the sample moment condition $\sum_{i=1}^n \hat{\pi}_{ij} \hat{u}_j = 0$ when the first order conditions for $\hat{\lambda}_i$ hold, and are positive when $\hat{\lambda}_i' \hat{u}_j$ is small uniformly in j ; see Lemma B.1 in Appendix B.

For unconditional moment restrictions the (unconditional) probabilities are $\hat{\pi}_i = \rho_1(\hat{\lambda}' \hat{g}_i) / \sum_{k=1}^n \rho_1(\hat{\lambda}' \hat{g}_k)$, ($i = 1, \dots, n$), see NS, equation (2.4), where $g(z, \beta) \equiv v(x, \beta) u(z, \beta)$ from (2.2), $\hat{g}_i \equiv g(z_i, \hat{\beta})$ and $\hat{\beta}$ and $\hat{\lambda}$ denote an unconditional GMM or GEL estimator and associated auxiliary parameter estimator respectively. In contrast, the empirical conditional probabilities $\hat{\pi}_{ij}$ employ the differential data-determined kernel weights w_{ij} , ($j = 1, \dots, n$), rather than equal weights $1/n$ resulting from the unconditional empirical distribution function.

For EL, see KTA, $\hat{\pi}_{ij} = w_{ij} / (1 - \hat{\lambda}_i' \hat{u}_j)$, cf. Owen (1988), for ET, $\hat{\pi}_{ij} = w_{ij} \exp(\hat{\lambda}_i' \hat{u}_j) / \sum_{k=1}^n w_{ik} \exp(\hat{\lambda}_i' \hat{u}_k)$, cf. Kitamura and Stutzer (1997), and for quadratic $\rho(v)$ or CUE, see Bonnal and Renault (2003), $\hat{\pi}_{ij} = w_{ij} (1 + \hat{\lambda}_i' \hat{u}_j) / \sum_{k=1}^n w_{ik} (1 + \hat{\lambda}_i' \hat{u}_k)$, cf. Back and Brown (1993) and Smith (2003). See also Brown and Newey (1992, 2002) and Smith (1997).

3.2 First Order Conditions

Like NS for the unconditional moment restrictions case, a re-interpretation of the first order conditions determining the local GEL estimator $\hat{\beta}$ is useful for gaining an intuitive understanding of the reason why $\hat{\beta}$ achieves the semi-parametric efficiency lower bound \mathcal{I}^{-1} .

Let $U_j(\beta) \equiv \partial u_j(\beta)/\partial \beta'$, ($j = 1, \dots, n$), and $\hat{D}(x_i, \beta) \equiv \sum_{j=1}^n w_{ij} U_j(\beta)$ the Nadaraya-Watson estimator of $E[U_i(\beta)|x_i]$. Initially, consider the first order conditions for the semi-parametric efficient two-step GMM estimator $\hat{\beta}_{GMM}$ defined above in (3.4); that is,

$$\sum_{i=1}^n \hat{D}(x_i, \hat{\beta}_{GMM}) \hat{V}(x_i, \tilde{\beta})^{-1} \hat{u}_i(\hat{\beta}_{GMM}) = 0, \quad (3.6)$$

which employs an explicit estimator for the efficient matrix of instrumental variables $v_*(x, \beta_0) = D(x, \beta_0)' V(x, \beta_0)^{-1}$. An analogous expression may also be provided for any GEL estimator $\hat{\beta}$ which mimics that given in NS, Theorem 2.3, for the unconditional moment restrictions case. Let $k(v) = [\rho_1(v) + 1]/v$, $v \neq 0$ and $k(0) = -1$. Also, let $\hat{v}_{ij} \equiv \hat{\lambda}'_i \hat{u}_j$.

Theorem 3.2 *The local GEL first order conditions for $\hat{\beta}$ imply*

$$\sum_{i=1}^n \mathbb{T}_{i,n} \left[\sum_{j=1}^n w_{ij} \rho_1(\hat{v}_{ij}) U_j(\hat{\beta}) \right]' \left[\sum_{j=1}^n w_{ij} k(\hat{v}_{ij}) u_j(\hat{\beta}) u_j(\hat{\beta})' \right]^{-1} \hat{u}_i(\hat{\beta}) = 0, \quad (3.7)$$

where $k(\hat{v}_{ij}) = -1/(1 - \hat{v}_{ij})$ for local EL and $k(\hat{v}_{ij}) = -1$ for local CUE.

See also Bonnal and Renault (2003) and Smith (2003) for analogous results for local CUE and efficient information theoretic estimators respectively.

A comparison of the first order conditions determining the semi-parametric efficient infeasible GMM estimator, (3.6), and those for local GEL, (3.7), is instructive. Let $\hat{k}_{ij} \equiv w_{ij} k(\hat{v}_{ij}) / \sum_{k=1}^n w_{ik} k(\hat{v}_{ik})$ and $\hat{\pi}_{ij} \equiv w_{ij} \rho_1(\hat{v}_{ij}) / \sum_{k=1}^n w_{ik} \rho_1(\hat{v}_{ik})$ as in (3.5), ($i, j = 1, \dots, n$). Similarly to $\hat{\pi}_{ij}$ (3.5), we may also interpret \hat{k}_{ij} as an empirical conditional probability. Now, Lemma B.1 of Appendix B shows that $\max_{1 \leq j \leq n} \sup_{\lambda_i \in \Lambda_n, \beta \in \mathcal{B}} \lambda'_i g_j(\beta) = o_p(1)$. Therefore, the implicit estimators for the conditional Jacobian and conditional variance matrices in (3.7) are consistent, i.e., $\sum_{j=1}^n \hat{\pi}_{ij} U_j(\hat{\beta}) \xrightarrow{P} D(x_i, \beta_0)$ and $\sum_{j=1}^n \hat{k}_{ij} u_j(\hat{\beta}) u_j(\hat{\beta})' \xrightarrow{P} V(x_i, \beta_0)$ with $\sum_{k=1}^n w_{ik} k(\hat{v}_{ik}) / \sum_{k=1}^n w_{ik} \rho_1(\hat{v}_{ik}) \xrightarrow{P} 1$. Comparing the GMM and GEL first order conditions, (3.6) and (3.7), we see straightforwardly that, asymptotically, local GEL estimators implicitly employ the semi-parametric efficient matrix of instrumental variables and thereby achieve the semi-parametric efficiency lower bound \mathcal{I}^{-1} .

It is also interesting to note that the local CUE uses the Nadaraya-Watson kernel regression estimator $\hat{V}(x_i, \hat{\beta})$ for the conditional variance matrix $V(x_i, \beta_0)$ whereas local EL employs the same weights for the estimation of $V(x_i, \beta_0)$ as for the conditional Jacobian matrix $D(x_i, \beta_0)$, that is, the empirical probabilities $\hat{\pi}_{ij} = 1/(1 - \hat{v}_{ij})$. The two-step semi-parametric efficient GMM estimator $\hat{\beta}_{GMM}$ described in (3.4) utilises Nadaraya-Watson regression estimators for both conditional Jacobian and variance matrices.

4 Asymptotic Theory for Local GEL

This section gives consistency and asymptotic normality results for the local GEL estimator $\hat{\beta}$.

We firstly, however, require some additional notation. Let $h(x)$ denote the density function of x . Elements of vectors and matrices are denoted by superscripts (i) and (ij) respectively.

Next, we provide some regularity conditions. Our assumptions are virtually identical to KTA, Assumptions 3.1-3.7. For a full discussion of these assumptions, see KTA, section 3.

Assumption 4.1 *For each $\beta \neq \beta_0$ there exists a set $\mathcal{X}_\beta \subseteq \mathcal{R}^s$ such that $\mathcal{P}\{x \in \mathcal{X}_\beta\} > 0$ and $E[u(z, \beta)|x] \neq 0$ for all $x \in \mathcal{X}_\beta$.*

This is the conditional identification condition given in KTA, Assumption 3.1. Together with (2.1) it crucially ensures that $E[\|E[u(z, \beta)|x]\|^2] = 0$ if and only if $\beta = \beta_0$.

Assumption 4.2 **(i)** $\rho(v)$ is twice continuously differentiable and concave on its domain, an open interval \mathcal{V} containing 0 and $\rho_1 = \rho_2 = -1$; **(ii)** $E[\sup_{\beta \in \mathcal{B}} \|u(z, \beta)\|^m] < \infty$ for some $m > 8$.

Assumption 4.2 (i) is the condition on $\rho(v)$ adapted from NS, Assumption 1 (f).

Assumption 4.3 *The kernel $\mathcal{K}(x) = \prod_{k=1}^s \kappa(x^{(k)})$, $x = (x^{(1)}, \dots, x^{(s)})'$, where $\kappa : \mathcal{R} \rightarrow \mathcal{R}$, is a continuously differentiable p.d.f. with support $[-1, 1]$, symmetric about 0 and bounded away from 0 on $[-a, a]$ for some $a \in (0, 1)$.*

Assumption 4.4 (i) $0 < h(x) \leq \sup_{x \in \mathcal{R}^s} h(x) < \infty$, $h(x)$ is twice continuously differentiable on \mathcal{R}^s , $\sup_{x \in \mathcal{R}^s} \|\partial h(x)/\partial x\| < \infty$ and $\sup_{x \in \mathcal{R}^s} \|\partial^2 h(x)/\partial x \partial x'\| < \infty$; (ii) $E[\|x\|^{\rho+1}] < \infty$ for some $\rho > 0$; (iii) $u(z, \beta)$ is continuous on \mathcal{B} w.p.1 and $E[\sup_{\beta \in \mathcal{B}} \|\partial u(z, \beta)/\partial \beta'\|] < \infty$; (iv) $\|\partial^2[E[u^{(i)}(z, \beta)|x]h(x)]/\partial x \partial x'\|$ is uniformly bounded on $\mathcal{B} \times \mathcal{R}^s$, ($i = 1, \dots, q$).

Let $\mathcal{S}^q \equiv \{\xi : \xi \in \mathcal{R}^q, \|\xi\| = 1\}$ be the unit sphere in \mathcal{R}^q .

Assumption 4.5 There exists a non-empty neighbourhood \mathcal{B}_0 of β_0 such that (i) $D(x, \beta)$ and $V(x, \beta)$ are continuous on \mathcal{B}_0 w.p.1; (ii) $\inf_{(\xi, x, \beta) \in \mathcal{S}^q \times \mathcal{R}^s \times \mathcal{B}_0} \xi' V(x, \beta) \xi > 0$ and $\sup_{(\xi, x, \beta) \in \mathcal{S}^q \times \mathcal{R}^s \times \mathcal{B}_0} \xi' V(x, \beta) \xi > 0$; (iii) $\sup_{\beta \in \mathcal{B}_0} |\partial u^{(i)}(z, \beta)/\partial \beta^{(j)}| < c(z)$ and $\sup_{\beta \in \mathcal{B}_0} |\partial^2 u^{(i)}(z, \beta)/\partial \beta^{(j)} \partial \beta^{(k)}| < d(z)$ w.p.1 for some functions $c(z)$ and $d(z)$ such that $E[c(z)^\eta] < \infty$ for some $\eta > 4$ and $E[d(z)] < \infty$; (iv) $\sup_{x \in \mathcal{R}^s} \|\partial E[D^{(ij)}(x, \beta_0)h(x)]/\partial x\| < \infty$ and $\sup_{(x, \beta) \in \mathcal{R}^s \times \mathcal{B}_0} \|\partial^2 E[D^{(ij)}(x, \beta_0)h(x)]/\partial x \partial x'\| < \infty$; (v) $\sup_{x \in \mathcal{R}^s} \|\partial E[V^{(ij)}(x, \beta_0)h(x)]/\partial x\| < \infty$ and $\sup_{(x, \beta) \in \mathcal{R}^s \times \mathcal{B}_0} \|\partial^2 E[V^{(ij)}(x, \beta_0)h(x)]/\partial x \partial x'\| < \infty$.

Assumption 4.6 The parameters λ_i , ($i = 1, \dots, n$), are constrained to lie in the set $\Lambda_n = \{\lambda_i : \|\lambda_i\| \leq Dn^{-1/m}\}$ for some $D > 0$.

Assumption 4.7 Let $\tau \in (0, 1)$, $\rho \geq \max(1/\eta + 1/2, 2/m + 1/2)$, $b_n \downarrow 0$ and $\sigma \in (0, 1/2)$. Then $n^{1-2\sigma-2/m} b_n^{2s+4\tau} \uparrow \infty$, $n^\rho b_n^{2\tau} \uparrow \infty$, $n^{\rho-1/\eta} b_n^\tau \uparrow \infty$, $n^{\rho-2/m} b_n^\tau \uparrow \infty$, $n^{1-2\sigma} b_n^{5s/2+6\tau} \uparrow \infty$, $n^{2\rho-1/\eta-1/m-1/2} b_n^{2\tau} \uparrow \infty$ and $n^{2\rho-3/m-1/2} b_n^{3\tau} \uparrow \infty$.

As noted by KTA, the presence of the parameter σ is required for the uniform convergence result for kernel estimators given in Ai (1997, Lemma B.1, p.955) which is central to the proofs of many of the subsidiary results presented in KTA, Appendix B.

These conditions lead to a consistency result.

Theorem 4.1 Let Assumptions 4.1-4.5 and 4.7 hold. Then $\hat{\beta} \xrightarrow{p} \beta_0$.

Asymptotic normality of the local GEL estimator $\hat{\beta}$ requires the additional regularity condition Assumption 4.6.

Theorem 4.2 *If Assumptions 4.1-4.7 are satisfied, then $n^{1/2}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, \mathcal{I}^{-1})$.*

Theorem 4.2 emphasises that all local GEL estimators $\hat{\beta}$ are first order equivalent and achieve the semi-parametric efficiency lower bound \mathcal{I}^{-1} . Lemma B.3 below provides a basis for the estimation of the asymptotic variance matrix \mathcal{I}^{-1} of $n^{1/2}(\hat{\beta} - \beta_0)$.

5 Hypothesis Tests

Consider the following null hypothesis which incorporates the parametric restrictions $r(\beta) = 0$:

$$H_0 : r(\beta_0) = 0, \quad (5.1)$$

where $r(\cdot)$ is an r -vector of twice continuously differentiable functions of β where $p > r$. The alternative hypothesis is defined by $H_1 : r(\beta_0) \neq 0$.

A standard Wald statistic based on the local GEL estimator could be used to test $H_0 : r(\beta_0) = 0$. This form of statistic like others requires the consistent estimation of the semi-parametric information matrix \mathcal{I} , cf. Lemma B.3 below. Estimators of \mathcal{I} may be unreliable with the samples typically available in applications. Unlike the Wald statistic the statistic described here is based on the local GEL criterion function (3.1) and therefore does not require estimation of the conditional Jacobian and variance matrices, $D(x, \beta_0)$ and $V(x, \beta_0)$, which are required for explicit estimation of \mathcal{I} .

The local GEL criterion function statistic LR_n^{GEL} is then defined as

$$LR_n^{GEL} = 2n[\hat{P}(\hat{\beta}, \hat{\lambda}(\hat{\beta})) - \hat{P}(\tilde{\beta}, \hat{\lambda}(\tilde{\beta}))], \quad (5.2)$$

where the restricted local GEL estimator $\tilde{\beta} = \arg \inf_{\beta \in \mathcal{B}^r} \sum_{i=1}^n T_{i,n} \sup_{\lambda_i \in \Lambda_n} \hat{P}_i(\beta, \lambda_i)/n$ with $\mathcal{B}^r = \{\beta : r(\beta) = 0, \beta \in \mathcal{B}\}$.

Similarly to KTA, section 4, to motivate the use of the statistic LR_n^{GEL} (5.2) consider the situation which arises when the null hypothesis is simple, that is, $H_0^c : \beta_0 = \beta^c$ where

β^c is known. By a Taylor expansion about $\hat{\beta}$, as $\partial\hat{P}(\hat{\beta}, \lambda(\hat{\beta}))/\partial\beta = 0$,

$$\begin{aligned} LR_n^{GEL} &= 2n[\hat{P}(\hat{\beta}, \hat{\lambda}(\hat{\beta})) - \hat{P}(\beta^c, \hat{\lambda}(\beta^c))] \\ &= -n(\hat{\beta} - \beta^c)' \frac{\partial^2 \hat{P}(\beta^*, \hat{\lambda}(\beta^*))}{\partial\beta\partial\beta'} (\hat{\beta} - \beta^c). \end{aligned}$$

Now $\partial^2 \hat{P}(\beta^*, \hat{\lambda}(\beta^*)) / \partial\beta\partial\beta' = -\mathcal{I} + o_p(1)$ by Lemma B.3 and $n^{1/2}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, \mathcal{I}^{-1})$ from Theorem 4.2. Therefore, $LR_n^{GEL} \xrightarrow{d} \chi^2(p)$ under H_0^c .

To deal with the general case (5.1), we make the following assumption:

Assumption 5.1 (i) $r : \mathcal{B} \rightarrow \mathcal{R}^r$ is twice continuously differentiable; (ii) $R \equiv \partial r(\beta_0) / \partial\beta'$ is full row rank r .

The following result describes the limiting distribution of the local GEL criterion function statistic LR_n^{GEL} (5.2).

Theorem 5.1 *Let Assumptions 4.1-4.7 and 5.1 be satisfied. Then $LR_n^{GEL} \xrightarrow{d} \chi^2(r)$ under $H_0 : r(\beta_0) = 0$.*

A test with given asymptotic size is obtained by comparing the statistic LR_n^{GEL} to an appropriate critical value from the chi-square distribution with r degrees of freedom. Valid asymptotic confidence regions for β_0 may be constructed by inversion of LR_n^{GEL} . In particular, a $(1-\alpha)$ confidence region is $\beta_0 \in \{\beta : 2n[\hat{P}(\hat{\beta}, \hat{\lambda}(\hat{\beta})) - \hat{P}(\beta, \hat{\lambda}(\beta))] \leq \chi_{1-\alpha}^2(p)\}$ where $\chi_{1-\alpha}^2(\cdot)$ is the $100(1-\alpha)$ -percentile from the $\chi^2(\cdot)$ distribution. Like KTA, a $(1-\alpha)$ confidence interval for a single parameter, $\beta_0^{(j)}$ say, is given by $\beta_0^{(j)} \in \{\beta : \min_{\beta^{(i)}, i \neq j} 2n[\hat{P}(\hat{\beta}, \hat{\lambda}(\hat{\beta})) - \hat{P}(\beta, \hat{\lambda}(\beta))] \leq \chi_{1-\alpha}^2(1)\}$.

Appendix A: Proofs of Results

Throughout these Appendices, C will denote a generic positive constant that may be different in different uses, and CS and T the Cauchy-Schwarz and triangle inequalities respectively. Also, with probability approaching one will be abbreviated as w.p.a.1, UWL will denote a uniform weak law of large numbers such as Lemma 2.4 of Newey and McFadden (1994), and CLT will refer to the Lindeberg-Lévy central limit theorem.

Let $\hat{u}_i(\beta) \equiv \sum_{j=1}^n w_{ij}u_j(\beta)$, $\hat{D}(x_i, \beta) \equiv \sum_{j=1}^n w_{ij}U_j(\beta)$ and $\hat{V}(x_i, \beta) \equiv \sum_{j=1}^n w_{ij}u_j(\beta)u_j(\beta)$. Also let $\hat{\lambda}_{i0} \equiv \hat{\lambda}_i(\beta_0)$, ($i = 1, \dots, n$), $\hat{\lambda}_0 = (\hat{\lambda}'_{10}, \dots, \hat{\lambda}'_{n0})'$ and $u_{i0} \equiv u_i(\beta_0)$, $U_{i0} \equiv U_i(\beta_0)$, ($i = 1, \dots, n$).

Proof of Theorem 3.1: The proof is very similar to that for NS, Theorem 2.1. By $\rho(v)$ quadratic, a second order Taylor expansion is exact, giving

$$\hat{P}(\beta, \lambda) = - \sum_{i=1}^n \mathbf{T}_{i,n} \hat{u}_i(\beta)' \lambda_i - \frac{1}{2} \sum_{i=1}^n \mathbf{T}_{i,n} \lambda_i' \hat{V}(x_i, \beta) \lambda_i.$$

By concavity of $\hat{P}_i(\beta, \lambda_i) = -\mathbf{T}_{i,n} \hat{u}_i(\beta)' \lambda_i - \mathbf{T}_{i,n} \lambda_i' \hat{V}(x_i, \beta) \lambda_i / 2$ in λ_i , any solution $\hat{\lambda}_i(\beta)$ to the first order conditions

$$0 = -\mathbf{T}_{i,n} \hat{u}_i(\beta) - \mathbf{T}_{i,n} \hat{V}(x_i, \beta) \lambda_i$$

will maximize $\hat{P}_i(\beta, \lambda_i)$ with respect to λ_i holding β fixed. Then, $\hat{\lambda}_i(\beta) = -\hat{V}(x_i, \beta)^{-1} \hat{u}_i(\beta)$ solves the first order conditions. Since

$$\hat{P}_i(\beta, \hat{\lambda}_i(\beta)) = \frac{1}{2} \mathbf{T}_{i,n} \hat{u}_i(\beta)' \hat{V}(x_i, \beta)^{-1} \hat{u}_i(\beta),$$

the GEL objective function $\hat{P}(\beta, \hat{\lambda}(\beta))$ is a monotonic increasing function of the CUE objective function. ■

Proof of Theorem 3.2: Let $\hat{u}_i \equiv u_i(\hat{\beta})$ and $\hat{U}_i \equiv U_i(\hat{\beta})$. The first order conditions for $\hat{\lambda}_i \equiv \arg \sup_{\lambda_i \in \Lambda_n} \sum_{j=1}^n w_{ij} \rho(\lambda_i' \hat{u}_j)$ are $\mathbf{T}_{i,n} \sum_{j=1}^n w_{ij} \rho_1(\lambda_i' \hat{u}_j(\hat{\beta})) u_j(\hat{\beta}) = 0$. By the implicit function theorem there is a neighborhood of $\hat{\beta}$ where the solution $\hat{\lambda}_i(\beta)$ to

$\mathbb{T}_{i,n} \sum_{j=1}^n w_{ij} \rho_1(\hat{\lambda}'_i u_j(\beta)) u_j(\beta) = 0$ exists and is continuously differentiable. Then by the envelope theorem the first order conditions for GEL are

$$\begin{aligned} \sum_{i=1}^n \mathbb{T}_{i,n} \sum_{j=1}^n w_{ij} \rho_1(\hat{\lambda}'_i u_j(\hat{\beta})) U_j(\hat{\beta})' \hat{\lambda}_i &= 0, \\ \mathbb{T}_{i,n} \sum_{j=1}^n w_{ij} \rho_1(\hat{\lambda}'_i u_j(\hat{\beta})) u_j(\hat{\beta}) &= 0, (i = 1, \dots, n). \end{aligned} \quad (\text{A.1})$$

By eq. (A.1) and the definition of $k(v)$,

$$\begin{aligned} 0 &= \mathbb{T}_{i,n} \sum_{j=1}^n \rho_1(\hat{v}_{ij}) w_{ij} \hat{u}_j = \mathbb{T}_{i,n} \left[\sum_{j=1}^n (\rho_1(\hat{v}_{ij}) + 1) w_{ij} \hat{u}_j - \sum_{j=1}^n w_{ij} \hat{u}_j \right] \\ &= \mathbb{T}_{i,n} \left[\sum_{j=1}^n k(\hat{v}_{ij}) w_{ij} \hat{u}_j \hat{u}_j' \hat{\lambda}_i - \sum_{j=1}^n w_{ij} \hat{u}_j \right], (i = 1, \dots, n). \end{aligned}$$

Plugging the solutions $\mathbb{T}_{i,n} \hat{\lambda}_i = \mathbb{T}_{i,n} [\sum_{j=1}^n k(\hat{v}_{ij}) w_{ij} \hat{u}_j \hat{u}_j']^{-1} \sum_{j=1}^n w_{ij} \hat{u}_j$, ($i = 1, \dots, n$), into the first part of eq. (A.1) $\sum_{i=1}^n \mathbb{T}_{i,n} \sum_{j=1}^n \rho_1(\hat{v}_{ij}) w_{ij} \hat{U}_j' \hat{\lambda}_i = 0$ gives the first result. Note that for EL $k(v) = [-(1-v)^{-1} + 1]/v = -(1-v)^{-1} = \rho_1(v)$ and for CUE $k(v) = [-(1+v) + 1]/v = -1$. ■

Proof of Theorem 4.1: The structure of the proof closely resembles that of KTA, Proof of Theorem 3.1. Let $c > 0$ such that $(-c, c) \in \mathcal{V}$. Define $C_n = \{z \in \mathcal{R}^d : \sup_{\beta \in \mathcal{B}} \|u(z, \beta)\| \leq cn^{1/m}\}$ and $u_{nj}(\beta) = I_j u_j(\beta)$, where $I_j = I\{z_j \in C_n\}$. Let $\bar{\lambda}_i(\beta) = -E[u_i(\beta)|x_i]/(1 + \|E[u_i(\beta)|x_i]\|)$. Then,

$$\begin{aligned} \sup_{\{\lambda_i \in \Lambda_n\}_{i=1}^n} \hat{P}(\beta, \lambda) &\geq Q_n(\beta) \\ &= \sum_{i=1}^n \mathbb{T}_{i,n} \sum_{j=1}^n w_{ij} [\rho(n^{-1/m} \bar{\lambda}_i(\beta)' u_{nj}(\beta)) - \rho(0)]/n. \end{aligned} \quad (\text{A.2})$$

Note that $n^{-1/m} \bar{\lambda}_i(\beta) \in \Lambda_n$.

Now

$$\rho(n^{-1/m} \bar{\lambda}_i(\beta)' u_{nj}(\beta)) = \rho(0) - n^{-1/m} \bar{\lambda}_i(\beta)' u_j(\beta) + r_{nij}(t),$$

for some $t \in (0, 1)$ and

$$\begin{aligned} r_{nij}(t) &= n^{-1/m} \bar{\lambda}_i(\beta)' u_j(\beta) (1 - I_j) \\ &\quad + n^{-1/m} \bar{\lambda}_i(\beta)' u_{nj}(\beta) [\rho_1(tn^{-1/m} \bar{\lambda}_i(\beta)' u_{nj}(\beta)) - \rho_1(0)]. \end{aligned} \quad (\text{A.3})$$

[A.2]

From Lemma B.1 $\sup_{\beta \in \mathcal{B}, n^{-1/m} \lambda_i \in \Lambda_n, 1 \leq j \leq n} |\rho_1(n^{-1/m} \lambda_i' u_j(\beta)) - \rho_1(0)| \xrightarrow{p} 0$. Also $\max_{1 \leq j \leq n} (1 - I_j) = o_p(1)$. Hence, from eq. (A.3),

$$\begin{aligned} n^{1/m} \sum_{i=1}^n \mathbb{T}_{i,n} \sum_{j=1}^n w_{ij} r_{nij}(t)/n &= o_p(1) \sum_{i=1}^n \mathbb{T}_{i,n} \bar{\lambda}_i(\beta)' \hat{u}_i(\beta)/n + o_p(1) \sum_{i=1}^n \mathbb{T}_{i,n} \bar{\lambda}_i(\beta)' \hat{u}_i(\beta)/n \\ &\quad - o_p(1) \sum_{i=1}^n \mathbb{T}_{i,n} \bar{\lambda}_i(\beta)' \sum_{j=1}^n w_{ij} u_j(\beta) (1 - I_j)/n \\ &= o_p(1) \sum_{i=1}^n \mathbb{T}_{i,n} \bar{\lambda}_i(\beta)' \hat{u}_i(\beta)/n \end{aligned}$$

uniformly $\beta \in \mathcal{B}$. Thus,

$$\begin{aligned} n^{1/m} \sup_{\beta \in \mathcal{B}} \left| \sum_{i=1}^n \mathbb{T}_{i,n} \sum_{j=1}^n w_{ij} r_{nij}(t)/n \right| &\leq o_p(1) \sum_{i=1}^n \mathbb{T}_{i,n} \sup_{\beta \in \mathcal{B}} \|\hat{u}_i(\beta)\|/n \\ &= o_p(1) O_p(1) = o_p(1) \end{aligned}$$

as $\max_{1 \leq i \leq n} \sup_{\beta \in \mathcal{B}} \mathbb{T}_{i,n} \|\hat{u}_i(\beta) - E[u_i(\beta)|x_i]\| = o_p(1)$; cf. KTA, Proof of Lemma B.8.

Therefore, substituting into eq. (A.2), $n^{1/m} Q_n(\beta) = - \sum_{i=1}^n \mathbb{T}_{i,n} \sum_{j=1}^n w_{ij} \bar{\lambda}_i(\beta)' u_j(\beta) + o_p(1)$ uniformly $\beta \in \mathcal{B}$ and, from KTA, Lemma B.8,

$$n^{1/m} \sup_{\beta \in \mathcal{B}} |Q_n(\beta) - \bar{Q}_n(\beta)| = o_p(1), \quad (\text{A.4})$$

as $\sum_{i=1}^n (\mathbb{T}_{i,n} - 1)/n = o_p(1)$, see KTA, Proof of Theorem 3.1, where

$$n^{1/m} \bar{Q}_n(\beta) = - \sum_{i=1}^n \bar{\lambda}_i(\beta)' E[u_i(\beta)|x_i]/n; \quad (\text{A.5})$$

see KTA, eqs. (A.4) and (A.5). Thus, as in KTA, eq. (A.6), from eqs. (A.2) and (A.4),

$$n^{1/m} \inf_{\beta \in \mathcal{B}} \sup_{\{\lambda_i \in \Lambda_n\}_{i=1}^n} \hat{P}(\beta, \lambda) \geq n^{1/m} \inf_{\beta \in \mathcal{B}} \bar{Q}(\beta) + o_p(1). \quad (\text{A.6})$$

From the definition of $\bar{\lambda}_i(\beta)$, ($i = 1, \dots, n$), and (A.5), a UWL gives

$$n^{1/m} \bar{Q}_n(\beta) = E[\|E[u_i(\beta)|x_i]\|^2 / (1 + \|E[u_i(\beta)|x_i]\|)] + o_p(1), \quad (\text{A.7})$$

uniformly $\beta \in \mathcal{B}$; see KTA, eq. (A.7). The function $E[\|E[u_i(\beta)|x_i]\|^2 / (1 + \|E[u_i(\beta)|x_i]\|)] = E[I\{x_i \in \mathcal{X}_\beta\} \|E[u_i(\beta)|x_i]\|^2 / (1 + \|E[u_i(\beta)|x_i]\|)]$ is continuous in β , has a unique zero β_0 and is strictly positive for all $\beta \neq \beta_0$ by Assumption 4.1.

[A.3]

Now

$$0 \leq n^{1/m} \sup_{\{\lambda_i \in \Lambda_n\}_{i=1}^n} \hat{P}(\hat{\beta}, \lambda) \leq n^{1/m} \sup_{\{\lambda_i \in \Lambda_n\}_{i=1}^n} \hat{P}(\beta_0, \lambda), \quad (\text{A.8})$$

as $\hat{P}(\hat{\beta}, 0) = 0$. By the concavity of $\rho(v)$, $\rho(\lambda'_{i0} u_{j0}) - \rho(0) \leq -\lambda'_{i0} u_{j0}$. Hence,

$$\begin{aligned} n^{1/m} \sup_{\{\lambda_i \in \Lambda_n\}_{i=1}^n} \hat{P}(\beta_0, \lambda) &= n^{1/m} \hat{P}(\beta_0, \hat{\lambda}_0) \\ &= n^{1/m} \sum_{i=1}^n \mathbb{T}_{i,n} \sum_{j=1}^n w_{ij} [\rho(\hat{\lambda}'_{i0} u_{j0}) - \rho(0)]/n \\ &\leq -n^{1/m} \sum_{i=1}^n \mathbb{T}_{i,n} \sum_{j=1}^n w_{ij} \hat{\lambda}'_{i0} u_{j0}/n \\ &= n^{1/m} [o_p(\sqrt{\frac{n^\sigma}{nb^{s+2\tau}}}) + o_p(\frac{1}{n^{\rho-1/m}})]^2 \\ &= o_p(1), \end{aligned}$$

by Assumption 4.7, by eq. (B.3) of Lemma B.2 and KTA, Lemma B.3. Therefore, combining (A.8) and (A.9),

$$n^{1/m} \sup_{\{\lambda_i \in \Lambda_n\}_{i=1}^n} \hat{P}(\hat{\beta}, \lambda) = o_p(1). \quad (\text{A.9})$$

By T and (A.6)

$$\begin{aligned} 0 &\leq E[-\bar{\lambda}_i(\beta)' E[u_i(\beta)|x_i]]|_{\beta=\hat{\beta}} \\ &\leq \sup_{\beta \in \mathcal{B}} \|n^{1/m} \bar{Q}_n(\beta) - E[-\bar{\lambda}_i(\beta)' E[u_i(\beta)|x_i]]\| + \|n^{1/m} \bar{Q}_n(\hat{\beta})\| \\ &\leq \left\| n^{1/m} \sup_{\{\lambda_i \in \Lambda_n\}_{i=1}^n} \hat{P}(\hat{\beta}, \lambda) \right\| \end{aligned}$$

as $\sup_{\{\lambda_i \in \Lambda_n\}_{i=1}^n} \hat{P}(\beta, \lambda) \geq \bar{Q}_n(\beta) + o_p(1)$ uniformly $\beta \in \mathcal{B}$ from eqs. (A.2) and (A.4). Hence, from eqs. (A.9) and (A.10), $E[-\bar{\lambda}_i(\beta)' E[u_i(\beta)|x_i]]|_{\beta=\hat{\beta}} = o_p(1)$. Therefore, $\hat{\beta}$ must lie in any neighbourhood of β_0 w.p.a.1, i.e. $\hat{\beta} \xrightarrow{p} \beta_0$, as $E[-\bar{\lambda}_i(\beta)' E[u_i(\beta)|x_i]]$ is continuous and has a unique zero β_0 . ■

Proof of Theorem 4.2: We consider the first order condition determining the local GEL estimator $\hat{\beta}$; viz. $\partial \hat{P}(\hat{\beta}, \hat{\lambda}(\hat{\beta}))/\partial \beta = 0$. Hence,

$$0 = n^{1/2} \frac{\partial \hat{P}(\beta_0, \hat{\lambda}_0)}{\partial \beta} + \frac{\partial^2 \hat{P}(\beta^*, \hat{\lambda}(\beta^*))}{\partial \beta \partial \beta'} n^{1/2} (\hat{\beta} - \beta_0) \quad (\text{A.10})$$

[A.4]

for some β^* on the line segment joining $\hat{\beta}$ and β_0 which may differ row by row. From Lemma B.2 and eq. (A.1),

$$n^{1/2} \frac{\partial \hat{P}(\beta_0, \hat{\lambda}_0)}{\partial \beta} = n^{1/2} \hat{A} + n^{-1/2} \sum_{i=1}^n \mathbb{T}_{i,n} \sum_{j=1}^n w_{ij} \rho_1(\hat{\lambda}'_{i0} u_{j0}) U'_{j0} r_i, \quad (\text{A.11})$$

where

$$\hat{A} = - \sum_{i=1}^n \mathbb{T}_{i,n} \left[\sum_{j=1}^n w_{ij} \rho_1(\hat{\lambda}'_{i0} u_{j0}) U_{j0} \right]' \hat{V}(x_i, \beta_0)^{-1} \hat{u}_i(\beta_0) / n. \quad (\text{A.12})$$

From Lemma B.1, $\sup_{\beta \in \mathcal{B}, \lambda_i \in \Lambda_n, 1 \leq j \leq n} |\rho_1(\lambda'_i u_j(\beta)) - \rho_1(0)| = o_p(1)$. Therefore, w.p.a.1,

$$\begin{aligned} n^{-1/2} \left\| \sum_{i=1}^n \mathbb{T}_{i,n} \sum_{j=1}^n w_{ij} \rho_1(\hat{\lambda}'_{i0} u_{j0}) U'_{j0} r_i \right\| &\leq O_p(n^{1/2}) \max_{1 \leq i \leq n} \mathbb{T}_{i,n} \|r_i\| \sum_{i=1}^n \sum_{j=1}^n w_{ij} d(z_j) / n \\ &= o_p\left(\sqrt{\frac{n^{2\sigma+2/m}}{nb_n^{2s+4\tau}}}\right) + o_p\left(\frac{1}{n^{2\rho-3/m-1/2}}\right) \\ &= o_p(1) \end{aligned} \quad (\text{A.13})$$

uniformly i, j and $\beta \in \mathcal{B}_0$ by Assumption 4.7.

Let

$$\hat{A} = A + \sum_{i=1}^n \mathbb{T}_{i,n} \sum_{j=1}^n w_{ij} [\rho_1(\hat{\lambda}'_{i0} u_{j0}) - \rho_1(0)] U'_{j0} \hat{V}(x_i, \beta_0)^{-1} \hat{u}_i(\beta_0) / n, \quad (\text{A.14})$$

from (A.12), where

$$A = \sum_{i=1}^n \mathbb{T}_{i,n} \left[\sum_{j=1}^n w_{ij} U_{j0} \right]' \hat{V}(x_i, \beta_0)^{-1} \hat{u}_i(\beta_0) / n. \quad (\text{A.15})$$

Now, by Lemma B.1,

$$\begin{aligned} \left\| \mathbb{T}_{i,n} \sum_{j=1}^n w_{ij} [\rho_1(\hat{\lambda}'_{i0} u_{j0}) - \rho_1(0)] U_{j0} \right\| &\leq O_p(1) \mathbb{T}_{i,n} \left\| \hat{\lambda}_{i0} \right\| \\ &\quad \times \sum_{j=1}^n w_{ij} \sup_{\beta \in \mathcal{B}} \|u(z_j, \beta)\| d(z_j) \end{aligned}$$

uniformly i, j and $\beta \in \mathcal{B}_0$. Moreover, as $\max_{1 \leq i \leq n} \left\| \hat{V}(x_i, \beta_0)^{-1} \right\| = O_p(1)$, KTA, Lemma B.7, and $\max_{1 \leq i \leq n} \mathbb{T}_{i,n} \left\| \sum_{j=1}^n w_{ij} \sup_{\beta \in \mathcal{B}} \|u(z_j, \beta)\| d(z_j) \right\| = O_p(1)$ by Assumptions 4.2,

[A.5]

4.5 (iii) and 4.7, from (A.14)-(A.16),

$$\begin{aligned} n^{1/2} \left\| \hat{A} - A \right\| &\leq O_p(n^{1/2}) \max_{1 \leq i \leq n} \mathbf{T}_{i,n} \|\hat{u}_i(\beta_0)\| \left[\sum_{i=1}^n \mathbf{T}_{i,n} \|\hat{\lambda}_{i0}\| \right] / n \quad (\text{A.16}) \\ &= o_p\left(\sqrt{\frac{n^{2\sigma}}{nb_n^{2+4\tau}}}\right) + o_p\left(\frac{1}{n^{2\rho-2/m-1/2}}\right) = o_p(1), \end{aligned}$$

from eq. (B.3) and by KTA, Lemma B.3. Therefore, substituting (A.13) and (A.16) into (A.11), $n^{1/2} \partial \hat{P}(\beta_0, \hat{\lambda}_0) / \partial \beta = n^{1/2} A + o_p(1)$ and the result follows from Lemma B.3 and the continuity of $\mathcal{I}(\beta)$ from Assumption 4.5 (ii) on \mathcal{B}_0 as $n^{1/2} A \xrightarrow{d} N(0, \mathcal{I})$ from CLT by KTA, Lemma B.2. ■

Proof of Theorem 5.1: As $R(\beta_0)$ is full row rank r , by the implicit function theorem there exists neighbourhood \mathcal{N} of β_0 , an open set $\mathcal{U} \subseteq \mathcal{R}^{p-r}$ and a twice continuously differentiable function $\beta(\cdot) : \mathcal{U} \rightarrow \mathcal{R}^p$ such that $\{\beta \in \mathcal{N} : r(\beta) = 0\} = \{\beta = \beta(\alpha) : \alpha \in \mathcal{U}\}$. Therefore, any $\beta \in \mathcal{N}$ may be expressed as $\beta = \beta(\alpha)$ for some $\alpha \in \mathcal{U}$. In particular, $\beta_0 = \beta(\alpha_0)$ where $\alpha_0 \in \mathcal{U}$. Moreover, $\beta(\cdot)$ is twice continuously differentiable on \mathcal{U} and B is full column rank $p - r$ where $B \equiv B(\alpha_0)$ and $B(\alpha) \equiv \partial \beta(\alpha) / \partial \alpha'$. Cf. the Proof of Theorem 4.1 of KTA.

The restricted local GEL criterion under $H_0 : r(\beta_0) = 0$ (5.1) becomes

$$\hat{P}(\beta(\alpha), \lambda) = \sum_{i=1}^n \mathbf{T}_{i,n} \sum_{j=1}^n w_{ij} [\rho(\lambda'_i u_j(\beta(\alpha))) - \rho(0)] / n, \quad (\text{A.17})$$

cf. (3.1). The restricted GEL estimator is then given by $\tilde{\beta} \equiv \beta(\tilde{\alpha})$ where $\tilde{\alpha} = \arg \inf_{\alpha \in \mathcal{U}} \sum_{i=1}^n \mathbf{T}_{i,n} \sup_{\lambda_i \in \Lambda_n} \hat{P}_i(\beta(\alpha), \lambda_i) / n$. Therefore, as $n^{1/2}(\hat{\beta} - \beta_0) = -\mathcal{I}^{-1} n^{1/2} A + o_p(1) = -\mathcal{I}^{-1} n^{-1/2} \sum_{i=1}^n D(x_i, \beta_0)' V(x_i, \beta_0)^{-1} u_i(\beta_0) + o_p(1)$ from the Proof of Theorem 4.2 and KTA, eq. (B.6), under $H_0 : r(\beta_0) = 0$,

$$n^{1/2}(\tilde{\alpha} - \alpha_0) = -(B' \mathcal{I} B)^{-1} B' n^{-1/2} \sum_{i=1}^n D(x_i, \beta(\alpha_0))' V(x_i, \beta(\alpha_0))^{-1} u_i(\beta(\alpha_0)) + o_p(1), \quad (\text{A.18})$$

where $\mathcal{I} = E[D(x_i, \beta(\alpha_0))' V(x_i, \beta(\alpha_0))^{-1} D(x_i, \beta(\alpha_0))]$, cf. KTA, eq. (A.18).

Using a second order Taylor expansion of $\hat{P}(\beta_0, \hat{\lambda}(\beta_0))$ around $\hat{\beta}$, by Lemma B.3,

$$\begin{aligned} 2n[\hat{P}(\hat{\beta}, \hat{\lambda}(\hat{\beta})) - \hat{P}(\beta_0, \hat{\lambda}(\beta_0))] &= -n(\hat{\beta} - \beta_0)' \frac{\partial^2 \hat{P}(\beta^*, \hat{\lambda}(\beta^*))}{\partial \beta \partial \beta'} (\hat{\beta} - \beta_0) \\ &= n(\hat{\beta} - \beta_0)' \mathcal{I} (\hat{\beta} - \beta_0) + o_p(1) \end{aligned}$$

[A.6]

for some β^* between $\hat{\beta}$ and β_0 . Similarly, a Taylor expansion of $\hat{P}(\beta(\alpha_0), \hat{\lambda}(\beta(\alpha_0)))$ around $\tilde{\alpha}$ yields

$$2n[\hat{P}(\beta(\tilde{\alpha}), \hat{\lambda}(\beta(\tilde{\alpha}))) - \hat{P}(\beta(\alpha_0), \hat{\lambda}(\beta(\alpha_0)))] = -n(\tilde{\alpha} - \alpha_0)' \frac{\partial^2 \hat{P}(\beta(\alpha^*), \hat{\lambda}(\beta(\alpha^*)))}{\partial \alpha \partial \alpha'} (\tilde{\alpha} - \alpha_0) \quad (\text{A.19})$$

for some α^* between $\tilde{\alpha}$ and α_0 . Now

$$\frac{\partial^2 \hat{P}(\beta(\alpha), \hat{\lambda}(\beta(\alpha)))}{\partial \alpha \partial \alpha'} = B(\alpha)' \frac{\partial^2 \hat{P}(\beta(\alpha), \hat{\lambda}(\beta(\alpha)))}{\partial \beta \partial \beta'} B(\alpha) + \sum_{k=1}^p \frac{\partial^2 \beta^{(k)}(\alpha)}{\partial \alpha \partial \alpha'} \frac{\partial \hat{P}(\beta(\alpha), \hat{\lambda}(\beta(\alpha)))}{\partial \alpha^{(k)}},$$

and, by Lemmata B.3 and B.8 and Assumption 5.1 (i),

$$\frac{\partial^2 \hat{P}(\beta(\alpha^*), \hat{\lambda}(\beta(\alpha^*)))}{\partial \alpha \partial \alpha'} = -B' \mathcal{I} B + o_p(1). \quad (\text{A.20})$$

Combining eqs. (A.19) and (A.20),

$$2n[\hat{P}(\beta(\tilde{\alpha}), \hat{\lambda}(\beta(\tilde{\alpha}))) - \hat{P}(\beta(\alpha_0), \hat{\lambda}(\beta(\alpha_0)))] = n(\tilde{\alpha} - \alpha_0)' B' \mathcal{I} B (\tilde{\alpha} - \alpha_0) + o_p(1). \quad (\text{A.21})$$

Therefore, from eqs. (A.15), (A.18), (A.19), (A.21) and KTA, eq. (B.6),

$$\begin{aligned} LR_n^{GEL} &= 2n[\hat{P}(\hat{\beta}, \hat{\lambda}(\hat{\beta})) - \hat{P}(\beta(\tilde{\alpha}), \hat{\lambda}(\beta(\tilde{\alpha})))] \\ &= nA'(\mathcal{I}^{-1} - B(B' \mathcal{I} B)^{-1} B')A + o_p(1) \\ &\xrightarrow{d} \chi^2(r), \end{aligned}$$

Rao and Mitra (1971, Theorem 9.2.1, p.171), as $\mathcal{I}(\mathcal{I}^{-1} - B(B' \mathcal{I} B)^{-1} B') \mathcal{I}(\mathcal{I}^{-1} - B(B' \mathcal{I} B)^{-1} B') \mathcal{I} = \mathcal{I}(\mathcal{I}^{-1} - B(B' \mathcal{I} B)^{-1} B') \mathcal{I}$ and $tr(\mathcal{I}(\mathcal{I}^{-1} - B(B' \mathcal{I} B)^{-1} B')) = p - (p - r) = r$. ■

Appendix B: Auxiliary Results

The following Lemma is used extensively in the Proofs of Theorems 4.1, 4.2 and various of the Lemmata given below.

Let $\widehat{T}_{i,n} \equiv T_{i,n}h(x_i)/\widehat{h}(x_i)$.

Lemma B.1 *Suppose Assumptions 4.2 and 4.6 are satisfied. Then for any ζ with $1/m \leq \zeta < 1/2$ and $\Lambda_n = \{\lambda : \|\lambda\| \leq Cn^{-\zeta}, C > 0\}$, $\sup_{\beta \in \mathcal{B}, \lambda_i \in \Lambda_n, 1 \leq j \leq n} |\lambda'_i u_j(\beta)| \xrightarrow{p} 0$ and w.p.a.1, $\lambda'_i u_j(\beta) \in \mathcal{V}$ for all $1 \leq j \leq n$, $\lambda \in \Lambda_n$ and $\beta \in \mathcal{B}$.*

Proof. By Assumption 4.2 and KTA, Lemma D.2, $\max_{1 \leq j \leq n} \sup_{\beta \in \mathcal{B}} \|u_j(\beta)\| = o_p(n^{1/m})$; also see Owen (1990, Lemma 3). It therefore follows from Assumption 4.6 that

$$\sup_{\beta \in \mathcal{B}} \sup_{\lambda_i \in \Lambda_n} \max_{1 \leq j \leq n} |\lambda'_i u_j(\beta)| \leq Cn^{-\zeta} \max_{1 \leq j \leq n} \sup_{\beta \in \mathcal{B}} \|u_j(\beta)\| \xrightarrow{p} 0.$$

Therefore, w.p.a.1, $\lambda'_i u_j(\beta) \in \mathcal{V}$ for all $1 \leq j \leq n$, $\lambda_i \in \Lambda_n$ and $\beta \in \mathcal{B}$. ■

The next Lemma parallels KTA, Lemma B.1, which provides a similar result for local EL.

Lemma B.2 *Let Assumptions 4.2-4.5 be satisfied. Also let $n^{1-\sigma-2/m} b_n^{s+2\tau} \uparrow \infty$, $n^{\rho-2/m} \uparrow \infty$ and $n^{1-\sigma} b_n^{\frac{(m+2)s}{m-2}} \uparrow \infty$ for some $\sigma \in (0, 1)$ and $b_n \downarrow 0$. Then $T_{i,n} \widehat{\lambda}_{i0} = -T_{i,n} \widehat{V}(x_i, \beta_0)^{-1} \widehat{u}_i(\beta_0) + T_{i,n} r_i$, where $\max_{1 \leq i \leq n} T_{i,n} \|r_i\| = o_p\left(\frac{n^{\sigma+1/m}}{nb_n^{s+2\tau}}\right) + o_p\left(\frac{1}{n^{2\rho-3/m}}\right)$.*

Proof. From eq. (A.1)

$$\begin{aligned} 0 &= T_{i,n} \sum_{j=1}^n w_{ij} \rho_1(\widehat{\lambda}'_{i0} u_{j0}) u_{j0} \\ &= -T_{i,n} \widehat{u}_i(\beta_0) - T_{i,n} \widehat{V}(x_i, \beta_0) \lambda_{i0} + T_{i,n} r_{1i}(t), \end{aligned} \tag{B.1}$$

for some $t \in (0, 1)$, where $r_{1i}(t) = \sum_{j=1}^n w_{ij} [\rho_2(t \widehat{\lambda}'_{i0} u_{j0}) - \rho_2(0)] u_{j0} u'_{j0} \widehat{\lambda}_{i0}$. From Lemma B.1 $\sup_{\beta \in \mathcal{B}, \lambda_i \in \Lambda_n, 1 \leq j \leq n} |\rho_2(\lambda'_i u_j(\beta)) - \rho_2(0)| \xrightarrow{p} 0$. Thus, $r_{1i}(t) = o_p(1) \widehat{V}(x_i, \beta_0) \widehat{\lambda}_{i0}$ uni-

[B.1]

formly i and j and

$$\begin{aligned}
\mathbb{T}_{i,n} \|r_{1i}(t)\| &\leq o_p(1) \max_{1 \leq j \leq n} \|u_{j0}\| \mathbb{T}_{i,n} \left| \hat{u}_i(\beta_0)' \hat{\lambda}_{i0} \right| \\
&\leq o_p(n^{1/m}) \max_{1 \leq j \leq n} \|u_{j0}\| \mathbb{T}_{i,n} \|\hat{u}_i(\beta_0)\| \|\hat{\lambda}_{i0}\| \\
&= o_p(n^{1/m}) [o_p(\sqrt{\frac{n^\sigma}{nb_n^{s+2\tau}}}) + o_p(\frac{1}{n^{\rho-1/m}})] \mathbb{T}_{i,n} \|\hat{\lambda}_{i0}\|
\end{aligned} \tag{B.2}$$

where the second inequality follows from CS and $\max_{1 \leq j \leq n} \|u_{j0}\| = o(n^{1/m})$ by KTA, Lemma D.2, and the equality by KTA, Lemma B.3; see KTA, eq. (B.3).

Let $\xi_{i0} = \hat{\lambda}_{i0} / \|\hat{\lambda}_{i0}\|$. Then, multiplying eq. (A.1) by $\hat{\lambda}_{i0}$ yields

$$\begin{aligned}
0 &= \mathbb{T}_{i,n} \|\hat{\lambda}_{i0}\| \sum_{j=1}^n w_{ij} \rho_1(\hat{\lambda}'_{i0} u_{j0}) \xi'_{i0} u_{j0} \\
&= -\mathbb{T}_{i,n} \|\hat{\lambda}_{i0}\| \xi'_{i0} \hat{u}_i(\beta_0) - \mathbb{T}_{i,n} \|\hat{\lambda}_{i0}\|^2 (1 + o_p(1)) \xi'_{i0} \hat{V}(x_i, \beta_0) \xi_{i0}
\end{aligned}$$

uniformly i and j . As $\xi'_{i0} \hat{V}(x_i, \beta_0) \xi_{i0}$ is bounded below by Assumption 4.5 (ii) solving

$$\begin{aligned}
\mathbb{T}_{i,n} \|\hat{\lambda}_{i0}\| &= -\mathbb{T}_{i,n} \xi'_{i0} \hat{u}_i(\beta_0) / (1 + o_p(1)) \xi'_{i0} \hat{V}(x_i, \beta_0) \xi_{i0} \\
&= o_p(\sqrt{\frac{n^\sigma}{nb_n^{s+2\tau}}}) + o_p(\frac{1}{n^{\rho-1/m}})
\end{aligned} \tag{B.3}$$

uniformly i by KTA, Lemma B.3, as $\hat{V}(x_i, \beta_0) = O_p(1)$ from KTA, Lemma B.6. Therefore, from eq. (B.2),

$$\mathbb{T}_{i,n} \|r_{1i}(t)\| = o_p(\frac{n^{\sigma+1/m}}{nb_n^{s+2\tau}}) + o_p(\frac{1}{n^{2\rho-3/m}}) \tag{B.4}$$

uniformly i .

By Assumption 4.5 (ii), from eqs. (B.1) and (B.4), as $\max_{1 \leq i \leq n} \mathbb{T}_{i,n} \|\hat{V}(x_i, \beta_0)^{-1}\| = O_p(1)$ by KTA, Lemma B.7,

$$\begin{aligned}
\mathbb{T}_{i,n} \hat{\lambda}_{i0} &= -\mathbb{T}_{i,n} \hat{V}(x_i, \beta_0)^{-1} \hat{u}_i(\beta_0) + \mathbb{T}_{i,n} \hat{V}(x_i, \beta_0)^{-1} r_{1i}(t) \\
&= -\mathbb{T}_{i,n} \hat{V}(x_i, \beta_0)^{-1} \hat{u}_i(\beta_0) + \mathbb{T}_{i,n} r_i.
\end{aligned}$$

■

Lemmata B.3-B.8 given below are the local GEL counterparts of KTA, Lemmas C.1-C.6. Our proofs follow closely those in KTA.

[B.2]

Lemma B.3 *Let Assumptions 4.1-4.7 hold. Then $\sup_{\beta \in \mathcal{B}_0} \left\| -\partial^2 \hat{P}(\beta, \hat{\lambda}(\beta)) / \partial \beta \partial \beta' - \mathcal{I}(\beta) \right\| = o_p(1)$.*

Proof. As $\mathbb{T}_{i,n} \sum_{j=1}^n w_{ij} \rho_1(\hat{\lambda}_i(\beta)' u_j(\beta)) u_j(\beta) = 0$, ($i = 1, \dots, n$), for all $\beta \in \mathcal{B}$ from (A.1),

$$\frac{\partial \hat{P}(\beta, \hat{\lambda}(\beta))}{\partial \beta} = \sum_{i=1}^n \mathbb{T}_{i,n} \sum_{j=1}^n w_{ij} \rho_1(\hat{\lambda}_i(\beta)' u_j(\beta)) U_j(\beta)' \hat{\lambda}_i(\beta) / n. \quad (\text{B.5})$$

Therefore, $\partial^2 \hat{P}(\beta, \hat{\lambda}(\beta)) / \partial \beta \partial \beta' = T_1(\beta) + T_2(\beta) + T_3(\beta)$ where

$$\begin{aligned} T_1(\beta) &= \sum_{i=1}^n \mathbb{T}_{i,n} \sum_{j=1}^n w_{ij} \rho_2(\hat{\lambda}_i(\beta)' u_j(\beta)) U_j(\beta)' \hat{\lambda}_i(\beta) \partial[\hat{\lambda}_i(\beta)' u_j(\beta)] / \partial \beta' / n, \\ T_2(\beta) &= \sum_{i=1}^n \mathbb{T}_{i,n} \sum_{j=1}^n w_{ij} \rho_1(\hat{\lambda}_i(\beta)' u_j(\beta)) [U_j(\beta)' \partial \hat{\lambda}_i(\beta) / \partial \beta'] / n, \\ T_3(\beta) &= \sum_{i=1}^n \mathbb{T}_{i,n} \sum_{j=1}^n w_{ij} \rho_1(\hat{\lambda}_i(\beta)' u_j(\beta)) \left[\sum_{k=1}^q \hat{\lambda}_i^{(k)}(\beta) \partial^2 u_j^{(k)}(\beta) / \partial \beta \partial \beta' \right] / n. \end{aligned}$$

From Lemmata B.4-B.6 the desired result follows. ■

Lemma B.4 *If Assumptions 4.2-4.7 are satisfied, then $\sup_{\beta \in \mathcal{B}_0} \|T_1(\beta)\| = o_p(1)$.*

Proof. As $\partial[\hat{\lambda}_i(\beta)' u_j(\beta)] / \partial \beta' = \hat{\lambda}_i(\beta)' U_j(\beta) + u_j(\beta)' \partial \hat{\lambda}_i(\beta) / \partial \beta'$, consider

$$\begin{aligned} T_{1,a}(\beta) &= \sum_{i=1}^n \mathbb{T}_{i,n} \sum_{j=1}^n w_{ij} \rho_2(\hat{\lambda}_i(\beta)' u_j(\beta)) U_j(\beta)' \hat{\lambda}_i(\beta) \hat{\lambda}_i(\beta)' U_j(\beta) / n, \\ T_{1,b}(\beta) &= \sum_{i=1}^n \mathbb{T}_{i,n} \sum_{j=1}^n w_{ij} \rho_2(\hat{\lambda}_i(\beta)' u_j(\beta)) U_j(\beta)' \hat{\lambda}_i(\beta) u_j(\beta)' \partial \hat{\lambda}_i(\beta) / \partial \beta' / n. \end{aligned}$$

By Lemma B.1, Assumptions 4.5 (iii) and 4.6 $\sup_{\beta \in \mathcal{B}_0} \|T_{1,a}(\beta)\| \leq o_p(1) \sum_{i=1}^n \mathbb{T}_{i,n} \sum_{j=1}^n w_{ij} d(z_j)^2 / n = o_p(1)$ as $\mathbb{T}_{i,n} \sum_{j=1}^n w_{ij} d(z_j)^2 = O_p(1)$ uniformly i by KTA, Lemma D.4. Also, $\sup_{\beta \in \mathcal{B}_0} \|T_{1,b}(\beta)\| \leq o_p(1) \sum_{i=1}^n \mathbb{T}_{i,n} \left\| \partial \hat{\lambda}_i(\beta) / \partial \beta' \right\| \sum_{j=1}^n w_{ij} d(z_j) \|u_j(\beta)\| / n = o_p(1) \sum_{i=1}^n \mathbb{T}_{i,n} \left\| \partial \hat{\lambda}_i(\beta) / \partial \beta' \right\| / n$ since similarly $\mathbb{T}_{i,n} \sum_{j=1}^n w_{ij} d(z_j) \|u_j(\beta)\| \leq (\mathbb{T}_{i,n} \sum_{j=1}^n w_{ij} d(z_j)^2) (\mathbb{T}_{i,n} \sum_{j=1}^n w_{ij} \|u_j(\beta)\|^2) = O_p(1)$ uniformly i, j and $\beta \in \mathcal{B}_0$. Now, from Lemma B.7 below, $\sup_{\beta \in \mathcal{B}_0} \sum_{i=1}^n \mathbb{T}_{i,n} \left\| \partial \hat{\lambda}_i(\beta) / \partial \beta' \right\| / n = o_p(1)$. Hence, $\sup_{\beta \in \mathcal{B}_0} \|T_{1,b}(\beta)\| \leq o_p(1)$ from Lemma B.7 as $\sum_{i=1}^n \mathbb{T}_{i,n} \left\| \partial \hat{\lambda}_i(\beta) / \partial \beta' \right\| / n = O_p(1)$. ■

Lemma B.5 *If Assumptions 4.2-4.7 are satisfied, then $\sup_{\beta \in \mathcal{B}_0} \|-T_2(\beta) - \mathcal{I}(\beta)\| = o_p(1)$.*

Proof. Using Lemma B.7 below, by a similar argument to that above KTA, eq. (C.3),

as $\sup_{\beta \in \mathcal{B}, \lambda_i \in \Lambda_n, 1 \leq j \leq n} |\rho_1(\lambda_i' u_j(\beta)) - \rho_1(0)| = o_p(1)$ from Lemma B.1,

$$\begin{aligned} & \sum_{i=1}^n \mathbb{T}_{i,n} \sum_{j=1}^n w_{ij} \rho_1(\hat{\lambda}_i(\beta)' u_j(\beta)) [U_j(\beta)' \partial \hat{\lambda}_i(\beta) / \partial \beta'] / n \\ &= - \sum_{i=1}^n \mathbb{T}_{i,n} D(x_i, \beta)' \partial \hat{\lambda}_i(\beta) / \partial \beta' / n + o_p(1), \end{aligned}$$

uniformly $\beta \in \mathcal{B}_0$. Again using Lemma B.7,

$$n^{-1} \sup_{\beta \in \mathcal{B}_0} \left\| \sum_{i=1}^n \hat{\mathbb{T}}_{i,n} D(x_i, \beta)' \partial \hat{\lambda}_i(\beta) / \partial \beta' - \sum_{i=1}^n \hat{\mathbb{T}}_{i,n}^2 D(x_i, \beta)' V(x_i, \beta) D(x_i, \beta) \right\| = o_p(1).$$

Therefore, from KTA, below eq. (C.3),

$$\begin{aligned} & \sum_{i=1}^n \mathbb{T}_{i,n} \sum_{j=1}^n w_{ij} \rho_1(\hat{\lambda}_i(\beta)' u_j(\beta)) [U_j(\beta)' \partial \hat{\lambda}_i(\beta) / \partial \beta'] / n \\ &= - \sum_{i=1}^n D(x_i, \beta)' V(x_i, \beta) D(x_i, \beta) / n + o_p(1), \end{aligned}$$

uniformly $\beta \in \mathcal{B}_0$, cf. KTA, eq. (C.4). The result follows by UWL. ■

Lemma B.6 *If Assumptions 4.3, 4.5 and 4.6 are satisfied, then $\sup_{\beta \in \mathcal{B}_0} \|T_3(\beta)\| = o_p(1)$.*

Proof. By Assumptions 4.5 (iii) and 4.6 and Lemma B.1, from KTA, Lemma D.4,

$$\begin{aligned} n^{-1} \sup_{\beta \in \mathcal{B}_0} \left\| \sum_{i=1}^n \mathbb{T}_{i,n} \sum_{j=1}^n w_{ij} \rho_1(\hat{\lambda}_i(\beta)' u_j(\beta)) \left[\sum_{k=1}^q \hat{\lambda}_i^{(k)}(\beta) \partial^2 u_j^{(k)}(\beta) / \partial \beta \partial \beta' \right] \right\| &\leq o_p(1) \\ &\times \sum_{i=1}^n \mathbb{T}_{i,n} \sum_{j=1}^n w_{ij} l(z_j) / n \\ &= o_p(1). \end{aligned}$$

■

Lemma B.7 *If Assumptions 4.2-4.7 are satisfied, then, for each i and $\beta \in \mathcal{B}_0$,*

$$\begin{aligned} \mathbb{T}_{i,n} \partial \hat{\lambda}_i(\beta) / \partial \beta' &= \hat{\mathbb{T}}_{i,n} V(x_i, \beta)^{-1} D(x_i, \beta) + \hat{\mathbb{T}}_{i,n} M_{1,i}(\beta) D(x_i, \beta) \\ &\quad + \hat{\mathbb{T}}_{i,n} M_{2,i}(\beta) E[d(z_i) | x_i] + M_{3,i}(\beta) \sum_{j=1}^n w_{ij} d(z_j) + M_{4,i}(\beta), \end{aligned}$$

where $\max_{1 \leq i \leq n} \sup_{\beta \in \mathcal{B}_0} \|M_{k,i}(\beta)\| = o_p(1)$, $k = 1, \dots, 4$.

Proof. Firstly, from differentiating (A.1), we have

$$\begin{aligned} \sum_{j=1}^n w_{ij} \rho_2(\hat{\lambda}_i(\beta)' u_j(\beta)) u_j(\beta) u_j(\beta)' \partial \hat{\lambda}_i(\beta) / \partial \beta' &= - \sum_{j=1}^n w_{ij} \rho_1(\hat{\lambda}_i(\beta)' u_j(\beta)) U_j(\beta) \\ &\quad - \sum_{j=1}^n w_{ij} \rho_2(\hat{\lambda}_i(\beta)' u_j(\beta)) u_j(\beta) \hat{\lambda}_i(\beta)' U_j(\beta). \end{aligned} \quad (\text{B.6})$$

By Lemma B.1, from Assumption 4.5 (ii) and KTA, Lemma B.6,

$$\max_{1 \leq i \leq n} \sup_{\beta \in \mathcal{B}_0} \mathbb{T}_{i,n} \left\| - \sum_{j=1}^n w_{ij} \rho_2(\hat{\lambda}_i(\beta)' u_j(\beta)) u_j(\beta) u_j(\beta)' - V(x_i, \beta) \right\| = o_p(1). \quad (\text{B.7})$$

Thus, with $\max_{1 \leq i \leq n} \sup_{\beta \in \mathcal{B}_0} \|R_{1,i}(\beta)\| = o_p(1)$, from Assumption 4.5 (ii),

$$\mathbb{T}_{i,n} \left[\sum_{j=1}^n w_{ij} \rho_2(\hat{\lambda}_i(\beta)' u_j(\beta)) u_j(\beta) u_j(\beta)' \right]^{-1} = -\mathbb{T}_{i,n} V(x_i, \beta)^{-1} + R_{1,i}(\beta); \quad (\text{B.8})$$

cf. KTA, eq. (C.6).

Similarly, by Lemma B.1 and KTA, Lemma B.5, an argument like that for KTA, eq. (C.7), yields

$$\mathbb{T}_{i,n} \sum_{j=1}^n w_{ij} \rho_1(\hat{\lambda}_i(\beta)' u_j(\beta)) U_j(\beta) = \hat{\mathbb{T}}_{i,n} \rho_1(0) D(x_i, \beta) + \hat{\mathbb{T}}_{i,n} E[d(z_i) | x_i] R_{2,i}(\beta) + R_{3,i}(\beta), \quad (\text{B.9})$$

where $\max_{1 \leq i \leq n} \sup_{\beta \in \mathcal{B}_0} \|R_{2,i}(\beta)\| = o_p(1)$ and $\max_{1 \leq i \leq n} \sup_{\beta \in \mathcal{B}_0} \|R_{3,i}(\beta)\| = o_p(1)$.

Finally, by Lemma B.1, Assumptions 4.5 (iii) and 4.6, from KTA, Lemma D.2,

$$\begin{aligned} \mathbb{T}_{i,n} \left\| \sum_{j=1}^n w_{ij} \rho_2(\hat{\lambda}_i(\beta)' u_j(\beta)) u_j(\beta) \hat{\lambda}_i(\beta)' U_j(\beta) \right\| &\leq O_p(1) \mathbb{T}_{i,n} \sum_{j=1}^n w_{ij} d(z_j) \left\| \hat{\lambda}_i(\beta) \right\| \|u_j(\beta)\| \\ &= o_p(1) \sum_{j=1}^n w_{ij} d(z_j), \end{aligned}$$

[B.5]

uniformly i, j and $\beta \in \mathcal{B}_0$. Therefore, for $\max_{1 \leq i \leq n} \sup_{\beta \in \mathcal{B}_0} \|R_{4,i}(\beta)\| = o_p(1)$,

$$\mathbb{T}_{i,n} \sum_{j=1}^n w_{ij} \rho_2(\hat{\lambda}_i(\beta)' u_j(\beta)) u_j(\beta) \hat{\lambda}_i(\beta)' U_j(\beta) = R_{4,i}(\beta) \sum_{j=1}^n w_{ij} d(z_j), \quad (\text{B.10})$$

cf. KTA, eq. (C.8).

Substituting eqs. (B.7)-(B.10) in eq. (B.6) and solving for $\mathbb{T}_{i,n} \partial \hat{\lambda}_i(\beta) / \partial \beta'$ yields the desired result. ■

Lemma B.8 *Let Assumptions 4.4, 4.5 and 4.6 hold. Then $\sup_{\beta \in \mathcal{B}_0} \left\| \partial \hat{P}(\beta, \hat{\lambda}(\beta)) / \partial \beta \right\| = o_p(1)$.*

Proof. From eq. (B.5), Assumptions 4.5 (ii) and 4.6, $\sup_{\beta \in \mathcal{B}_0} \left\| \partial \hat{P}(\beta, \hat{\lambda}(\beta)) / \partial \beta \right\| \leq o_p(1) \sum_{j=1}^n w_{ij} d(z_j) = o_p(1)$ uniformly i, j and $\beta \in \mathcal{B}_0$. ■

References

- Altonji, J. and L.M. Segal (1996): “Small Sample Bias in GMM Estimation of Covariance Structures,” *Journal of Economic and Business Statistics* 14, 353-366.
- Back, K., and D. Brown (1993): “Implied Probabilities in GMM Estimators,” *Econometrica* 61, 971-976.
- Bonnal, H., and E. Renault (2003): “On the Efficient Use of the Informational Content of Estimating Equations: Implied Probabilities and Maximum Euclidean Likelihood,” working paper, Département de Sciences Économiques, Université de Montréal.
- Brown, B.W. and W.K. Newey (1992): “Bootstrapping for GMM”, working paper, Department of Economics, M.I.T.
- Brown, B.W. and W.K. Newey (2002): “Generalized Method of Moments, Efficient Bootstrapping, and Improved Inference,” *Journal of Economic and Business Statistics*, 20, 507-517.
- Chamberlain, G. (1987): “Asymptotic Efficiency in Estimation with Conditional Moment Restrictions,” *Journal of Econometrics* 34, 305-334.
- Corcoran, S.A. (1998): “Bartlett Adjustment of Empirical Discrepancy Statistics”, *Biometrika* 85, 967-972.
- Cressie, N., and T. Read (1984): “Multinomial Goodness-of-Fit Tests”, *Journal of the Royal Statistical Society Series B* 46, 440-464.
- Donald, S.G., G.W. Imbens, and W.K. Newey (2003): “Empirical Likelihood Estimation and Consistent Tests with Conditional Moment Restrictions,” *Journal of Econometrics*, 117, 55-93.

- Hansen, L. P. (1982): “Large Sample Properties of Generalized Method of Moments Estimators”, *Econometrica* 50, 1029-1054.
- Hansen, L.P., J. Heaton and A. Yaron (1996): “Finite-Sample Properties of Some Alternative GMM Estimators,” *Journal of Business and Economic Statistics*, 14, 262-280.
- Imbens, G.W. (1997): “One-Step Estimators for Over-Identified Generalized Method of Moments Models,” *Review of Economic Studies*, 64, 359-383.
- Imbens, G.W. and R.H. Spady (2005): “The Performance of Empirical Likelihood and Its Generalizations.” Chapter 10 in *Identification and Inference in Econometric Models: Essays in Honor of Thomas J. Rothenberg*, eds. D.W.K. Andrews and J.H. Stock, 216-244. Cambridge University Press: Cambridge.
- 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. (2001): “Asymptotic Optimality of Empirical Likelihood for Testing Moment Restriction,” *Econometrica*, 69, 1661-1672.
- Kitamura, Y., and M. Stutzer (1997): “An Information-Theoretic Alternative to Generalized Method of Moments Estimation,” *Econometrica*, 65, 861-874.
- Kitamura, Y., G. Tripathi, and H. Ahn (2004): “Empirical Likelihood-Based Inference in Conditional Moment Restriction Models,” *Econometrica* 72, 1667-1714.
- Mittelhammer, R.C., G.G. Judge and R. Schoenberg (2005): ”Empirical Evidence Concerning the Finite Sample Performance of EL-Type Structural Equation Estimation and Inference Methods.” Chapter 12 in *Identification and Inference in Econometric Models: Essays in Honor of Thomas J. Rothenberg*, eds. D.W.K. Andrews and J.H. Stock, 282-305. Cambridge University Press: Cambridge.

- Newey, W.K. (1990): “Efficient Instrumental Variables Estimation of Nonlinear Models,” *Econometrica*, 58, 809-837.
- Newey, W.K. (1993): “Efficient Estimation of Models with Conditional Moment Restrictions.” Chapter 16 in *Handbook of Statistics*, Vol. 11, eds. G.S. Maddala, C.R. Rao and H. Vinod, 419-454. Amsterdam: North Holland.
- Newey, W.K., and D.L. McFadden (1994): “Large Sample Estimation and Hypothesis Testing.” Chapter 36 in *Handbook of Econometrics*, Vol. 4, eds. R.F. Engle and D.L. McFadden, 2111-2245. New York: North Holland.
- Newey, W.K., J.J.S. Ramalho, and R.J. Smith (2005): “Asymptotic Bias for GMM and GEL Estimators with Estimated Nuisance Parameters.” Chapter 11 in *Identification and Inference in Econometric Models: Essays in Honor of Thomas J. Rothenberg*, eds. D.W.K. Andrews and J.H. Stock, 245-281. Cambridge University Press: Cambridge.
- Newey, W.K., and R.J. Smith (2004): “Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators,” *Econometrica*, 72, 219-255.
- Owen, A. (1988): “Empirical Likelihood Ratio Confidence Intervals for a Single Functional,” *Biometrika*, 75, 237-249.
- Owen, A. (1990): “Empirical Likelihood Ratio Confidence Regions,” *Annals of Statistics*, 18, 90-120.
- Owen, A. (2001): *Empirical Likelihood*, New York: Chapman and Hall.
- Qin, J. and Lawless, J. (1994): “Empirical Likelihood and General Estimating Equations,” *Annals of Statistics*, 22, 300-325.
- Ramalho, J.J.S. (2001): *Alternative Estimation Methods and Specification Tests for Moment Condition Models*, unpublished Ph.D. thesis, Department of Economics, University of Bristol.

- Robinson, P.M. (1987): “Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form,” *Econometrica*, 55, 875-891.
- Smith, R.J. (1997): “Alternative Semi-Parametric Likelihood Approaches to Generalized Method of Moments Estimation,” *Economic Journal*, 107, 503-519.
- Smith, R.J. (2001): “GEL Methods for Moment Condition Models”, working paper, Department of Economics, University of Bristol. Revised version CWP 19/04, cemmap, I.F.S. and U.C.L. <http://cemmap.ifs.org.uk/wps/cwp0419.pdf>
- Smith, R.J. (2003): “Efficient Information Theoretic Inference for Conditional Moment Restrictions”, working paper, Department of Economics, University of Warwick. Revised version CWP 14/05, cemmap, I.F.S. and U.C.L. <http://cemmap.ifs.org.uk/wps/cwp1405.pdf>