

Smith, Richard J.

Working Paper

Efficient information theoretic inference for conditional moment restrictions

cemmap working paper, No. CWP14/05

Provided in Cooperation with:

The Institute for Fiscal Studies (IFS), London

Suggested Citation: Smith, Richard J. (2005) : Efficient information theoretic inference for conditional moment restrictions, cemmap working paper, No. CWP14/05, Centre for Microdata Methods and Practice (cemmap), London, <https://doi.org/10.1920/wp.cem.2005.1405>

This Version is available at:

<https://hdl.handle.net/10419/79264>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

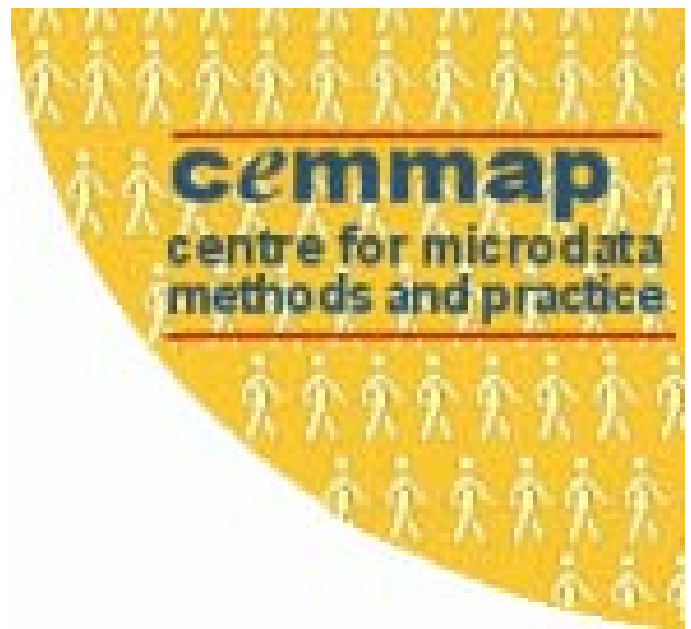
Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



EFFICIENT INFORMATION THEORETIC INFERENCE FOR CONDITIONAL MOMENT RESTRICTIONS

Richard J. Smith

THE INSTITUTE FOR FISCAL STUDIES
DEPARTMENT OF ECONOMICS, UCL
cemmap working paper CWP14/05

EFFICIENT INFORMATION THEORETIC INFERENCE FOR CONDITIONAL MOMENT RESTRICTIONS*

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

First Draft: August 2003
This Revision: October 2005

Abstract

The generalized method of moments estimator may be substantially biased in finite samples, especially so when there are large numbers of unconditional moment conditions. This paper develops a class of first order equivalent semi-parametric efficient estimators and tests for conditional moment restrictions models based on a local or kernel-weighted version of the Cressie-Read power divergence family of discrepancies. This approach is similar in spirit to the empirical likelihood methods of Kitamura, Tripathi and Ahn (2004) and Tripathi and Kitamura (2003). These efficient local methods avoid the necessity of explicit estimation of the conditional Jacobian and variance matrices of the conditional moment restrictions and provide empirical conditional probabilities for the observations.

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

Keywords: Conditional Moment Restrictions, Local Cressie-Read Minimum Discrepancy, GMM, Semi-Parametric Efficiency.

*This is a revision of aspects of an Invited Paper entitled "Local GEL Estimation with Conditional Moment Restrictions" presented at the A.S.A. Business and Economics Section conference "Recent Developments in the Theory, Method and Application of Information and Entropy Econometrics" held in honour of Arnold Zellner at the American University, Washington D.C., 19-21 September 2003. I am grateful to Y. Kitamura for his discussion of that paper. An earlier version of this paper also formed 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. This research benefitted from the financial support of a 2002 Leverhulme Major Research Fellowship.

1 Introduction

Evidence of substantial biases in finite samples of the standard generalized method of moments (GMM) estimator, Hansen (1982), for models specified by unconditional moment restrictions is becoming increasingly prevalent, especially so when there are large numbers of moment conditions. See, for example, the theoretical discussion in Newey and Smith (2004), henceforth NS, and the simulation evidence in Altonji and Segal (1996), Imbens and Spady (2001), Judge and Mittelhammer (2001), Ramalho (2001) and Newey, Ramalho and Smith (2001). A number of alternative estimators have therefore been suggested which are first order asymptotically equivalent to GMM, including 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). As demonstrated by NS, these estimators and those from the Cressie and Read (1984) power divergence family of discrepancies share a common structure, being members of a class of generalized empirical likelihood (GEL) estimators; see Brown and Newey (1992, 2002) and Smith (1997, 2001). Correspondingly NS show that GEL and GMM estimators display the same first order asymptotic properties. For the unconditional context, NS also describe the higher order efficiency of bias-corrected EL. Also see Kitamura (2001).

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. A principal aim of this paper is to adapt to the conditional moment context the information theoretic methods based on the Cressie and Read (1984) power divergence family of discrepancies discussed in Imbens, Spady and Johnson (1998) for unconditional moment restriction models and, thereby, to describe a class of information theoretic estimators which achieve the semi-parametric efficiency lower bound. Tripathi and Kitamura (2003), henceforth TK, propose an EL-type statistic for testing the validity of conditional moment restrictions. A further objective of this

paper is to extend the analysis of TK to provide a class of alternative test statistics based on the Cressie-Read family of discrepancy measures.

KTA employ a kernel weighted version of EL. The resultant EL criterion may be regarded as a form of local EL. Similar ideas have been employed elsewhere, for example, in nonparametric regression. For an excellent exposition of these ideas and applications, see Fan and Gijbels (1996). Like KTA for EL we define a class of estimation criteria based on local or kernel weighting of the Cressie-Read power divergence family of discrepancies. We term the consequent estimators local Cressie-Read minimum discrepancy (MD) estimators. We show that local Cressie-Read MD estimators are asymptotically first order equivalent to the local EL estimator proposed by KTA. Consequently local Cressie-Read MD estimators achieve the semi-parametric efficiency lower bound; see Chamberlain (1987). A reformulation of the first order conditions defining the local Cressie-Read MD estimator facilitates intuition for the semi-parametric efficiency of the local Cressie-Read MD estimator. The structure of these conditions conforms to those describing a semi-parametric efficient GMM estimator, incorporating an estimator of the efficient matrix of instrumental variables formed from implicit consistent estimators of the conditional Jacobian and conditional variance matrices associated with the conditional moment restrictions. The class of local Cressie-Read MD estimators includes local versions of EL as in KTA, the ET estimator and the CUE, the last of which is related to the estimator suggested by Bonnal and Renault (2003); cf. NS, Theorem 2.3, for unconditional GEL. Like TK for EL, the optimised local Cressie-Read MD criterion function suitably centred and scaled yields an asymptotically pivotal test statistic for the validity of the conditional moment restrictions.

Because of their one-step nature a particular advantage of efficient local methods is the avoidance of the necessity of providing explicit nonparametric estimators for the conditional Jacobian and variance matrices arising from the conditional moment restrictions which may be inaccurately estimated unless large numbers of observations are available. See, for example, Robinson (1987) and Newey (1990, 1993) for semi-parametric approaches based on explicit conditional Jacobian and variance matrix estimation. More-

over, efficient local methods display an invariance to normalisation of the conditional moment restrictions, a property absent for two-step semi-parametric efficient estimators. In contrast to local EL and the local Cressie-Read MD methods developed here, Donald, Imbens and Newey (2003) employs a sequence of unconditional moment restrictions, obtained, for example, using spline or series approximants, in order to approximate the efficient matrix of instrumental variables. This sequence of unconditional moment restrictions is then used within the standard GEL set-up discussed in NS from which a semi-parametric efficient estimator also results. Their method has the computational virtue of requiring estimation of a nuisance parameter vector whose dimension increases at a slower rate than the sample size whereas the dimension of that associated with efficient local methods increases in direct proportion. A disadvantage though is not producing an explicit estimator for the conditional distribution of the data.

The outline of the paper is as follows. In Section 2 the conditional moment restrictions model is described together with some other preliminaries. Section 3 details the class of local Cressie-Read MD criteria and estimators. Local EL, ET estimators and CUE are obtained as special cases of the local Cressie-Read MD estimator. This section also gives some intuition for their semi-parametric efficiency. Various regularity conditions are provided for the consistency, asymptotic normality and semi-parametric efficiency of local Cressie-Read MD estimation in section 4. Section 5 considers test statistics for the conditional moment restrictions based on optimised Cressie-Read MD criteria together with moment- and Lagrange multiplier-based statistics. Asymptotic results and associated regularity conditions for the null distribution of these statistics are presented in section 6. Proofs are given in Appendix A with certain subsidiary results and proofs in Appendices B and C for estimation and inference respectively.

2 The Model and Preliminaries

Let (x_i, z_i) , $(i = 1, \dots, n)$, be a random sample of observations on the s - and d -dimensional data vectors x and z . Following KTA, x is assumed to be continuously distributed with

Lebesgue density $h(\cdot)$ whereas z may be discrete, mixed or continuous.¹

The conditional moment indicator vector $u(z, \beta)$ is a known q -vector of functions of the data observation z and the p -dimensional parameter vector β which lies in a compact parameter space \mathcal{B} . In many contexts, the vector $u(z, \beta)$ would be interpreted as a vector of residuals from some econometric model. We assume there exists a value of the parameter vector β_0 in the interior of the parameter space \mathcal{B} satisfying the conditional moment restriction

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

Here $E[\cdot|x]$ denotes expectation taken with respect to the conditional distribution of z given x .

Efficient estimation of the parameter β_0 under (2.1) is one of the principal objectives of this paper. From (2.1), any measurable function of the conditioning vector x is uncorrelated with $u(z, \beta_0)$. A standard approach then to constructing a consistent estimator for β_0 would be to formulate a set of unconditional moment restrictions from (2.1) by specifying a $m \times q$ matrix of instrumental variables, $v(x, \beta)$ say, with $m \geq p$. The m -vector of unconditional moment indicators is defined as

$$g(z, \beta) = v(x, \beta)u(z, \beta) \quad (2.2)$$

with the consequent unconditional moment restrictions obtained by iterated expectations as

$$\begin{aligned} E[g(z, \beta_0)] &= E[v(x, \beta_0)u(z, \beta_0)] \\ &= E_x[v(x, \beta_0)E[u(z, \beta_0)|x]] = 0, \end{aligned} \quad (2.3)$$

where $E[\cdot]$ and $E_x[\cdot]$ denote expectation taken with respect to the joint unconditional distribution of x and z and the marginal distribution of x respectively. Under appropriate regularity conditions, see, inter alia, Newey and McFadden (1994) and NS, GMM or GEL estimation using $g(z, \beta)$ (2.2) as the vector of moment indicators will deliver consistent estimators of β_0 . In general, however, because the instrumental variables $v(x, \beta)$

¹The following analysis may be straightforwardly adapted for x discrete or mixed distributed. See KTA, section 3.

are nonunique, neither GMM nor GEL estimators based on the unconditional moment restrictions (2.3) will achieve the semi-parametric efficiency bound.

Let $D(x) = E[\partial u(z, \beta_0)/\partial \beta' | x]$ and $V(x) = E[u(z, \beta_0)u(z, \beta_0)' | x]$ denote the conditional Jacobian and conditional variance matrices arising from the conditional moment restrictions (2.1). In a seminal paper, 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 \mathcal{I}^{-1} where

$$\mathcal{I} = E_x[D(x)'V(x)^{-1}D(x)]. \quad (2.4)$$

The matrix \mathcal{I} may be regarded as the semi-parametric equivalent of the classical information matrix and is, in fact, derived from consideration of a particular classical parametric problem. An optimal GMM or GEL estimator based on the unconditional moment restrictions (2.3), therefore, requires the implementation of the (infeasible) instrumental variable matrix $v_*(x, \beta) = D(x, \beta)'V(x, \beta)^{-1}$, where $D(x, \beta) = E[\partial u(z, \beta)/\partial \beta' | x]$ and $V(x, \beta) = E[u(z, \beta)u(z, \beta)' | x]$.

3 Information Theoretic Estimation

A main concern of this paper, like KTA, then is the development of feasible estimators for β_0 which achieve the semi-parametric efficiency bound \mathcal{I}^{-1} under (2.1) but which avoid explicit estimation of the conditional Jacobian and conditional variance matrices, $D(x)$ and $V(x)$, of the moment indicator $u(z, \beta_0)$.

The approach adopted here considers a class of information theoretic criteria based on the Cressie-Read power divergence family of discrepancies, see Cressie and Read (1984). Imbens, Spady and Johnson (1998) developed estimation methods and test statistics based on the Cressie-Read family for the unconditional moment context, see (2.3) above. Essentially, in the unconditional case, the sample is treated as arising from a discretely distributed population with each data point i , ($i = 1, \dots, n$), treated as a single cell of a

n -cell contingency table. The Cressie-Read discrepancies are then given by

$$\frac{1}{\gamma(\gamma + 1)} \sum_{i=1}^n [(n\pi_i)^{\gamma+1} - 1] \quad (3.1)$$

where expressions are interpreted as limits for $\gamma = -1$ or $\gamma = 0$. In (3.1), π_i is interpreted as the (unconditional) probability associated with observation i , ($i = 1, \dots, n$), and, thus, $\pi_i \geq 0$, ($i = 1, \dots, n$), and $\sum_{i=1}^n \pi_i = 1$. Correspondingly, the (unconditional) moment condition (2.3) becomes $\sum_{i=1}^n \pi_i u(z_i, \beta) = 0$. For a given γ , the Cressie-Read minimum discrepancy (MD) estimator for β_0 minimises the criterion (3.1) with respect to π_i , ($i = 1, \dots, n$), and β subject to the constraints $\sum_{i=1}^n \pi_i u(z_i, \beta) = 0$ and $\sum_{i=1}^n \pi_i = 1$. In effect, the Cressie-Read discrepancy criterion contrasts probabilities π_i , ($i = 1, \dots, n$), which incorporate the moment restrictions $\sum_{i=1}^n \pi_i u(z_i, \beta) = 0$, with their (unrestricted) empirical distribution function counterparts $1/n$, which solve the minimisation problem in the absence of these moment restrictions.

The Cressie-Read discrepancies (3.1) are also known as Renyi's α -class of generalized measures of entropy in which the function $[v^{\gamma+1} - 1]/\gamma(\gamma + 1)$ is a particular form of entropy, see Renyi (1961). For suitable choices for γ many familiar entropy measures are obtained. For example, if $\gamma = 0$, then the entropy takes the standard Shannon form $v \log(v)$, see Shannon and Weaver (1949), whereas $-\log v$ results if $\gamma = -1$. In an interesting recent contribution Kitamura (2005) characterises a class of criteria based on a general entropy function using an information theoretic argument which exploits Fenchel duality, see Borwein and Lewis (1991). Kitamura's (2005) treatment which includes the Cressie-Read family as a special case has as its empirical counterpart Corcoran's (1998) formulation of a general class of MD estimators.²

To adapt the Cressie-Read criteria (3.1) to the conditional moment restrictions (2.1) context, we employ a local version of the Cressie-Read discrepancy family (3.1) similar in spirit to that adopted by KTA for EL and extended for GEL in Smith (2003). In

²Corcoran's (1998) MD criterion is defined by $\sum_{i=1}^n h(n\pi_i)$ where $h(\cdot)$ is a convex function which may also be interpreted as entropy. This general MD criterion rather than the Cressie-Read family could be adopted in the following analysis with little alteration to the results, see fn.3 below. NS, see also Newey and Smith (2001), compared GEL with Corcoran's (1998) MD estimator.

particular, we are interested in the first order large sample properties of the estimator for β_0 obtained from optimising the resultant local information theoretic criterion. We term the estimator a local Cressie-Read MD estimator for β_0 .

To describe the local Cressie-Read MD estimator, let the weights $w_{ij} = \mathcal{K}_{ij} / \sum_{k=1}^n \mathcal{K}_{ik}$, where $\mathcal{K}_{ij} = \mathcal{K}(\frac{x_i - x_j}{b_n})$, $\mathcal{K}(\cdot)$ is a symmetric positive kernel function and b_n a bandwidth parameter, the properties of which are described below in Assumption 4.6 of section 4. Note that $\sum_{j=1}^n w_{ij} = 1$ is automatically satisfied. We need to replace the unconditional weights π_i , ($i = 1, \dots, n$), defining the Cressie-Read criterion (3.1) by their conditional counterparts π_{ij} , ($i, j = 1, \dots, n$), which in a discrete setting may be interpreted as conditional probabilities for cell i given cell j . Correspondingly, the unrestricted empirical distribution function weights n^{-1} are substituted by the data-determined kernel weights w_{ij} , ($i, j = 1, \dots, n$). Finally, to create the local Cressie-Read discrepancy criterion, we smooth the consequently adjusted criterion around each data point i , ($i = 1, \dots, n$), using the kernel weights w_{ij} , ($i, j = 1, \dots, n$), in a manner similar to that used in kernel regression for estimation of a conditional mean based on the Nadaraya-Watson estimator. That is, the local Cressie-Read discrepancy criterion is given from (3.1) by

$$\frac{1}{\gamma(\gamma + 1)} \sum_{i=1}^n \sum_{j=1}^n w_{ij} \left[\left(\frac{\pi_{ij}}{w_{ij}} \right)^{\gamma+1} - 1 \right]. \quad (3.2)$$

Let $u_j(\beta) = u(z_j, \beta)$, ($j = 1, \dots, n$). Also let $\pi_i = (\pi_{i1}, \dots, \pi_{in})'$, ($i = 1, \dots, n$), and $\pi = (\pi'_1, \dots, \pi'_n)'$. The local Cressie-Read MD estimator $\hat{\beta}$ is then given as

$$\hat{\beta} = \arg \inf_{\beta \in \mathcal{B}, \pi} \frac{1}{\gamma(\gamma + 1)} \sum_{i=1}^n \sum_{j=1}^n w_{ij} \left[\left(\frac{\pi_{ij}}{w_{ij}} \right)^{\gamma+1} - 1 \right] \quad (3.3)$$

subject to

$$\sum_{j=1}^n \pi_{ij} u_j(\beta) = 0, \sum_{j=1}^n \pi_{ij} = 1, (i = 1, \dots, n). \quad (3.4)$$

The Lagrangian arising from the local Cressie-Read MD criterion (3.2) corresponding to the associated optimisation problem (3.3) and (3.4) is given by

$$\mathcal{L}(\beta, \pi, \mu, \lambda) = \frac{1}{\gamma(\gamma + 1)} \sum_{i=1}^n \sum_{j=1}^n w_{ij} \left[\left(\frac{\pi_{ij}}{w_{ij}} \right)^{\gamma+1} - 1 \right] \quad (3.5)$$

$$-\sum_{i=1}^n \lambda'_i \sum_{j=1}^n \pi_{ij} u_j(\beta) - \sum_{i=1}^n \mu_i \left(\sum_{j=1}^n \pi_{ij} - 1 \right),$$

where $\mu = (\mu_1, \dots, \mu_n)'$ and $\lambda = (\lambda'_1, \dots, \lambda'_n)'$ collect together the scalar and q -vectors of Lagrangian multipliers associated with the respective constraints $\sum_{j=1}^n \pi_{ij} = 1$ and $\sum_{j=1}^n \pi_{ij} u_j(\beta) = 0$, ($i = 1, \dots, n$).³

Like the local EL estimator discussed by KTA, the local Cressie-Read MD estimator $\hat{\beta}$ possesses a particular normalisation-invariance property. Let $A(x, \beta)$ denote a (q, q) matrix which is non-singular w.p.1 on \mathcal{B} . Clearly, the conditional mean restriction (2.1) remains unaltered by premultiplication of $u(z, \beta_0)$ by $A(x, \beta_0)$. Likewise, the local Cressie-Read MD estimator is invariant to such renormalisations as the additional factor $A(x_i, \beta)$ is merely absorbed into the Lagrange multiplier λ_i , ($i = 1, \dots, n$). This property is lacking for two-step semi-parametric efficient GMM estimators.

Similar to the unconditional moment condition setting examined by Imbens, Spady and Johnson (1998), see also NS and Smith (1997, 2001) for GEL, suitable choices of the parameter γ yield more familiar forms of the local MD criterion (3.2). The local or smoothed empirical likelihood (EL) criterion $-\sum_{i=1}^n \sum_{j=1}^n w_{ij} \log(\pi_{ij}/w_{ij})$ considered by KTA is a special case when $\gamma = -1$, cf. Imbens (1997) and Qin and Lawless (1994), and a local exponential tilting (ET) criterion $\sum_{i=1}^n \sum_{j=1}^n \pi_{ij} \log(\pi_{ij}/w_{ij})$ obtains if $\gamma = 0$, cf. Imbens, Spady and Johnson (1998) and Kitamura and Stutzer (1997). Similarly to the unconditional context, in comparison to EL, ET substitutes the weight π_{ij} for the unrestricted weight w_{ij} . See Imbens, Spady and Johnson (1998, section 2.2, pp.336-338) for a more detailed discussion.

Like NS (Theorem 2.1, p.223), which demonstrates an analogous result for uncondi-

³The Lagrangean based on Corcoran's (1998) general class of discrepancies, see fn.2 above, when adapted for the conditional moment context is given from (3.5) by

$$\begin{aligned} \mathcal{L}(\beta, \pi, \mu, \lambda) &= \sum_{i=1}^n \sum_{j=1}^n w_{ij} h \left(\frac{\pi_{ij}}{w_{ij}} \right) \\ &\quad - \sum_{i=1}^n \lambda'_i \sum_{j=1}^n \pi_{ij} u_j(\beta) - \sum_{i=1}^n \mu_i \left(\sum_{j=1}^n \pi_{ij} - 1 \right). \end{aligned}$$

tional moment restrictions, a local Cressie-Read MD criterion (3.2) similar to that for the continuous updating estimator (CUE), Hansen, Heaton and Yaron (1996), arises in the quadratic case when $\gamma = 1$, see also Bonnal and Renault (2003). Let $\hat{m}(x_i, \beta) = \sum_{j=1}^n w_{ij} u_j(\beta)$ and $\hat{V}(x_i, \beta) = \sum_{j=1}^n w_{ij} u_j(\beta) u_j(\beta)'$, the Nadaraya-Watson estimators of $E[u_i(\beta)|x_i]$ and $E[u_i(\beta)u_i(\beta)|x_i]$ respectively. The local CUE is constructed as

$$\hat{\beta}_{CUE} = \arg \min_{\beta \in \mathcal{B}} \sum_{i=1}^n \hat{m}(x_i, \beta)' [\hat{V}(x_i, \beta) - \hat{m}(x_i, \beta) \hat{m}(x_i, \beta)]^{-1} \hat{m}(x_i, \beta). \quad (3.6)$$

Theorem 3.1 *If $\gamma = 1$ or the local Cressie Read MD criterion (3.2) 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 both $\hat{m}(x_i, \beta)$ and $\hat{V}(x_i, \beta)$, a local GMM estimator is given by⁴

$$\hat{\beta}_{GMM} = \arg \min_{\beta \in \mathcal{B}} \sum_{i=1}^n \hat{m}(x_i, \beta)' \hat{V}(x_i, \tilde{\beta})^{-1} \hat{m}(x_i, \beta), \quad (3.7)$$

where $\tilde{\beta}$ denotes an initial consistent estimator for β_0 ; see, for example, Newey (1990, 1993).⁵

3.1 Implied Probabilities

As in the unconditional context, see NS, p.223, empirical conditional probabilities for the observations may be defined for each member of the Cressie-Read class.

Let $\gamma \neq 0$. For fixed β , the first order conditions for the solution $\hat{\pi}_{ij}(\beta)$ from (3.5) are

$$\frac{1}{\gamma} \left(\frac{\hat{\pi}_{ij}(\beta)}{w_{ij}} \right)^\gamma - \hat{\mu}_i(\beta) - \hat{\lambda}_i(\beta)' u_j(\beta) = 0 \quad (3.8)$$

where $\hat{\mu}_i(\beta)$ and $\hat{\lambda}_i(\beta)$ denote the Lagrange multiplier estimators associated with the respective constraints $\sum_{j=1}^n \hat{\pi}_{ij}(\beta) = 1$ and $\sum_{j=1}^n \hat{\pi}_{ij}(\beta) u_j(\beta) = 0$, ($i = 1, \dots, n$). Solving

⁴A local GMM estimator paralleling (3.6) based on a centred conditional variance matrix estimator is obtained from the minimisation over $\beta \in \mathcal{B}$ of $\sum_{i=1}^n \hat{m}(x_i, \beta)' [\hat{V}(x_i, \tilde{\beta}) - \hat{m}(x_i, \beta) \hat{m}(x_i, \beta)]^{-1} \hat{m}(x_i, \beta)$.

⁵An alternative local CUE more in the spirit of local GMM would minimize $\sum_{i=1}^n \hat{m}(x_i, \beta)' \hat{V}(x_i, \beta)^{-1} \hat{m}(x_i, \beta)$. In contrast to the unconditional moment case, see NS, fn.1, the resultant CUE does not coincide with $\hat{\beta}_{CUE}$.

gives

$$\hat{\pi}_{ij}(\beta) = w_{ij}[\gamma(\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)'u_j(\beta))]^{1/\gamma}. \quad (3.9)$$

The local MD first order condition for $\hat{\mu}_i(\beta)$ is $\sum_{j=1}^n \hat{\pi}_{ij}(\beta) = 1$, from which by (3.9) $\sum_{k=1}^n w_{ik}[\gamma(\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)'u_k(\beta))]^{1/\gamma} = 1$, ($i = 1, \dots, n$). The empirical local MD conditional probabilities are then given by

$$\hat{\pi}_{ij}(\beta) = \frac{w_{ij}(\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)'u_j(\beta))^{1/\gamma}}{\sum_{k=1}^n w_{ik}(\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)'u_k(\beta))^{1/\gamma}}, (i, j = 1, \dots, n). \quad (3.10)$$

If $\gamma = 0$, the first order condition is $\log(\hat{\pi}_{ij}(\beta)/w_{ij}) + 1 - \hat{\mu}_i(\beta) - \hat{\lambda}_i(\beta)'u_j(\beta) = 0$ from which $\hat{\pi}_{ij}(\beta) = \exp(\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)'u_j(\beta))/e$, cf. (3.8) and (3.9). Thus, $\hat{\pi}_{ij}(\beta) = \exp(\hat{\lambda}_i(\beta)'u_j(\beta))/\sum_{k=1}^n \exp(\hat{\lambda}_i(\beta)'u_k(\beta))$, cf. (3.10).

Let $\hat{\pi}_{ij} = \hat{\pi}_{ij}(\hat{\beta})$, ($i, j = 1, \dots, n$), denote the solutions to the MD optimization problem (3.3), (3.4) and $\hat{\mu}_i = \hat{\mu}_i(\hat{\beta})$, and $\hat{\lambda}_i = \hat{\lambda}_i(\hat{\beta})$, ($i = 1, \dots, n$), the Lagrange multiplier estimators along with the corresponding local Cressie-Read MD estimator $\hat{\beta}$. Also let $\hat{u}_j = u_j(\hat{\beta})$, ($j = 1, \dots, n$).

If $\gamma \neq 0$, the empirical conditional probabilities then are defined from (3.9) and (3.10) by

$$\begin{aligned} \hat{\pi}_{ij} &= w_{ij}[\gamma(\hat{\mu}_i + \hat{\lambda}_i'\hat{u}_j)]^{\frac{1}{\gamma}} \\ &= \frac{w_{ij}(\hat{\mu}_i + \hat{\lambda}_i'\hat{u}_j)^{\frac{1}{\gamma}}}{\sum_{k=1}^n w_{ij}(\hat{\mu}_i + \hat{\lambda}_i'\hat{u}_k)^{\frac{1}{\gamma}}}, (j = 1, \dots, n), \end{aligned} \quad (3.11)$$

where $\hat{\mu}_i$ satisfies $\sum_{k=1}^n w_{ij}[\gamma(\hat{\mu}_i + \hat{\lambda}_i'\hat{u}_k)]^{\frac{1}{\gamma}} = 1$. If $\gamma = 0$, $\hat{\pi}_{ij} = w_{ij} \exp(\hat{\lambda}_i'\hat{u}_j)/\sum_{k=1}^n w_{ik} \exp(\hat{\lambda}_i'\hat{u}_k)$, ($j = 1, \dots, n$).

The empirical conditional probabilities $\hat{\pi}_{ij}$ sum to one over $j = 1, \dots, n$, and are positive by construction. They also satisfy the sample moment condition $\sum_{i=1}^n \hat{\pi}_{ij}\hat{u}_j = 0$ when the first order conditions for $\hat{\mu}_i$ and $\hat{\lambda}_i$ hold. For unconditional moment restrictions the (unconditional) probabilities are $\hat{\pi}_i = \exp(\hat{\lambda}'\hat{g}_i)/\sum_{k=1}^n \exp(\hat{\lambda}'\hat{g}_k)$ if $\gamma = 0$ or $(1 + \hat{\lambda}'\hat{g}_i)^{\frac{1}{\gamma}}/\sum_{k=1}^n (1 + \hat{\lambda}'\hat{g}_k)^{\frac{1}{\gamma}}$ otherwise, ($i = 1, \dots, n$), see NS, equation (2.4), where $\hat{g}_i = g(z_i, \hat{\beta})$ from (2.2) and $\hat{\beta}$ denotes an unconditional GMM or GEL estimator. 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 empirical distribution function weights n^{-1} .

For EL, $\hat{\pi}_{ij} = w_{ij}/(1 + \hat{\lambda}'_i \hat{u}_j)$, 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 CUE, $\hat{\pi}_{ij} = w_{ij}(1 + \hat{\lambda}'_i(\hat{u}_j - \hat{m}(x_i, \hat{\beta})))$, cf. Back and Brown (1993), see Bonnal and Renault (2003). See also Brown and Newey (1992, 2002) and Smith (1997).

3.2 First Order Conditions

A re-interpretation of the first order conditions determining the local Cressie-Read MD estimator provides some intuition for why the semi-parametric efficiency lower bound \mathcal{I}^{-1} is achieved by $\hat{\beta}$.

Let $U_j(\beta) = \partial u_j(\beta) / \partial \beta'$, ($j = 1, \dots, n$), and $\hat{D}(x_i, \beta) = \sum_{j=1}^n w_{ij} U_j(\beta)$ the Nadaraya-Watson estimator of $E[U_i(\beta) | x_i]$.

Consider the first order conditions for a semi-parametric efficient GMM estimator $\hat{\beta}_{GMM}$ for β_0 obtained from the GMM minimisation problem (3.7); that is,

$$\sum_{i=1}^n \hat{D}(x_i, \hat{\beta}_{GMM}) \hat{V}(x_i, \tilde{\beta})^{-1} \hat{m}(x_i, \hat{\beta}_{GMM}) = 0. \quad (3.12)$$

An analogous expression to (3.12) can also be provided for any local Cressie-Read MD estimator $\hat{\beta}$, a result which mirrors NS, Theorem 2.3, for the unconditional moment restrictions case.

Let $\hat{k}_{ij} = w_{ij}[\exp(\hat{\lambda}'_i \hat{u}_j) - 1] / \hat{\lambda}'_i \hat{u}_j$ if $\gamma = 0$ and $w_{ij}[(\hat{\mu}_i + \hat{\lambda}'_i \hat{u}_j)^{1/\gamma} - (\hat{\mu}_i)^{1/\gamma}] / \hat{\lambda}'_i \hat{u}_j (\hat{\mu}_i)^{1/\gamma}$ otherwise. Also let $\hat{U}_j = U_j(\hat{\beta})$, ($j = 1, \dots, n$).

Theorem 3.2 *The local Cressie-Read MD first order conditions for $\hat{\beta}$ imply*

$$\sum_{i=1}^n \left[\sum_{j=1}^n \hat{\pi}_{ij} \hat{U}_j \right]' \left[\sum_{j=1}^n \hat{k}_{ij} \hat{u}_j \hat{u}'_j \right]^{-1} \hat{m}(x_i, \hat{\beta}) = 0, \quad (3.13)$$

where $\hat{k}_{ij} = \hat{\pi}_{ij}$ for local EL and $\hat{k}_{ij} = w_{ij} / \hat{\mu}_i$ for local CUE.

A comparison of the first order conditions determining the semi-parametric efficient GMM estimator, (3.12), and those for local Cressie-Read MD, (3.13), is instructive. Similarly to $\hat{\pi}_{ij}$ in (3.11), \hat{k}_{ij} may also be interpreted as an empirical conditional probability.

Lemma B.1 of Appendix B shows that $\hat{\lambda}'_i \hat{u}_j \xrightarrow{p} 0$ and, hence, $\gamma \hat{\mu}_i \xrightarrow{p} 1$ if $\gamma \neq 1$, as $\hat{\mu}_i$ satisfies $\sum_{k=1}^n w_{ij} [\gamma(\hat{\mu}_i + \hat{\lambda}'_i \hat{u}_k)]^{\frac{1}{\gamma}} = 1$. Thus, $\hat{k}_{ij}/w_{ij} \xrightarrow{p} 1$. Therefore, the implicit estimators for the conditional Jacobian and conditional variance matrices are consistent, i.e. $\sum_{j=1}^n \hat{\pi}_{ij} \hat{U}_j \xrightarrow{p} D(x_i)$ and $\sum_{j=1}^n \hat{k}_{ij} \hat{u}_j \hat{u}'_j \xrightarrow{p} V(x_i)$. By comparing the local GMM and local Cressie-Read MD first order conditions, (3.12) and (3.13), it is clear that, asymptotically, local Cressie-Read MD estimators implicitly employ the semi-parametric efficient matrix of instrumental variables $v_*(x, \beta_0) = D(x)'V(x)^{-1}$ and thereby achieve the semi-parametric efficiency lower bound \mathcal{I}^{-1} .

It is also interesting to note that local CUE uses the Nadaraya-Watson kernel regression estimator $\sum_{j=1}^n w_{ij} \hat{u}_j \hat{u}'_j$ for the conditional variance matrix $V(x_i)$ with differential data-determined weights over $i = 1, \dots, n$, that is, the first order conditions for local CUE are $\sum_{i=1}^n \hat{\mu}_i [\sum_{j=1}^n \hat{\pi}_{ij} \hat{U}_j]' \hat{V}(x_i, \hat{\beta})^{-1} \hat{m}(x_i, \hat{\beta}) = 0$, where $\hat{\mu}_i = 1 - \hat{\lambda}'_i \hat{m}(x_i, \hat{\beta})$. In contrast, local EL employs the same implied probabilities $\hat{\pi}_{ij} = w_{ij}/(1 + \hat{\lambda}'_i \hat{u}_j)$ for the estimation of both $V(x_i)$ and the conditional Jacobian matrix $D(x_i)$ which parallels the unconditional case; see NS, Theorem 2.3. The two-step semi-parametric efficient GMM estimator $\hat{\beta}_{GMM}$ described in (3.7) utilises Nadaraya-Watson regression estimators for both conditional Jacobian and variance matrices, see (3.12).⁶

3.3 Duality

For each local MD member we may describe a dual estimator which is similar in spirit to the local GEL estimator considered in Smith (2003). This result mirrors the duality of members of the GEL class for Cressie-Read MD estimators with unconditional moment restrictions given in NS, Theorem 2.2.

⁶Bonnal and Renault (2003) give an alternative expression for the local Cressie-Read MD first order conditions (3.13). Multiplying (3.8) by $\hat{\pi}_{ij}(\beta)u_j(\beta)$ and summing over $j = 1, \dots, n$, $\frac{1}{\gamma} \sum_{j=1}^n w_{ij} [\hat{\pi}_{ij}(\beta)/w_{ij}]^{\gamma+1} u_j(\beta) - [\sum_{j=1}^n \hat{\pi}_{ij}(\beta)u_j(\beta)u_j(\beta)'] \hat{\lambda}_i(\beta) = 0$. Substitution for $\hat{\lambda}_i(\beta)$ into (A.2) yields

$$\sum_{i=1}^n [\sum_{j=1}^n \hat{\pi}_{ij} \hat{U}_j]' [\sum_{j=1}^n \hat{\pi}_{ij} \hat{u}_j \hat{u}'_j]^{-1} \sum_{j=1}^n w_{ij} \left(\frac{\hat{\pi}_{ij}}{w_{ij}} \right)^{\gamma+1} \hat{u}_j = 0,$$

which involves the implicit estimators $\sum_{j=1}^n \hat{\pi}_{ij} \hat{U}_j$ and $\sum_{j=1}^n \hat{\pi}_{ij} \hat{u}_j \hat{u}'_j$ for the conditional Jacobian and variance matrices $D(x_i)$ and $V(x_i)$ but $\hat{m}(x_i, \hat{\beta}) = \sum_{j=1}^n w_{ij} \hat{u}_j$ is replaced by the re-weighted expression $\sum_{j=1}^n (\hat{\pi}_{ij}/w_{ij})^{\gamma+1} w_{ij} \hat{u}_j$.

Let⁷

$$\begin{aligned}
\hat{P}_i(\beta, \mu_i, \lambda_i) &= - \left(\frac{1}{(\gamma + 1)} \sum_{j=1}^n w_{ij} [\gamma(\mu_i + \lambda'_i u_j(\beta))]^{\frac{\gamma+1}{\gamma}} \right. \\
&\quad \left. - [\mu_i - \frac{1}{\gamma(\gamma + 1)}] \right) \text{ if } \gamma \neq -1, 0, \\
&= - \left(\sum_{j=1}^n w_{ij} \log(\mu_i + \lambda'_i u_j(\beta)) - \mu_i + 1 \right) \text{ if } \gamma = -1, \\
&= - \left(\sum_{j=1}^n w_{ij} \exp(\mu_i + \lambda'_i u_j(\beta)) / e - \mu_i - 1/e \right) \text{ if } \gamma = 0.
\end{aligned} \tag{3.14}$$

Theorem 3.3 *If $u(z, \beta)$ is continuously differentiable in β , then the first order conditions to the saddle point problem*

$$\min_{\beta \in \mathcal{B}} \sum_{i=1}^n \sup_{\mu_i, \lambda_i} \hat{P}_i(\beta, \mu_i, \lambda_i) / n \tag{3.15}$$

and to local Cressie-Read MD, (3.3) and (3.4), coincide at the local Cressie-Read MD estimator $\hat{\beta}$, $\hat{\pi}_{ij} = w_{ij} \exp(\hat{\mu}_i + \hat{\lambda}'_i \hat{u}_j) / e$ if $\gamma = 0$ and $w_{ij} [\gamma(\hat{\mu}_i + \hat{\lambda}'_i \hat{u}_j)]^{1/\gamma}$ if $\gamma \neq 0$, ($i, j = 1, \dots, n$), $\hat{\mu}_i$ and $\hat{\lambda}_i$, ($i = 1, \dots, n$).

The dual reformulation offered in (3.14) and (3.15) of the local Cressie-Read optimisation problem is particularly useful in the analysis of the asymptotic properties of the local Cressie-Read MD estimator $\hat{\beta}$ given in Appendix A.⁸

Smith (2003) proposes a different form of dual local GEL criterion

$$\sum_{i=1}^n \sum_{j=1}^n w_{ij} [\rho(\lambda'_i u_j(\beta)) - \rho(0)] / n, \tag{3.16}$$

⁷The criterion (3.14) employs the Nadaraya-Watson estimator $\sum_{j=1}^n w_{ij} [\gamma(\mu_i + \lambda'_i u_j(\beta))]^{\frac{\gamma+1}{\gamma}}$ of the conditional expectation of $[\gamma(\mu_i + \lambda'_i u_i(\beta))]^{\frac{\gamma+1}{\gamma}}$ given x_i , i.e. $E[[\gamma(\mu_i + \lambda'_i u_i(\beta))]^{\frac{\gamma+1}{\gamma}} | x_i]$, ($i = 1, \dots, n$). Hence, the resultant criterion obtained by averaging (3.14) over $i = 1, \dots, n$, utilises a re-scaled and re-centred estimator of the average conditional expectation $\sum_{i=1}^n E[[\gamma(\mu_i + \lambda'_i u_i(\beta))]^{\frac{\gamma+1}{\gamma}} | x_i] / n$.

⁸The first order conditions for μ_i and λ_i are $\sum_{j=1}^n \hat{\pi}_{ij} = 1$ and $\sum_{j=1}^n \hat{\pi}_{ij} \hat{u}_j = 0$. Therefore, if $\gamma \neq -1, 0$, $\hat{\pi}_{ij} = w_{ij} [\gamma(\hat{\mu}_i + \hat{\lambda}'_i \hat{u}_j)]^{1/\gamma} / \sum_{k=1}^n w_{ik} [\gamma(\hat{\mu}_i + \hat{\lambda}'_i \hat{u}_k)]^{1/\gamma}$. If $\gamma = -1$, $\hat{\pi}_{ij} \hat{\mu}_i + \hat{\lambda}'_i (\hat{\pi}_{ij} \hat{u}_j) = w_{ij}$ from which $\hat{\mu}_i = 1$. Therefore, $\hat{\pi}_{ij} = w_{ij} / (1 + \hat{\lambda}'_i \hat{u}_j)$. If $\gamma = 0$, then $\exp(\hat{\mu}_i) \sum_{j=1}^n w_{ij} \exp(\hat{\lambda}'_i \hat{u}_j) = 1$ from which $\hat{\pi}_{ij} = \exp(\hat{\lambda}'_i \hat{u}_j) / \sum_{k=1}^n w_{ik} \exp(\hat{\lambda}'_i \hat{u}_k)$.

to that in (3.14) and (3.15) which has a similar structure to the unconditional case given in NS and Smith (1997, 2001).⁹ For given β , the local Cressie-Read MD first order condition for $\lambda_i(\beta)$ is identical to that from (3.16) when $\rho(v)$ takes the Cressie-Read form $\rho(v) = -(1 + \gamma v)^{(\gamma+1)/\gamma}/(\gamma + 1)$; see (A.1) in Appendix A. However, in contradistinction to the unconditional moment restrictions case described in NS, Theorem 2.2, the first order condition determining the local Cressie-Read MD estimator $\hat{\beta}$ differs from that for the local GEL estimator from (3.16). Let $\rho_1(v) = d\rho(v)/dv$. The first order condition for local GEL is $\sum_{i=1}^n \sum_{j=1}^n \rho_1(\hat{\lambda}_i(\hat{\beta})'u_j(\hat{\beta}))U_j(\hat{\beta})'\hat{\lambda}_i(\hat{\beta}) = 0$ and is similar in structure to that for local Cressie-Read MD $\sum_{i=1}^n \sum_{j=1}^n \hat{\pi}_{ij}(\hat{\beta})U_j(\hat{\beta})'\hat{\lambda}_i(\hat{\beta}) = 0$, see (A.2) in Appendix A. In general, however, for given γ , Cressie-Read $\rho_1(\hat{\lambda}_i(\hat{\beta})'u_j(\hat{\beta}))$ and $\hat{\pi}_{ij}(\hat{\beta})$, see section 3.1, are proportional but depend on i , ($i = 1, \dots, n$). Thus, the first order conditions for β will differ. For local Cressie-Read MD, the implied probabilities are $\hat{\pi}_{ij}(\beta) = w_{ij} \exp(\hat{\lambda}_i(\beta)'u_j(\beta)) / \sum_{k=1}^n \exp(\hat{\lambda}_i(\beta)'u_k(\beta))$ if $\gamma = 0$ and $w_{ij}(\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)'u_j(\beta))^{1/\gamma} / \sum_{k=1}^n (\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)'u_k(\beta))^{1/\gamma}$ if $\gamma \neq 0$. If $\rho(\cdot)$ takes the Cressie-Read form $-(1 + \gamma v)^{(\gamma+1)/\gamma}/(\gamma + 1)$, $\rho_1(\hat{\lambda}_i(\beta)'u_j(\beta)) = w_{ij} \exp(\hat{\lambda}_i(\beta)'u_j(\beta))$ if $\gamma = 0$ and $w_{ij}[1 + \hat{\lambda}_i(\beta)'u_j(\beta)]^{1/\gamma}$ if $\gamma \neq 0$. Local GEL and local Cressie-Read MD employ different self-weighting schemes over observations $i = 1, \dots, n$. It is only when $\gamma = -1$, that is, for local EL, when $\hat{\pi}_{ij}(\beta) = w_{ij}/(1 + \hat{\lambda}_i(\beta)'u_j(\beta))$, that local GEL and local Cressie-Read MD estimators coincide; cf. KTA, section 2.

4 Asymptotic Theory for Estimation

This section gives consistency and asymptotic normality results for the local Cressie-Read MD estimator $\hat{\beta}$. Firstly, however, we provide some regularity conditions for the large sample analysis. Our assumptions by and large follow KTA, Assumptions 3.1-3.7. However, we assume bounded support for the conditioning vector x . This assumption is primarily made for analytical simplicity. It avoids the necessity of the trimming device employed in KTA and Smith (2003) and enables a rather less technically complex devel-

⁹The criterion (3.16) employs a Nadaraya-Watson estimator of the centred average conditional expectation $\sum_{i=1}^n E[\rho(\mu_i + \lambda_i' u_i(\beta)) | x_i] / n - \rho(0)$.

opment. The reader is referred to KTA for the modifications entailed by trimming and for a fuller discussion of these assumptions.¹⁰

Assumption 4.1 (i) $\{x_i, z_i\}_{i=1}^n$ is a random sample on $\mathcal{S} \times \mathcal{R}^d$; (ii) x is continuously distributed with Lebesgue density $h(\cdot)$ whereas the distribution of z is continuous, discrete or mixed; (iii) $u(z, \beta) : \mathcal{R}^d \times \mathcal{B} \rightarrow \mathcal{R}^q$ is a known function.

Assumption 4.2 (i) $\beta_0 \in \text{int}(\mathcal{B})$ is such that $E[u(z, \beta_0)|x] = 0$; (ii) for each $\beta \neq \beta_0$ there exists a set $\mathcal{X}_\beta \subseteq \mathcal{S}$ such that $\mathcal{P}\{x \in \mathcal{X}_\beta\} > 0$ and $E[u(z, \beta)|x] \neq 0$ for all $x \in \mathcal{X}_\beta$; (iii) $\mathcal{B} \subseteq \mathcal{R}^p$ is compact; (iv) $E[\sup_{\beta \in \mathcal{B}} \|u(z, \beta)\|^m] < \infty$ for some $m > 8$.

Assumptions 4.2 (i) and (ii) are the conditional identification condition given in KTA, Assumption 3.1. Together they crucially ensure that $E[\|E[u(z, \beta)|x]\|^2] = 0$ if and only if $\beta = \beta_0$.

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) \mathcal{S} is a proper compact subset of \mathcal{R}^s such that $0 < \inf_{x \in \mathcal{S}} h(x) \leq \sup_{x \in \mathcal{S}} h(x) < \infty$, $h(x)$ is twice continuously differentiable on \mathcal{S} , $\sup_{x \in \mathcal{S}} \|\partial h(x)/\partial x\| < \infty$ and $\sup_{x \in \mathcal{S}} \|\partial^2 h(x)/\partial x \partial x'\| < \infty$; (ii) $u(z, \beta)$ is continuous on \mathcal{B} w.p.1 and $E[\sup_{\beta \in \mathcal{B}} \|\partial u(z, \beta)/\partial \beta'\|] < \infty$; (iii) $\|\partial^2 [E[u^{(i)}(z, \beta)|x]h(x)]/\partial x \partial x'\|$ is uniformly bounded on $\mathcal{B} \times \mathcal{R}^s$, ($i = 1, \dots, q$).

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) $\sup_{x \in \mathcal{R}^s} \|\partial E[D^{(ij)}(x)h(x)]/\partial x\| < \infty$ and $\sup_{(x, \beta) \in \mathcal{R}^s \times \mathcal{B}_0} \|\partial^2 E[D^{(ij)}(x)h(x)]/\partial x \partial x'\| < \infty$; (v) $\sup_{x \in \mathcal{R}^s} \|\partial E[V^{(ij)}(x)h(x)]/\partial x\| < \infty$ and $\sup_{(x, \beta) \in \mathcal{R}^s \times \mathcal{B}_0} \|\partial^2 E[V^{(ij)}(x)h(x)]/\partial x \partial x'\| < \infty$, ($i, j = 1, \dots, q$).

¹⁰Elements of vectors and matrices are denoted by superscripts (i) and (ij) respectively.

Assumption 4.6 *Let $b_n \downarrow 0$ and $\tau \in (0, 1/2)$. Then $n^{1-2\tau-2/m}b_n^{2s} \uparrow \infty$ and $n^{1-2\tau}b_n^{5s/2} \uparrow \infty$.*

As noted by KTA, the parameter τ 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, and used here. Because of the compact support Assumption 4.4 (i), the additional rate assumptions in KTA, Assumption 3.7, arising from trimming and a mild moment existence assumption on the distribution of x are obviated.

These conditions lead to a consistency result for the local Cressie-Read MD estimator $\hat{\beta}$.

Theorem 4.1 *Let Assumptions 4.1-4.6 be satisfied. Then $\hat{\beta} \xrightarrow{p} \beta_0$.*

Asymptotic normality of the local Cressie-Read MD estimator $\hat{\beta}$ requires that the Lagrange multiplier parameters λ_i be restricted to a set which shrinks more slowly than the parametric rate $n^{-1/2}$.

Theorem 4.2 *If Assumptions 4.1-4.6 are satisfied and the Lagrange multiplier parameters λ_i , ($i = 1, \dots, n$), are each constrained to lie in the set $\Lambda_n = \{\lambda : \|\lambda\| \leq Cn^{-1/m}\}$ for some $C > 0$, then $n^{1/2}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, \mathcal{I}^{-1})$.*

Theorem 4.2 emphasises that all local Cressie-Read MD estimators $\hat{\beta}$ are first order equivalent and achieve the semi-parametric efficiency lower bound \mathcal{I}^{-1} confirming the intuition of section 3.2.

5 Information Theoretic Inference

This section is concerned with tests for the validity of the conditional moment restrictions arising from (2.1). TK suggests the use of the optimised local EL criterion as a test statistic. In particular, TK examine a local EL test statistic for the null hypothesis

$$H_0 : \mathcal{P}\{E[u(z, \beta_0)|x] = 0\} = 1 \tag{5.1}$$

against the alternative that H_0 of (5.1) is false. Our concern here is with the efficacy of the optimised scaled local Cressie-Read discrepancy criterion (3.2)

$$\begin{aligned}\mathcal{CR}(\hat{\beta}) &= 2\frac{1}{(\gamma+1)}\sum_{i=1}^n\mathbb{I}(x_i\in S_*)\sum_{j=1}^nw_{ij}\left[\left(\frac{\hat{\pi}_{ij}}{w_{ij}}\right)^{\gamma+1}-1\right] \\ &= 2\frac{1}{(\gamma+1)}\sum_{i=1}^n\mathbb{I}(x_i\in S_*)\sum_{j=1}^nw_{ij}([\gamma(\hat{\mu}_i+\hat{\lambda}'_iu_j(\hat{\beta}))]^{\frac{\gamma+1}{\gamma}}-1),\end{aligned}\tag{5.2}$$

where $\mathbb{I}(\cdot)$ denotes the indicator function. From their respective first order conditions, see (A.1), for a given estimator $\hat{\beta}$ of β_0 , $\hat{\mu}_i = \hat{\mu}_i(\hat{\beta})$ and $\hat{\lambda}_i = \hat{\lambda}_i(\hat{\beta})$ are the respective Lagrange multiplier estimators obtained from the adding-up constraint $\sum_{j=1}^n\hat{\pi}_{ij}(\hat{\beta}) = 1$ and the conditional moment constraint $\sum_{j=1}^n\hat{\pi}_{ij}(\hat{\beta})u_j(\hat{\beta}) = 0$, ($i = 1, \dots, n$), where $\hat{\pi}_{ij}(\beta)$ is given in (3.9), ($i, j = 1, \dots, n$). As in TK, we only require that the estimator $\hat{\beta}$ be $n^{1/2}$ -consistent for β_0 . Particular choices of $\hat{\beta}$ are efficient (or otherwise) GMM or GEL estimators based on unconditional moment restrictions (2.3) or local Cressie-Read MD estimators as described in earlier sections. The set S_* is a compact subset in the support \mathcal{S} of the conditioning variable x . An advantage of the fixed trimming device $\mathbb{I}(x_i \in S_*)$ is that the test statistic may be employed over regions of the sample space of x which are believed to be of particular empirical relevance and, thus, importance.

6 Asymptotic Theory for Inference

We provide an additional regularity condition adapted from TK, Assumption 3.5, which is required for the following limiting distribution theory of the test statistic $\mathcal{CR}(\hat{\beta})$ under the null hypothesis H_0 of (5.1).

Assumption 6.1 (i) $D(x)$ is continuous on \mathcal{S} w.p.1; (ii) $h(x)$ and $V(x)$ are twice continuously differentiable on \mathcal{S} w.p.1; (iii) $\inf_{(\xi,x)\in\mathcal{S}^q\times\mathcal{S}_*}\xi'V(x,\beta)\xi > 0$; (iv) $h(x)$ and $E[\sup_{\beta\in\mathcal{B}}\|u(z,\beta)\|^m|x]h(x)$ are uniformly bounded on \mathcal{S} .

Let $I_i = \mathbb{I}(x_i \in S_*)$. Lemma C.1 in Appendix C demonstrates that we may re-express the local Cressie-Read MD criterion as

$$\mathcal{CR}(\hat{\beta}) = \hat{T}_n + o_p\left[\left(\frac{\log n}{n^{\frac{1}{2}-\frac{1}{m}}b_n^s}\right)^2\right] + o_p\left(\frac{1}{n^{1-2/m}}\right)\tag{6.3}$$

$$+O_p\left[\left(\frac{\log n}{n^{\frac{1}{3}}b_n^s}\right)^{3/2}\right] + O_p\left(\frac{1}{n^{1-4/\eta}}\right),$$

where $\hat{T}_n = \sum_{i=1}^n \mathbf{I}_i \hat{m}(x_i, \hat{\beta})' \hat{V}(x_i, \hat{\beta})^{-1} \hat{m}(x_i, \hat{\beta})$. The leading term \hat{T}_n is a form of local CUE statistic and is identical to that given in TK, (4.1). As detailed in TK, section 4, the statistic \hat{T}_n may be decomposed as $\hat{T}_n = \sum_{k=1}^5 \hat{T}_{n,k}$ where

$$\hat{T}_{n,1} = \mathcal{K}(0)^2 \sum_{i=1}^n \mathbf{I}_i \frac{u_i(\hat{\beta})' \hat{V}(x_i, \hat{\beta})^{-1} u_i(\hat{\beta})}{\left(\sum_{k=1}^n \mathcal{K}_{ik}\right)^2}, \hat{T}_{n,2} = \sum_{i=1}^n \mathbf{I}_i \sum_{j \neq i, j=1}^n w_{ij}^2 u_j(\hat{\beta})' \hat{V}(x_i, \hat{\beta})^{-1} u_j(\hat{\beta}),$$

$$\hat{T}_{n,3} = \mathcal{K}(0) \sum_{i=1}^n \mathbf{I}_i \sum_{j \neq i, j=1}^n \frac{u_i(\hat{\beta})' \hat{V}(x_i, \hat{\beta})^{-1} u_j(\hat{\beta}) w_{ij}}{\sum_{k=1}^n \mathcal{K}_{ik}}, \hat{T}_{n,4} = \hat{T}_{n,3},$$

$$\hat{T}_{n,5} = \sum_{i=1}^n \mathbf{I}_i \sum_{j \neq i, j=1}^n \sum_{k \neq j \neq i, k=1}^n w_{ij} u_j(\hat{\beta})' \hat{V}(x_i, \hat{\beta})^{-1} u_k(\hat{\beta}) w_{ik}.$$

Define $\mathfrak{R}(\mathcal{K}) = \int_{[-1,1]^s} \mathcal{K}(u)^2 du$, $\mathcal{K}^*(v) = \int_{[-1,1]^s} \mathcal{K}(u) \mathcal{K}(v-u) du$ and $\mathcal{K}^{**} = \int_{[-2,2]^s} \mathcal{K}^*(u)^2 du$.

Now, under our regularity conditions, from TK, Lemmata A.2-A.4, under H_0 ,

$$\hat{T}_{n,1} = O_p\left(\frac{1}{nb_n^{2s}}\right), \hat{T}_{n,2} = \frac{1}{b_n^s} [q\mathfrak{R}(\mathcal{K}) \text{vol}(S_*) + O_p\left(\sqrt{\frac{\log n}{nb_n^s}} + b_n^2\right) + o_p\left(\frac{1}{n^{1/2-1/m-1/\eta}}\right)]$$

$$\hat{T}_{n,3} = [O_p\left(\sqrt{\frac{\log n}{nb_n^s}}\right) + o_p\left(\frac{1}{n^{1/2-1/m-1/\eta}}\right)] O_p\left(\sqrt{\frac{1}{nb_n^{3s}}}\right) + O_p\left(\sqrt{\frac{1}{nb_n^{2s}}}\right),$$

and from TK, Lemma A.5, from a CLT due to de Jong (1987),

$$b_n^{s/2} \hat{T}_{n,5} \xrightarrow{d} N(0, \sigma^2),$$

where $\sigma^2 = 2q\mathcal{K}^{**} \text{vol}(S_*)$ and $\text{vol}(S_*) = \int_{S_*} dx$ is the Lebesgue measure of S_* .

Now $b_n^s \hat{T}_{n,2} = q\mathfrak{R}(\mathcal{K}) \text{vol}(S_*) + o_p(1)$ and thus $b_n^{s/2} \hat{T}_{n,2}$ is explosive asymptotically under H_0 . We therefore need to centre $\mathcal{CR}(\hat{\beta})$ by subtracting $\hat{T}_{n,2}$ to obtain an asymptotically pivotal statistic. As noted in TK, section 4, the asymptotic behaviour of $\hat{T}_{n,2}$ does not alter under local alternatives to H_0 .

The following theorem is then immediate from the above discussion and Lemma C.1 of Appendix C.

Theorem 6.1 *Let Assumptions 4.1, 4.2 (i) and (iv) with $m \geq 6$, 4.3, 4.5 (iii) with $\eta \geq 6$ and $E[d(z)^2] < \infty$ and Assumption 6.1 be satisfied. Then, if $\|\hat{\beta} - \beta_0\| = O_p(n^{-1/2})$, S_* is*

a compact subset of \mathcal{S} , $b_n = n^{-\delta}$ for $0 < \delta < \min[\frac{1}{s}(1 - \frac{4}{m}), \frac{1}{3s}]$, and $\lambda_i \in \Lambda_n$, ($i = 1, \dots, n$), $b_n^{s/2}[\mathcal{CR}(\hat{\beta}) - \hat{T}_{n,2}]/\sigma \xrightarrow{d} N(0, 1)$ under H_0 of (5.1).

A test with given asymptotic size is obtained by comparing the statistic $b_n^{s/2}[\mathcal{CR}(\hat{\beta}) - \hat{T}_{n,2}]/\sigma$ to appropriate critical values from the standard normal distribution.

Alternative asymptotically equivalent statistics under H_0 to $b_n^{s/2}[\mathcal{CR}(\hat{\beta}) - \hat{T}_{n,2}]/\sigma$ are the moment-based statistic $b_n^{s/2}[\hat{T}_n - \hat{T}_{n,2}]/\sigma$ and from (C.9) the Lagrange multiplier statistic $b_n^{s/2}[\sum_{i=1}^n \mathbf{1}_i \hat{\lambda}_i' \hat{V}(x_i, \hat{\beta}) \hat{\lambda}_i - \hat{T}_{n,2}]/\sigma$. A potential disadvantage of both of these statistics is their increased dependence on the conditional variance estimator $\hat{V}(x_i, \hat{\beta})$ which may lead to poorer small sample properties than a test based on the statistic $b_n^{s/2}[\mathcal{CR}(\hat{\beta}) - \hat{T}_{n,2}]/\sigma$ in which $\mathcal{CR}(\hat{\beta})$ is self-studentized.

Appendix A: Proofs of Results

Throughout these Appendices, C will denote a generic positive constant that may be different in different uses, and CS, J and T the Cauchy-Schwarz, Jensen 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.

For ease of reference, we collect together some notation used in the text and these Appendices. Let $\hat{m}(x_i, \beta) = \sum_{j=1}^n w_{ij} u_j(\beta)$, $\hat{m}(x) = \hat{m}(x, \beta_0)$, $\hat{D}(x_i, \beta) = \sum_{j=1}^n w_{ij} U_j(\beta)$, $\hat{D}(x) = \hat{D}(x, \beta_0)$, $\hat{V}(x_i, \beta) = \sum_{j=1}^n w_{ij} u_j(\beta) u_j(\beta)'$ and $\hat{V}(x) = \hat{V}(x, \beta_0)$. Also, let $u_{i0} = u_i(\beta_0)$, $U_{j0} = U_j(\beta_0)$, $\hat{\mu}_{i0} = \hat{\mu}_i(\beta_0)$ and $\hat{\lambda}_{i0} = \hat{\lambda}_i(\beta_0)$, ($i = 1, \dots, n$). Finally, $m(x_i, \beta) = E[u_i(\beta)|x_i]$, $\mathcal{I}(\beta) = E_x[D(x, \beta)'[V(x, \beta) - m(x, \beta)m(x, \beta)']^{-1}D(x, \beta)]$ and $\mathcal{I} = \mathcal{I}(\beta_0)$.

Let $c_n = \sqrt{\frac{\log n}{nb_n^s}}$.

Proof of Theorem 3.1: For given β , the first order conditions determining $\hat{\lambda}_i(\beta)$ are

$$\sum_{j=1}^n \hat{\pi}_{ij}(\beta) u_j(\beta) = 0, (i = 1, \dots, n). \quad (\text{A.1})$$

whereas those for the local Cressie-Read MD estimator $\hat{\beta}$ are

$$\sum_{i=1}^n \sum_{j=1}^n \hat{\pi}_{ij}(\hat{\beta}) U_j(\hat{\beta})' \hat{\lambda}_i(\hat{\beta}) = 0. \quad (\text{A.2})$$

When $\gamma = 1$, $\sum_{k=1}^n w_{ik}(\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)' u_k(\beta)) = 1$, ($i = 1, \dots, n$). Therefore, $\hat{\mu}_i(\beta) = 1 - \hat{\lambda}_i(\beta)' \hat{m}(x_i, \beta)$ and, from eq. (3.10),

$$\hat{\pi}_{ij}(\beta) = w_{ij} [1 + \hat{\lambda}_i(\beta)' (u_j(\beta) - \hat{m}(x_i, \beta))], (i, j = 1, \dots, n). \quad (\text{A.3})$$

Substituting eq. (A.3) into eq. (A.1) yields

$$\hat{\lambda}_i(\beta) = -[\hat{V}(x_i, \beta) - \hat{m}(x_i, \beta) \hat{m}(x_i, \beta)']^{-1} \hat{m}(x_i, \beta). \quad (\text{A.4})$$

Therefore, from eqs. (A.2) and (A.4),

$$\sum_{i=1}^n \hat{D}(x_i, \hat{\beta})' [\hat{V}(x_i, \hat{\beta}) - \hat{m}(x_i, \hat{\beta}) \hat{m}(x_i, \hat{\beta})']^{-1} \hat{m}(x_i, \hat{\beta}) = 0.$$

■

Proof of Theorem 3.2: Let $\gamma \neq 0$. By eq. (A.1) and the definition of \hat{k}_{ij} ,

$$\begin{aligned} 0 &= \sum_{j=1}^n \hat{\pi}_{ij} \hat{u}_j = \sum_{j=1}^n (\hat{\pi}_{ij} - w_{ij}) \hat{u}_j + \hat{m}(x_i, \hat{\beta}) \\ &= \sum_{j=1}^n w_{ij} \left(\frac{(\hat{\mu}_i + \hat{\lambda}'_i \hat{u}_j)^{1/\gamma} - (\hat{\mu}_i)^{1/\gamma}}{\hat{\lambda}'_i \hat{u}_j (\hat{\mu}_i)^{1/\gamma}} \right) \hat{u}_j \hat{u}'_j \hat{\lambda}_i + \hat{m}(x_i, \hat{\beta}) \\ &= \sum_{j=1}^n \hat{k}_{ij} \hat{u}_j \hat{u}'_j \hat{\lambda}_i + \hat{m}(x_i, \hat{\beta}). \end{aligned}$$

Solving

$$\hat{\lambda}_i = - \left[\sum_{j=1}^n \hat{k}_{ij} \hat{u}_j \hat{u}'_j \right]^{-1} \hat{m}(x_i, \hat{\beta})$$

gives the first result. If $\gamma = 0$, $\hat{\pi}_{ij} = \exp(\hat{\mu}_i + \hat{\lambda}'_i \hat{u}_j)/e$. Therefore,

$$\begin{aligned} 0 &= \sum_{j=1}^n w_{ij} \left(\frac{\exp(\hat{\lambda}'_i \hat{u}_j) - 1}{\hat{\lambda}'_i \hat{u}_j} \right) \hat{u}_j \hat{u}'_j \hat{\lambda}_i + \hat{m}(x_i, \hat{\beta}) \\ &= \sum_{j=1}^n \hat{k}_{ij} \hat{u}_j \hat{u}'_j \hat{\lambda}_i + \hat{m}(x_i, \hat{\beta}) \end{aligned}$$

and the conclusion then follows as before. Note that for local EL $\hat{k}_{ij} = w_{ij} [-(\hat{\mu}_i + \hat{\lambda}'_i \hat{u}_j)^{-1} + (\hat{\mu}_i)^{-1}] / \hat{\lambda}'_i \hat{u}_j (\hat{\mu}_i)^{-1} = w_{ij} (\hat{\mu}_i + \hat{\lambda}'_i \hat{u}_j)^{-1} = \hat{\pi}_{ij}$ and for local CUE $\hat{k}_{ij} = w_{ij} [(\hat{\mu}_i + \hat{\lambda}'_i \hat{u}_j) - (\hat{\mu}_i)] / \hat{\lambda}'_i \hat{u}_j (\hat{\mu}_i) = w_{ij} / \hat{\mu}_i$. ■

Proof of Theorem 3.3: The first order conditions for local Cressie-Read MD are given in eqs. (A.1) and (A.2) together with the adding-up constraint $\sum_{j=1}^n \hat{\pi}_{ij} = 1$, ($i = 1, \dots, n$). The implied probabilities $\hat{\pi}_{ij}$, ($i, j = 1, \dots, n$), are defined in eq. (3.9).

Let $\gamma \neq -1, 0$. Differentiating (3.14) with respect to μ_i and λ_i for fixed β yields the adding-up constraint $\sum_{j=1}^n w_{ij} [\gamma(\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)' u_j(\beta))]^{1/\gamma} - 1 = 0$ and the conditional moment restriction $\sum_{j=1}^n w_{ij} [\gamma(\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)' u_j(\beta))]^{1/\gamma} u_j(\beta) = 0$, ($i = 1, \dots, n$). Differentiating with respect to β , $\sum_{j=1}^n w_{ij} [\gamma(\hat{\mu}_i(\hat{\beta}) + \hat{\lambda}_i(\hat{\beta})' u_j(\hat{\beta}))]^{1/\gamma} U_j(\hat{\beta})' \hat{\lambda}_i(\hat{\beta}) = 0$, ($i = 1, \dots, n$).

If $\gamma = -1$, the derivatives with respect to μ_i and λ_i for fixed β are respectively $\sum_{j=1}^n w_{ij} (\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)' u_j(\beta))^{-1} - 1 = 0$ and $\sum_{j=1}^n w_{ij} (\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)' u_j(\beta))^{-1} u_j(\beta) = 0$, ($i = 1, \dots, n$). Define $\hat{\pi}_{ij}(\beta) = w_{ij} (\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)' u_j(\beta))^{-1}$. Hence, $\hat{\pi}_{ij}(\beta) \hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)' \hat{\pi}_{ij}(\beta) u_j(\beta) = w_{ij}$. Summing over $j = 1, \dots, n$, $\hat{\mu}_i(\beta) = 1$ and, thus, $\hat{\pi}_{ij}(\beta) = w_{ij} (1 + \hat{\lambda}_i(\beta)' u_j(\beta))^{-1}$. Differentiating with respect to β yields $\sum_{j=1}^n \hat{\pi}_{ij}(\hat{\beta}) U_j(\hat{\beta})' \hat{\lambda}_i(\hat{\beta}) = 0$, ($i = 1, \dots, n$).

If $\gamma = 0$, the derivatives with respect to μ_i and λ_i for fixed β are $\sum_{j=1}^n w_{ij} \exp(\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)' u_j(\beta))/e - 1 = 0$ and $\sum_{j=1}^n w_{ij} \exp(\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)' u_j(\beta)) u_j(\beta)/e = 0$ respectively. Let $\hat{\pi}_{ij}(\beta) = w_{ij} \exp(\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)' u_j(\beta))$. Differentiating with respect to β yields $\sum_{j=1}^n \hat{\pi}_{ij}(\hat{\beta}) U_j(\hat{\beta})' \hat{\lambda}_i(\hat{\beta}) = 0$, ($i = 1, \dots, n$).

The result of the theorem then follows. ■

Proof of Theorem 4.1: Let $C_n = \{z \in \mathcal{R}^d : \sup_{\beta \in \mathcal{B}} \|u(z, \beta)\| \leq Cn^{1/m}\}$ and $u_{nj}(\beta) = \mathbf{I}_j u_j(\beta)$, where $\mathbf{I}_j = \mathbf{I}(z_j \in C_n)$. Define $\bar{\lambda}_i(\beta) = -E[g_i(\beta)|x_i]/(1 + \|E[g_i(\beta)|x_i]\|)$. Then,

$$\begin{aligned} \sum_{i=1}^n \sup_{\mu_i, \lambda_i} \hat{P}_i(\beta, \mu_i, \lambda_i)/n &\geq Q_n(\beta) \\ &= -\sum_{i=1}^n \left(\frac{1}{(\gamma+1)} \sum_{j=1}^n w_{ij} [\gamma(\bar{\mu}_i(\beta) + n^{-1/m} \bar{\lambda}_i(\beta)' u_{nj}(\beta))]^{\frac{\gamma+1}{\gamma}} \right. \\ &\quad \left. - [\bar{\mu}_i(\beta) - \frac{1}{\gamma(\gamma+1)}] \right) / n \end{aligned} \quad (\text{A.5})$$

where $\bar{\mu}_i(\beta)$ solves $\sum_{j=1}^n w_{ij} [\gamma(\bar{\mu}_i(\beta) + n^{-1/m} \bar{\lambda}_i(\beta)' u_{nj}(\beta))]^{\frac{1}{\gamma}} = 1$. Note that $n^{-1/m} \bar{\lambda}_i(\beta) \in \Lambda_n$ defined in Theorem 4.2.

Now

$$(\bar{\mu}_i(\beta) + n^{-1/m} \bar{\lambda}_i(\beta)' u_{nj}(\beta))^{\frac{\gamma+1}{\gamma}} = (\bar{\mu}_i(\beta))^{\frac{\gamma+1}{\gamma}} + \frac{\gamma+1}{\gamma} (\bar{\mu}_i(\beta))^{\frac{1}{\gamma}} n^{-1/m} \bar{\lambda}_i(\beta)' u_j(\beta) + \bar{r}_{nij}(t)$$

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

$$\begin{aligned} \bar{r}_{nij}(t) &= \frac{\gamma+1}{\gamma} (\bar{\mu}_i(\beta))^{\frac{1}{\gamma}} n^{-1/m} \bar{\lambda}_i(\beta)' u_j(\beta) (1 - \mathbf{I}_j) \\ &\quad + n^{-1/m} \bar{\lambda}_i(\beta)' u_{nj}(\beta) \left[\frac{\gamma+1}{\gamma} (\bar{\mu}_i(\beta) + t n^{-1/m} \bar{\lambda}_i(\beta)' u_{nj}(\beta))^{\frac{1}{\gamma}} - \frac{\gamma+1}{\gamma} (\bar{\mu}_i(\beta))^{\frac{1}{\gamma}} \right]. \end{aligned} \quad (\text{A.6})$$

From Lemma B.1 $\sup_{\beta \in \mathcal{B}, n^{-1/m} \bar{\lambda}_i \in \Lambda_n, 1 \leq j \leq n} \left| (\bar{\mu}_i(\beta) + t n^{-1/m} \bar{\lambda}_i(\beta)' u_{nj}(\beta))^{\frac{1}{\gamma}} - (\bar{\mu}_i(\beta))^{\frac{1}{\gamma}} \right| \xrightarrow{p} 0$. Also $\max_{1 \leq j \leq n} (1 - \mathbf{I}_j) = o_p(1)$ from KTA, Lemma D.2. As $\bar{\mu}_i(\beta)$ solves $\sum_{j=1}^n w_{ij} [\gamma(\bar{\mu}_i(\beta) + n^{-1/m} \bar{\lambda}_i(\beta)' u_{nj}(\beta))]^{\frac{1}{\gamma}} = 1$, a similar expansion yields

$$\begin{aligned} 1 &= \sum_{j=1}^n w_{ij} [(\gamma \bar{\mu}_i(\beta))^{\frac{1}{\gamma}} + \gamma^{\frac{1}{\gamma}-1} (\bar{\mu}_i(\beta) + t n^{-1/m} \bar{\lambda}_i(\beta)' u_{nj}(\beta))^{\frac{1}{\gamma}-1} n^{-1/m} \bar{\lambda}_i(\beta)' u_{nj}(\beta)] \\ &= (\gamma \bar{\mu}_i(\beta))^{\frac{1}{\gamma}} + [\gamma(\bar{\mu}_i(\beta) + o_p(1))]^{\frac{1}{\gamma}-1} (1 + o_p(1)) n^{-1/m} \bar{\lambda}_i(\beta)' \hat{m}(x_i, \beta) \\ &= (\gamma \bar{\mu}_i(\beta))^{\frac{1}{\gamma}} + (\gamma \bar{\mu}_i(\beta))^{\frac{1}{\gamma}-1} O_p(n^{-1/m}) \end{aligned}$$

uniformly i and $\beta \in \mathcal{B}$ as $\max_{1 \leq i \leq n} \sup_{\beta \in \mathcal{B}} \|\hat{m}(x_i, \beta) - E[u_i(\beta)|x_i]\| = o_p(1)$, cf. KTA, Proof of Lemma B.8. Therefore, as $\gamma \bar{\mu}_i(\beta) = 1 + O_p(n^{-1/m})$,

$$\gamma \bar{\mu}_i(\beta) = [1 - n^{-1/m} \bar{\lambda}_i(\beta)' \hat{m}(x_i, \beta) + o_p(n^{-1/m})]^\gamma \quad (\text{A.7})$$

uniformly i and $\beta \in \mathcal{B}$.

Hence, from eqs. (A.6) and (A.7),

$$\begin{aligned} n^{1/m} \sum_{i=1}^n \sum_{j=1}^n w_{ij} \bar{r}_{nij}(t)/n &\leq o_p(1) \sum_{i=1}^n \bar{\lambda}_i(\beta)' \hat{m}(x_i, \beta)/n + o_p(1) \sum_{i=1}^n \bar{\lambda}_i(\beta)' \hat{m}(x_i, \beta)/n \\ &\quad - o_p(1) \sum_{i=1}^n \bar{\lambda}_i(\beta)' \hat{m}(x_i, \beta) (1 - \mathbf{I}_j)/n \\ &= o_p(1) \sum_{i=1}^n \bar{\lambda}_i(\beta)' \hat{m}(x_i, \beta)/n \end{aligned}$$

[A.3]

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

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

as $\hat{m}(x_i, \beta) \xrightarrow{p} E[u_i(\beta)|x_i]$ uniformly i and $\beta \in \mathcal{B}$ from above. Substituting into eq. (A.5),

$$\begin{aligned} Q_n(\beta) &= - \sum_{i=1}^n \left(\frac{1}{(\gamma+1)} \sum_{j=1}^n w_{ij} ([(\gamma \bar{\mu}_i(\beta))^{\frac{\gamma+1}{\gamma}} + \gamma^{\frac{1}{\gamma}} (\gamma+1) (\bar{\mu}_i(\beta))^{\frac{1}{\gamma}} n^{-1/m} \bar{\lambda}_i(\beta)' u_j(\beta))] \right. \\ &\quad \left. - [\bar{\mu}_i(\beta) - \frac{1}{\gamma(\gamma+1)}] \right) / n + o_p(n^{-1/m}) \end{aligned} \quad (\text{A.8})$$

uniformly $\beta \in \mathcal{B}$. Now, from (A.7)

$$(\gamma \bar{\mu}_i(\beta))^{\frac{\gamma+1}{\gamma}} = 1 - (\gamma+1) n^{-1/m} \bar{\lambda}_i(\beta)' \hat{m}(x_i, \beta) + o_p(n^{-1/m}) \quad (\text{A.9})$$

uniformly i and $\beta \in \mathcal{B}$. Therefore, substituting (A.9) into (A.8),

$$n^{1/m} Q_n(\beta) = - \sum_{i=1}^n n^{-1/m} \bar{\lambda}_i(\beta)' \hat{m}(x_i, \beta) / n + o_p(1).$$

From KTA, eqs. (A.4), (A.5) and Lemma B.8,

$$n^{1/m} \sup_{\beta \in \mathcal{B}} |Q_n(\beta) - \bar{Q}_n(\beta)| = o_p(1), \quad (\text{A.10})$$

see KTA, Proof of Theorem 3.1, where

$$n^{1/m} \bar{Q}_n(\beta) = - \sum_{i=1}^n \bar{\lambda}_i(\beta)' E[u_i(\beta)|x_i] / n. \quad (\text{A.11})$$

Thus, as in KTA, eq. (A.6), from eqs. (A.5) and (A.10),

$$n^{1/m} \inf_{\beta \in \mathcal{B}} \sum_{i=1}^n \sup_{\mu_i, \lambda_i} \hat{P}_i(\beta, \mu_i, \lambda_i) / n \geq n^{1/m} \inf_{\beta \in \mathcal{B}} \bar{Q}(\beta) + o_p(1). \quad (\text{A.12})$$

From the definition of $\bar{\lambda}_i(\beta)$, ($i = 1, \dots, n$), and (A.11), a UWL gives

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

uniformly $\beta \in \mathcal{B}$; see KTA, eq. (A.7). The function $E[\|E[u_i(\beta)|x_i]\|^2 / (1 + \|E[u_i(\beta)|x_i]\|)] = E[\mathbf{I}(x_i \in \mathcal{X}_\beta) \|E[u_i(\beta)|x_i]\|^2 / (1 + \|E[u_i(\beta)|x_i]\|)]$ is continuous in β , has a unique zero β_0 and is strictly positive for all $\beta \neq \beta_0$ by Assumptions 4.2 (i) and (ii).

Now

$$0 \leq n^{1/m} \sum_{i=1}^n \sup_{\mu_i, \lambda_i} \hat{P}_i(\hat{\beta}, \mu_i, \lambda_i) / n \leq n^{1/m} \sum_{i=1}^n \sup_{\mu_i, \lambda_i} \hat{P}_i(\beta_0, \mu_i, \lambda_i) / n, \quad (\text{A.13})$$

[A.4]

as $\sum_{i=1}^n \hat{P}_i(\hat{\beta}, \frac{1}{\gamma}, 0)/n = 0$. By concavity, evaluating $[\gamma(\mu_i + \lambda'_i u_j(\beta))]^{\frac{\gamma+1}{\gamma}}$ at $\gamma\mu_i = 1$ and $\lambda_i = 0$,

$$-[\gamma(\mu_i + \lambda'_i u_j(\beta))]^{\frac{\gamma+1}{\gamma}} \leq -1 - \frac{\gamma+1}{\gamma} [\gamma(\mu_i + \lambda'_i u_j(\beta)) - 1].$$

Hence,

$$\begin{aligned} n^{1/m} \sum_{i=1}^n \sup_{\mu_i, \lambda_i} \hat{P}_i(\beta_0, \mu_i, \lambda_i)/n &= n^{1/m} \sum_{i=1}^n \hat{P}(\beta_0, \mu_i(\beta_0), \lambda_i(\beta_0))/n & (A.14) \\ &= -n^{1/m} \sum_{i=1}^n \left(\frac{1}{(\gamma+1)} \sum_{j=1}^n w_{ij} [\gamma(\mu_i(\beta_0) + \lambda_i(\beta_0)' u_j(\beta_0))]^{\frac{\gamma+1}{\gamma}} \right. \\ &\quad \left. - [\mu_i(\beta_0) - \frac{1}{\gamma(\gamma+1)}] \right) /n \\ &\leq -n^{1/m} \sum_{i=1}^n \lambda_i(\beta_0)' \hat{m}(x_i) /n \\ &= n^{1/m} [o_p(c_n)]^2 \\ &= o_p(1), \end{aligned}$$

by Assumption 4.6, eq. (B.3) of Lemma B.2 and TK, Lemma C.1, cf. KTA, Lemma B.3. Therefore, combining (A.13) and (A.14),

$$n^{1/m} \sum_{i=1}^n \sup_{\mu_i, \lambda_i} \hat{P}_i(\hat{\beta}, \mu_i, \lambda_i)/n = o_p(1). \quad (A.15)$$

By T and (A.12)

$$\begin{aligned} 0 &\leq E[-\bar{\lambda}_i(\beta)' E[u_i(\beta)|x_i]]|_{\beta=\hat{\beta}} & (A.16) \\ &\leq \sup_{\beta \in \mathcal{B}} \left| n^{1/m} \bar{Q}_n(\beta) - E[-\bar{\lambda}_i(\beta)' E[u_i(\beta)|x_i]] \right| + \left| n^{1/m} \bar{Q}_n(\hat{\beta}) \right| \\ &\leq \left| n^{1/m} \sum_{i=1}^n \sup_{\mu_i, \lambda_i} \hat{P}_i(\hat{\beta}, \mu_i, \lambda_i)/n \right| + o_p(1) \end{aligned}$$

as $n^{1/m} \sum_{i=1}^n \sup_{\mu_i, \lambda_i} \hat{P}_i(\beta, \mu_i, \lambda_i)/n \geq n^{1/m} \bar{Q}_n(\beta) + o_p(1)$ uniformly $\beta \in \mathcal{B}$ from eqs. (A.5) and (A.10). Hence, from eqs. (A.15) and (A.16), $E[-\bar{\lambda}_i(\beta)' E[u_i(\beta)|x_i]]|_{\beta=\hat{\beta}} = o_p(1)$. 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 Cressie-Read MD estimator $\hat{\beta}$; viz. $n^{-1} \sum_{i=1}^n \partial \hat{P}_i(\hat{\beta}, \hat{\mu}_i(\hat{\beta}), \hat{\lambda}_i(\hat{\beta}))/\partial \beta = 0$. By a Taylor expansion about β_0 ,

$$0 = n^{-1/2} \sum_{i=1}^n \frac{\partial \hat{P}_i(\beta_0, \hat{\mu}_{i0}, \hat{\lambda}_{i0})}{\partial \beta} + n^{-1} \sum_{i=1}^n \frac{\partial^2 \hat{P}_i(\beta^*, \hat{\mu}_i(\beta^*), \hat{\lambda}_i(\beta^*))}{\partial \beta \partial \beta'} n^{1/2} (\hat{\beta} - \beta_0)$$

[A.5]

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.2),

$$n^{-1/2} \sum_{i=1}^n \frac{\partial \hat{P}_i(\beta_0, \hat{\mu}_{i0}, \hat{\lambda}_{i0})}{\partial \beta} = n^{1/2} \hat{A} + n^{-1/2} \sum_{i=1}^n \sum_{j=1}^n w_{ij} [\gamma(\hat{\mu}_{i0} + \hat{\lambda}'_{i0} u_{j0})]^{\frac{1}{\gamma}} U'_{j0} r_i, \quad (\text{A.17})$$

where

$$\hat{A} = \sum_{i=1}^n \left[\sum_{j=1}^n w_{ij} [\gamma(\hat{\mu}_{i0} + \hat{\lambda}'_{i0} u_{j0})]^{\frac{1}{\gamma}} U'_{j0} \hat{V}(x_i)^{-1} \hat{m}(x_i) \right] / n. \quad (\text{A.18})$$

From Lemma B.1, $\sup_{\beta \in \mathcal{B}, \lambda_i \in \Lambda_n, 1 \leq j \leq n} \left| [\gamma(\hat{\mu}_{i0} + \lambda'_i u_j(\beta))]^{\frac{1}{\gamma}} - (\gamma \hat{\mu}_{i0})^{\frac{1}{\gamma}} \right| = o_p(1)$. Moreover, by eq. (B.5), $\gamma \hat{\mu}_{i0} = 1 + o_p(1)$ uniformly i . Therefore, w.p.a.1,

$$\begin{aligned} n^{-1/2} \left\| \sum_{i=1}^n \sum_{j=1}^n w_{ij} [\gamma(\hat{\mu}_{i0} + \hat{\lambda}'_{i0} u_{j0})]^{\frac{1}{\gamma}} U'_{j0} r_i \right\| &\leq O_p(n^{1/2}) \max_{1 \leq i \leq n} \|r_i\| \sum_{i=1}^n \sum_{j=1}^n w_{ij} c(z_j) / n \\ &= o_p(n^{1/2+1/m} c_n^2) \\ &= o_p(1) \end{aligned} \quad (\text{A.19})$$

by Assumption 4.6, uniformly i and $\beta \in \mathcal{B}_0$.

Let

$$\hat{A} = A + \sum_{i=1}^n \left[\sum_{j=1}^n w_{ij} ([\gamma(\hat{\mu}_{i0} + \hat{\lambda}'_{i0} u_{j0})]^{\frac{1}{\gamma}} - (\gamma \hat{\mu}_{i0})^{\frac{1}{\gamma}}) U_{j0}' \hat{V}(x_i)^{-1} \hat{m}(x_i) \right] / n, \quad (\text{A.20})$$

from (A.17), where

$$A = \sum_{i=1}^n \left[\sum_{j=1}^n w_{ij} (\gamma \hat{\mu}_{i0})^{\frac{1}{\gamma}} U_{j0}' \hat{V}(x_i)^{-1} \hat{m}(x_i) \right] / n. \quad (\text{A.21})$$

Now,

$$\begin{aligned} \left\| \sum_{j=1}^n w_{ij} ([\gamma(\hat{\mu}_{i0} + \hat{\lambda}'_{i0} u_{j0})]^{\frac{1}{\gamma}} - (\gamma \hat{\mu}_{i0})^{\frac{1}{\gamma}}) U_{j0}' \right\| &\leq O_p(1) \left\| \hat{\lambda}_{i0} \right\| \\ &\times \sum_{j=1}^n w_{ij} \sup_{\beta \in \mathcal{B}} \|u_j(\beta)\| c(z_j) \end{aligned} \quad (\text{A.22})$$

uniformly i and $\beta \in \mathcal{B}_0$. Moreover, as $\max_{1 \leq i \leq n} \left\| \hat{V}(x_i)^{-1} \right\| = O_p(1)$, cf. KTA, Lemma B.7 and TK, Lemma C.2 (ii), and $\max_{1 \leq i \leq n} \left\| \sum_{j=1}^n w_{ij} \sup_{\beta \in \mathcal{B}} \|u_j(\beta)\| c(z_j) \right\| = O_p(1)$ by Assumptions 4.2 and 4.5 (iii), from (A.19)-(A.21),

$$\begin{aligned} n^{1/2} \left\| \hat{A} - A \right\| &\leq O_p(n^{1/2}) \max_{1 \leq i \leq n} \|\hat{m}(x_i)\| \left[\sum_{i=1}^n \left\| \hat{\lambda}_{i0} \right\| \right] / n \\ &= O_p(n^{1/2} c_n^2) = o_p(1), \end{aligned} \quad (\text{A.23})$$

[A.6]

from eqs. (B.3), (B.5) and by TK, Lemma C.1. Therefore, substituting (A.18) and (A.22) into (A.16), $n^{1/2} \sum_{j=1}^n \partial \hat{P}_i(\beta_0, \hat{\mu}_{i0}, \hat{\lambda}_{i0}) / \partial \beta = n^{1/2} A + o_p(1)$. The result follows from Lemma B.3, the continuity of $\mathcal{I}(\beta)$ on \mathcal{B}_0 from Assumption 4.5 (i) and $n^{1/2} A \xrightarrow{d} N(0, \mathcal{I})$ from CLT by KTA, Lemma B.2, as $(\gamma \hat{\mu}_{i0})^{\frac{1}{\gamma}} = 1 + o_p(1)$ uniformly i . ■

Appendix B: Auxiliary Results for Estimation

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

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

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 that

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

Therefore, the result follows. ■

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

Lemma B.2 *Let Assumptions 4.1-4.5 be satisfied and $\lambda_i \in \Lambda_n$, ($i = 1, \dots, n$). Also let $\frac{\log n}{n^{1-4/m} b_n^s} \downarrow 0$ and $b_n \downarrow 0$. Then $\hat{\lambda}_{i0} = -\hat{V}(x_i)^{-1} \hat{m}(x_i) + r_i$, where $\max_{1 \leq i \leq n} \|r_i\| = o_p(n^{1/m} c_n^2)$.*

Proof. From eq. (A.1)

$$\begin{aligned} 0 &= \sum_{j=1}^n w_{ij} [\gamma(\hat{\mu}_{i0} + \hat{\lambda}'_{i0} u_{j0})]^{\frac{1}{\gamma}} u_{j0} \\ &= (\gamma \hat{\mu}_{i0})^{\frac{1}{\gamma}} \hat{m}(x_i) + (\gamma \hat{\mu}_{i0})^{\frac{1}{\gamma}-1} \hat{V}(x_i) \hat{\lambda}_{i0} + r_{1i}(t), \end{aligned} \tag{B.1}$$

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

$$r_{1i}(t) = \sum_{j=1}^n w_{ij} ([\gamma(\hat{\mu}_{i0} + t \hat{\lambda}'_{i0} u_{j0})]^{\frac{1}{\gamma}-1} - (\gamma \hat{\mu}_{i0})^{\frac{1}{\gamma}-1}) u_{j0} u'_{j0} \hat{\lambda}_{i0}.$$

From Lemma B.1 $\sup_{\beta \in \mathcal{B}, \lambda_i \in \Lambda_n, 1 \leq j \leq n} \left| [\gamma(\mu_i + t \lambda'_i u_j(\beta))]^{\frac{1}{\gamma}-1} - (\gamma \mu_i)^{\frac{1}{\gamma}-1} \right| \xrightarrow{p} 0$. Thus, $r_{1i}(t) = o_p(1) \hat{V}(x_i) \hat{\lambda}_{i0}$ uniformly i and j and

$$\begin{aligned} \|r_{1i}(t)\| &\leq o_p(1) \max_{1 \leq j \leq n} \|u_{j0}\| \left\| \hat{m}(x_i)' \hat{\lambda}_{i0} \right\| \\ &\leq o_p(n^{1/m}) \|\hat{m}(x_i)\| \left\| \hat{\lambda}_{i0} \right\| \\ &= o_p(n^{1/m}) [O_p(c_n)] \left\| \hat{\lambda}_{i0} \right\| \end{aligned} \tag{B.2}$$

[A.7]

where the second inequality follows from CS and $\max_{1 \leq j \leq n} \|u_{j0}\| = o_p(n^{1/m})$ by KTA, Lemma D.2, and the equality by TK, Lemma C.1, from Assumption 4.4 (i).

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

$$\begin{aligned} 0 &= \|\hat{\lambda}_{i0}\| \sum_{j=1}^n w_{ij} [\gamma(\hat{\mu}_{i0} + \hat{\lambda}'_{i0} u_{j0})]^{\frac{1}{\gamma}} \hat{\xi}'_{i0} u_{j0} \\ &= (\gamma \hat{\mu}_{i0})^{\frac{1}{\gamma}} \|\hat{\lambda}_{i0}\| \hat{\xi}'_{i0} \hat{m}(x_i) + (\gamma \hat{\mu}_{i0})^{\frac{1}{\gamma}-1} \|\hat{\lambda}_{i0}\|^2 (1 + o_p(1)) \hat{\xi}'_{i0} \hat{V}(x_i) \hat{\xi}_{i0} \end{aligned}$$

uniformly i and j . As $\max_{1 \leq i \leq n} \|\hat{V}(x_i) - V(x_i)\| = O_p(c_n) + O_p(b_n^2)$, cf. TK, Lemma C.2 (i), $\max_{1 \leq i \leq n} \|\hat{V}(x_i)\| = O_p(1)$. Hence, $\hat{\xi}'_{i0} \hat{V}(x_i) \hat{\xi}_{i0}$ is bounded below w.p.a.1 by Assumption 4.5 (ii). Solving

$$\begin{aligned} \|\hat{\lambda}_{i0}\| / (\gamma \hat{\mu}_{i0}) &= -\hat{\xi}'_{i0} \hat{m}(x_i) / [(1 + o_p(1)) \hat{\xi}'_{i0} \hat{V}(x_i) \hat{\xi}_{i0}] \\ &= O_p(c_n) \end{aligned} \tag{B.3}$$

uniformly i by TK, Lemma C.1. An expansion of $\sum_{j=1}^n w_{ij} [\gamma(\hat{\mu}_{i0} + \hat{\lambda}'_{i0} u_{j0})]^{\frac{1}{\gamma}} = 1$ yields

$$\begin{aligned} 1 &= \sum_{j=1}^n w_{ij} [(\gamma \hat{\mu}_{i0})^{\frac{1}{\gamma}} + \gamma^{\frac{1}{\gamma}-1} (\hat{\mu}_{i0} + t \hat{\lambda}'_{i0} u_{j0})^{\frac{1}{\gamma}-1} \hat{\lambda}'_{i0} u_{j0}] \\ &= (\gamma \hat{\mu}_{i0})^{\frac{1}{\gamma}} + \gamma^{\frac{1}{\gamma}-1} (\hat{\mu}_{i0} + o_p(1))^{\frac{1}{\gamma}-1} \hat{\lambda}'_{i0} \hat{m}(x_i) \\ &= (\gamma \hat{\mu}_{i0})^{\frac{1}{\gamma}} + (\gamma \hat{\mu}_{i0})^{\frac{1}{\gamma}-1} (1 + o_p(1)) \hat{\lambda}'_{i0} \hat{m}(x_i) \\ &= (\gamma \hat{\mu}_{i0})^{\frac{1}{\gamma}} [1 + (1 + o_p(1)) \|\hat{\lambda}_{i0}\| / (\gamma \hat{\mu}_{i0})] \hat{\xi}'_{i0} \hat{m}(x_i) \end{aligned} \tag{B.4}$$

uniformly i . Therefore, from eqs. (B.3), (B.4) and TK, Lemma C.1,

$$\begin{aligned} \gamma \hat{\mu}_{i0} &= 1 / [1 + (1 + o_p(1)) \|\hat{\lambda}_{i0}\| / (\gamma \hat{\mu}_{i0})] \hat{\xi}'_{i0} \hat{m}(x_i)]^\gamma \\ &= 1 + O_p(c_n^2) \end{aligned} \tag{B.5}$$

uniformly i .

Therefore, from eqs. (B.2), (B.3) and (B.5)

$$\|r_{1i}(t)\| = o_p(n^{1/m} c_n^2) \tag{B.6}$$

uniformly i .

By Assumption 4.5 (ii), from eqs. (B.1), (B.5) and (B.6), as $\max_{1 \leq i \leq n} \|\hat{V}(x_i)^{-1}\| = O_p(1)$, cf. TK, Lemma C.2 (ii),

$$\begin{aligned} \hat{\lambda}_{i0} &= -\hat{V}(x_i)^{-1} \hat{m}(x_i) + \hat{V}(x_i)^{-1} r_{1i}(t) \\ &= -\hat{V}(x_i)^{-1} \hat{m}(x_i) + r_i. \end{aligned}$$

■

Lemmata B.3-B.8 given below mirror KTA, Lemmas C.1-C.6. Our proofs follow closely those in KTA.

[A.8]

Lemma B.3 *Let Assumptions 4.2-4.6 hold and $\lambda_i \in \Lambda_n$, ($i = 1, \dots, n$). Then*

$$\sup_{\beta \in \mathcal{B}_0} \left\| n^{-1} \sum_{i=1}^n \partial^2 \hat{P}_i(\beta, \hat{\mu}_i(\beta), \hat{\lambda}_i(\beta)) / \partial \beta \partial \beta' - \mathcal{I}(\beta) \right\| = o_p(1).$$

Proof. As $\sum_{j=1}^n w_{ij}(\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)'u_j(\beta))^{\frac{1}{\gamma}} = 1$ and $\sum_{j=1}^n w_{ij}(\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)'u_j(\beta))^{\frac{1}{\gamma}} u_j(\beta) = 0$, ($i = 1, \dots, n$), for all $\beta \in \mathcal{B}$ from (A.2),

$$n^{-1} \sum_{i=1}^n \frac{\partial \hat{P}_i(\beta, \hat{\mu}_i(\beta), \hat{\lambda}_i(\beta))}{\partial \beta} = - \sum_{i=1}^n \sum_{j=1}^n w_{ij} [\gamma(\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)'u_j(\beta))]^{\frac{1}{\gamma}} U_j(\beta)' \hat{\lambda}_i(\beta) / n. \quad (\text{B.7})$$

Therefore, $n^{-1} \sum_{i=1}^n \partial^2 \hat{P}_i(\beta, \hat{\mu}_i(\beta), \hat{\lambda}_i(\beta)) / \partial \beta \partial \beta' = T_1(\beta) + T_2(\beta) + T_3(\beta)$ where

$$\begin{aligned} T_1(\beta) &= - \sum_{i=1}^n \sum_{j=1}^n w_{ij} [\gamma(\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)'u_j(\beta))]^{\frac{1}{\gamma}-1} U_j(\beta)' \hat{\lambda}_i(\beta) \partial[\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)'u_j(\beta)] / \partial \beta' / n, \\ T_2(\beta) &= - \sum_{i=1}^n \sum_{j=1}^n w_{ij} [\gamma(\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)'u_j(\beta))]^{\frac{1}{\gamma}} [U_j(\beta)' \partial \hat{\lambda}_i(\beta) / \partial \beta'] / n, \\ T_3(\beta) &= - \sum_{i=1}^n \sum_{j=1}^n w_{ij} [\gamma(\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)'u_j(\beta))]^{\frac{1}{\gamma}} \left[\sum_{k=1}^q \hat{\lambda}_i^{(k)}(\beta) \partial^2 u_j^{(k)}(\beta) / \partial \beta \partial \beta' \right] / n. \end{aligned}$$

The desired result follows from Lemmata B.4-B.6 given below. ■

Lemma B.4 *If Assumptions 4.2-4.6 are satisfied and $\lambda_i \in \Lambda_n$, ($i = 1, \dots, n$), then $\sup_{\beta \in \mathcal{B}_0} \|T_1(\beta)\| = o_p(1)$.*

Proof. As $\partial[\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)'u_j(\beta)] / \partial \beta' = \partial \hat{\mu}_i(\beta) / \partial \beta' + \hat{\lambda}_i(\beta)'U_j(\beta) + g_j(\beta)' \partial \hat{\lambda}_i(\beta) / \partial \beta'$, $T_1(\beta) = T_{1,a}(\beta) + T_{1,b}(\beta) + T_{1,c}(\beta)$, where

$$\begin{aligned} T_{1,a}(\beta) &= - \sum_{i=1}^n \sum_{j=1}^n w_{ij} [\gamma(\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)'u_j(\beta))]^{\frac{1}{\gamma}-1} U_j(\beta)' \hat{\lambda}_i(\beta) \hat{\lambda}_i(\beta)' U_j(\beta) / n, \\ T_{1,b}(\beta) &= - \sum_{i=1}^n \sum_{j=1}^n w_{ij} [\gamma(\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)'u_j(\beta))]^{\frac{1}{\gamma}-1} U_j(\beta)' \hat{\lambda}_i(\beta) \partial \hat{\mu}_i(\beta) / \partial \beta' / n, \\ T_{1,c}(\beta) &= - \sum_{i=1}^n \sum_{j=1}^n w_{ij} [\gamma(\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)'u_j(\beta))]^{\frac{1}{\gamma}-1} U_j(\beta)' \hat{\lambda}_i(\beta) u_j(\beta)' \partial \hat{\lambda}_i(\beta) / \partial \beta' / n. \end{aligned}$$

A similar expansion to that leading to eq. (B.5) results in

$$1 = [\gamma \hat{\mu}_i(\beta)]^{\frac{1}{\gamma}} (1 + (1 + o_p(1)) \hat{\lambda}_i(\beta)' \hat{m}(x_i, \beta) / [\gamma \hat{\mu}_i(\beta)]).$$

Now, $\hat{m}(x_i, \beta) = o_p(n^{1/m})$ uniformly i and $\beta \in \mathcal{B}_0$ by KTA, Lemma D.5. Hence, as $\hat{\lambda}_i(\beta) \in \Lambda_n$, $\gamma \hat{\mu}_i(\beta) = 1 + o_p(1)$ uniformly i and $\beta \in \mathcal{B}_0$ by Lemma B.1.

[A.9]

By Lemma B.1, eq. (B.5), and Assumption 4.5 (iii),

$$\begin{aligned} \sup_{\beta \in \mathcal{B}_0} \|T_{1,a}(\beta)\| &\leq o_p(1) \sum_{i=1}^n \sup_{\beta \in \mathcal{B}_0} |\gamma \hat{\mu}_i(\beta) + o_p(1)|^{\frac{1}{\gamma}-1} \sum_{j=1}^n w_{ij} c(z_j)^2 / n \\ &= o_p(1) \end{aligned}$$

as $\sum_{j=1}^n w_{ij} c(z_j)^2 = O_p(1)$ uniformly i by KTA, Lemma D.4. Also,

$$\begin{aligned} \sup_{\beta \in \mathcal{B}_0} \|T_{1,b}(\beta)\| &\leq o_p(1) \sum_{i=1}^n \sup_{\beta \in \mathcal{B}_0} \|\partial \hat{\mu}_i(\beta) / \partial \beta'\| \sum_{j=1}^n w_{ij} c(z_j) / n \\ &= o_p(1) \sum_{i=1}^n \sup_{\beta \in \mathcal{B}_0} \|\partial \hat{\mu}_i(\beta) / \partial \beta'\| / n \end{aligned}$$

uniformly i and $\beta \in \mathcal{B}_0$ since $\sum_{j=1}^n w_{ij} c(z_j) = O_p(1)$. Similarly,

$$\begin{aligned} \sup_{\beta \in \mathcal{B}_0} \|T_{1,c}(\beta)\| &\leq o_p(1) \sum_{i=1}^n \sup_{\beta \in \mathcal{B}_0} \left\| \partial \hat{\lambda}_i(\beta) / \partial \beta' \right\| \sum_{j=1}^n w_{ij} c(z_j) \|u_j(\beta)\| / n \\ &= o_p(1) \sum_{i=1}^n \sup_{\beta \in \mathcal{B}_0} \left\| \partial \hat{\lambda}_i(\beta) / \partial \beta' \right\| / n \end{aligned}$$

since $\sum_{j=1}^n w_{ij} c(z_j) \|u_j(\beta)\| \leq (\sum_{j=1}^n w_{ij} c(z_j)^2)^{1/2} (\sum_{j=1}^n w_{ij} \|u_j(\beta)\|^2)^{1/2} = O_p(1)$ uniformly i and $\beta \in \mathcal{B}_0$ by CS and J. Now, as $\sum_{j=1}^n w_{ij} [\gamma(\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)' u_j(\beta))]^{\frac{1}{\gamma}} = 1$, $\sum_{j=1}^n w_{ij} [\gamma(\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)' u_j(\beta))]^{\frac{1}{\gamma}-1} \partial[\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)' u_j(\beta)] / \partial \beta = 0$. Therefore, from Lemma B.1, $[\gamma \hat{\mu}_i(\beta)]^{\frac{1}{\gamma}-1} \partial \hat{\mu}_i(\beta) / \partial \beta' = -[\gamma \hat{\mu}_i(\beta)]^{\frac{1}{\gamma}-1} \sum_{j=1}^n w_{ij} \partial[\hat{\lambda}_i(\beta)' u_j(\beta)] / \partial \beta + o_p(1)$ uniformly i and $\beta \in \mathcal{B}_0$. From Lemma B.7 below, $\sup_{\beta \in \mathcal{B}_0} \sum_{i=1}^n \left\| \partial \hat{\lambda}_i(\beta) / \partial \beta' \right\| / n = O_p(1)$ and, thus, likewise $\sup_{\beta \in \mathcal{B}_0} \sum_{i=1}^n \|\partial \hat{\mu}_i(\beta) / \partial \beta'\| / n = O_p(1)$. Hence, $\sup_{\beta \in \mathcal{B}_0} \|T_{1,b}(\beta)\| \leq o_p(1)$ and $\sup_{\beta \in \mathcal{B}_0} \|T_{1,c}(\beta)\| \leq o_p(1)$. ■

Lemma B.5 *If Assumptions 4.2-4.6 are satisfied and $\lambda_i \in \Lambda_n$, ($i = 1, \dots, n$), 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 $\sup_{\beta \in \mathcal{B}, \lambda_i \in \Lambda_n, 1 \leq j \leq n} \left| [\gamma(\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)' u_j(\beta))]^{\frac{1}{\gamma}} - [\gamma \hat{\mu}_i(\beta)]^{\frac{1}{\gamma}} \right| = o_p(1)$ from Lemma B.1 and $\gamma \hat{\mu}_i(\beta) = 1 + o_p(1)$ uniformly i and $\beta \in \mathcal{B}_0$,

$$\begin{aligned} &\sum_{i=1}^n \sum_{j=1}^n w_{ij} [\gamma(\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)' u_j(\beta))]^{\frac{1}{\gamma}} [U_j(\beta)' \partial \hat{\lambda}_i(\beta) / \partial \beta'] / n \\ &= (1 + o_p(1)) \sum_{i=1}^n [D(x_i, \beta) + o_p(1)]' \partial \hat{\lambda}_i(\beta) / \partial \beta' / n, \end{aligned}$$

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

$$n^{-1} \sup_{\beta \in \mathcal{B}_0} \left\| \sum_{i=1}^n D(x_i, \beta)' \partial \hat{\lambda}_i(\beta) / \partial \beta' - \sum_{i=1}^n D(x_i, \beta)' [V(x_i, \beta) - m(x_i, \beta) m(x_i, \beta)']^{-1} D(x_i, \beta) \right\| = o_p(1).$$

[A.10]

Therefore, similarly to below KTA, eq. (C.3),

$$\begin{aligned} & \sum_{i=1}^n \sum_{j=1}^n w_{ij} [\gamma(\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)' u_j(\beta))]^{\frac{1}{\gamma}} [U_j(\beta)' \partial \hat{\lambda}_i(\beta) / \partial \beta'] / n \\ &= \sum_{i=1}^n D(x_i, \beta)' [V(x_i, \beta) - m(x_i, \beta) m(x_i, \beta)']^{-1} D(x_i, \beta) / n + o_p(1), \end{aligned}$$

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 $\lambda_i \in \Lambda_n$, ($i = 1, \dots, n$) are satisfied, then $\sup_{\beta \in \mathcal{B}_0} \|T_3(\beta)\| = o_p(1)$.*

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

$$\begin{aligned} n^{-1} \sup_{\beta \in \mathcal{B}_0} \left\| \sum_{i=1}^n \sum_{j=1}^n w_{ij} [\gamma(\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)' u_j(\beta))]^{\frac{1}{\gamma}} \left[\sum_{k=1}^q \hat{\lambda}_i^{(k)}(\beta) \partial^2 u_j^{(k)}(\beta) / \partial \beta \partial \beta' \right] \right\| &\leq o_p(1) \\ &\times \sum_{i=1}^n \sum_{j=1}^n w_{ij} d(z_j) / n \\ &= o_p(1). \end{aligned}$$

■

Lemma B.7 *If Assumptions 4.2-4.6 are satisfied and $\lambda_i \in \Lambda_n$, ($i = 1, \dots, n$), then, for each i and $\beta \in \mathcal{B}_0$,*

$$\begin{aligned} \partial \hat{\lambda}_i(\beta) / \partial \beta' &= -[V(x_i, \beta) - m(x_i, \beta) m(x_i, \beta)']^{-1} D(x_i, \beta) + M_{1,i}(\beta) D(x_i, \beta) \\ &\quad + M_{2,i}(\beta) E[c(z_i) | x_i] + M_{3,i}(\beta) \sum_{j=1}^n w_{ij} c(z_j) + M_{4,i}(\beta), \end{aligned}$$

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. Differentiating (A.2) with respect to β , we have

$$\begin{aligned} & \sum_{j=1}^n w_{ij} [\gamma(\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)' u_j(\beta))]^{\frac{1}{\gamma}-1} u_j(\beta) \partial [\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)' u_j(\beta)] / \partial \beta' \\ &= - \sum_{j=1}^n w_{ij} [\gamma(\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)' u_j(\beta))]^{\frac{1}{\gamma}} U_j(\beta). \end{aligned} \tag{B.8}$$

From the Proof of Lemma B.4, as $\gamma \mu_i(\beta) = 1 + o_p(1)$,

$$\begin{aligned} \partial \hat{\mu}_i(\beta) / \partial \beta' &= - \sum_{j=1}^n w_{ij} \partial [\hat{\lambda}_i(\beta)' u_j(\beta)] / \partial \beta' + o_p(1) \\ &= - \hat{m}(x_i, \beta)' \partial \hat{\lambda}_i(\beta) / \partial \beta' - \hat{\lambda}_i(\beta)' \hat{D}(x_i, \beta) + o_p(1) \end{aligned}$$

[A.11]

uniformly i and $\beta \in \mathcal{B}_0$. Hence,

$$\sum_{j=1}^n w_{ij} [\gamma(\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)' u_j(\beta))]^{\frac{1}{\gamma}-1} u_j(\beta) \partial \hat{\mu}_i(\beta) / \partial \beta' = -(1 + o_p(1)) \hat{m}(x_i, \beta) \times \\ [\hat{m}(x_i, \beta)' \partial \hat{\lambda}_i(\beta) / \partial \beta' + \hat{\lambda}_i(\beta)' \hat{D}(x_i, \beta) + o_p(1)]$$

uniformly i and $\beta \in \mathcal{B}_0$. Therefore, the left hand side of eq. (B.8) becomes

$$(1 + o_p(1)) \sum_{j=1}^n w_{ij} [u_j(\beta) u_j(\beta)' - \hat{m}(x_i, \beta) \hat{m}(x_i, \beta)'] \partial \hat{\lambda}_i(\beta) / \partial \beta' \quad (\text{B.9}) \\ + (1 + o_p(1)) \sum_{j=1}^n w_{ij} [u_j(\beta) \hat{\lambda}_i(\beta)' [U_j(\beta) - \hat{D}(x_i, \beta)] + o_p(1)]$$

uniformly i and $\beta \in \mathcal{B}_0$.

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

$$\max_{1 \leq i \leq n} \sup_{\beta \in \mathcal{B}_0} \left\| \hat{V}(x_i, \beta) - V(x_i, \beta) \right\| = o_p(1).$$

Similarly,

$$\max_{1 \leq i \leq n} \sup_{\beta \in \mathcal{B}_0} \left\| \hat{m}(x_i, \beta) - m(x_i, \beta) \right\| = o_p(1).$$

Thus, from Assumption 4.5 (ii),

$$\left(\sum_{j=1}^n w_{ij} [u_j(\beta) u_j(\beta)' - \hat{m}(x_i, \beta) \hat{m}(x_i, \beta)'] \right)^{-1} = [V(x_i, \beta) - m(x_i, \beta) m(x_i, \beta)']^{-1} + R_{1,i}(\beta), \quad (\text{B.10})$$

where $\max_{1 \leq i \leq n} \sup_{\beta \in \mathcal{B}_0} \|R_{1,i}(\beta)\| = o_p(1)$; cf. KTA, eq. (C.6).

By Lemma B.1, Assumptions 4.5 (iii) and $\hat{\lambda}_i(\beta) \in \Lambda_n$, from KTA, Lemma D.2,

$$\left\| \sum_{j=1}^n w_{ij} u_j(\beta) \hat{\lambda}_i(\beta)' U_j(\beta) \right\| \leq \max_{1 \leq j \leq n} \|u_j(\beta)\| \left\| \hat{\lambda}_i(\beta) \right\| \sum_{j=1}^n w_{ij} c(z_j) \\ = o_p(1) \sum_{j=1}^n w_{ij} c(z_j),$$

uniformly i and $\beta \in \mathcal{B}_0$. Likewise, $\left\| \sum_{j=1}^n w_{ij} u_j(\beta) \hat{\lambda}_i(\beta)' \hat{D}(x_i, \beta) \right\| \leq \max_{1 \leq j \leq n} \|u_j(\beta)\| \left\| \hat{\lambda}_i(\beta) \right\| \sum_{j=1}^n w_{ij} c(z_j) = o_p(1) \sum_{j=1}^n w_{ij} c(z_j)$. Therefore, for $\max_{1 \leq i \leq n} \sup_{\beta \in \mathcal{B}_0} \|R_{4,i}(\beta)\| = o_p(1)$,

$$\sum_{j=1}^n w_{ij} u_j(\beta) \hat{\lambda}_i(\beta)' [U_j(\beta) - \hat{D}(x_i, \beta)] = R_{4,i}(\beta) \sum_{j=1}^n w_{ij} c(z_j), \quad (\text{B.11})$$

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

[A.12]

Finally, by Lemma B.1, $\gamma\hat{\mu}_i(\beta) = 1 + o_p(1)$ uniformly i and $\beta \in \mathcal{B}_0$ and KTA, Lemma B.5, an argument like that for KTA, eq. (C.7), for the right hand side of eq. (B.8) yields

$$\sum_{j=1}^n w_{ij}[\gamma(\hat{\mu}_i(\beta) + \hat{\lambda}_i(\beta)'u_j(\beta))]^{\frac{1}{\gamma}}U_j(\beta) = D(x_i, \beta) + E[c(z_i)|x_i]R_{2,i}(\beta) + R_{3,i}(\beta), \quad (\text{B.12})$$

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

The desired result obtains from firstly substituting eqs. (B.10), (B.11) into eq. (B.9), then the subsequent result and eq. (B.12) into eq. (B.8) and finally solving for $\partial\hat{\lambda}_i(\beta)/\partial\beta'$. ■

Lemma B.8 *Let Assumptions 4.4, 4.5 and $\lambda_i \in \Lambda_n$, ($i = 1, \dots, n$), hold. Then*

$$\sup_{\beta \in \mathcal{B}_0} \left\| n^{-1} \sum_{i=1}^n \partial\hat{P}_i(\beta, \hat{\mu}_i(\beta), \hat{\lambda}_i(\beta))/\partial\beta \right\| = o_p(1).$$

Proof. From eq. (B.7), Assumption 4.5(ii) and $\lambda_i \in \Lambda_n$, ($i = 1, \dots, n$),

$$\begin{aligned} \sup_{\beta \in \mathcal{B}_0} \left\| n^{-1} \sum_{i=1}^n \partial\hat{P}_i(\beta, \hat{\mu}_i(\beta), \hat{\lambda}_i(\beta))/\partial\beta \right\| &\leq o_p(1) \sum_{j=1}^n w_{ij}c(z_j) \\ &= o_p(1). \end{aligned}$$

■

Appendix C: Auxiliary Results for Inference

Let $u_{*j} = \sup_{\beta \in \mathcal{B}} \|u(z_j, \beta)\|$, $\mathbf{I}_i = \mathbf{I}(x_i \in S_*)$ and $\mathbf{I}_* = \{1 \leq i \leq n : x_i \in S_*\}$.

The following Lemma parallels TK, Lemma A.1, which provides a similar result for the local EL statistic.

Lemma C.1 *Let Assumptions 4.1, 4.2 (i) and (iv) with $m \geq 6$, 4.3, 4.5 (iii) with $\eta \geq 6$ and $E[d(z)^2] < \infty$ and Assumption 6.1 be satisfied. Also let $\lambda_i \in \Lambda_n$, ($i = 1, \dots, n$). Then, if $b_n = n^{-\delta}$ for $0 < \delta < \frac{1}{s}(1 - \frac{4}{m})$,*

$$\begin{aligned} \mathcal{CR}(\hat{\beta}) &= \hat{T}_n + o_p(n^{1+2/m}c_n^4) + o_p(1/n^{1-2/m}) \\ &\quad + O_p(nc_n^3) + O_p(1/n^{1-4/\eta}), \end{aligned}$$

where $\hat{T}_n = \sum_{i=1}^n \mathbf{I}_i \hat{m}(x_i, \hat{\beta})' \hat{V}(x_i, \hat{\beta})^{-1} \hat{m}(x_i, \hat{\beta})$.

Proof. From eq. (A.1)

$$\begin{aligned} 0 &= \sum_{j=1}^n w_{ij}[\gamma(\hat{\mu}_i + \hat{\lambda}_i' \hat{u}_j)]^{\frac{1}{\gamma}} \hat{u}_j \\ &= (\gamma\hat{\mu}_i)^{\frac{1}{\gamma}} \hat{m}(x_i, \hat{\beta}) + (\gamma\hat{\mu}_i)^{\frac{1}{\gamma}-1} \hat{V}(x_i, \hat{\beta}) \hat{\lambda}_i + \hat{r}_{1,i} \end{aligned} \quad (\text{C.1})$$

[A.13]

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

$$\hat{r}_{1,i} = \sum_{j=1}^n w_{ij} [\gamma(\hat{\mu}_i + t\hat{\lambda}'_i \hat{u}_j)]^{\frac{1}{\gamma}-1} - (\gamma\hat{\mu}_i)^{\frac{1}{\gamma}-1} \hat{u}_j \hat{u}'_j \hat{\lambda}_i$$

From Lemma B.1 $\sup_{\beta \in \mathcal{B}, \lambda_i \in \Lambda_n, 1 \leq j \leq n} \left| [\gamma(\hat{\mu}_i + t\lambda_i u_j(\beta))]^{\frac{1}{\gamma}-1} - (\gamma\hat{\mu}_i)^{\frac{1}{\gamma}-1} \right| \xrightarrow{p} 0$. Thus, $\hat{r}_{1,i} = o_p(1) \sum_{j=1}^n w_{ij} \hat{u}_j \hat{u}'_j \hat{\lambda}_i$ uniformly i and j and

$$\begin{aligned} \|\hat{r}_{1,i}\| &\leq o_p(1) \max_{1 \leq j \leq n} \|\hat{u}_j\| \left\| \hat{\lambda}'_i \hat{m}(x_i, \hat{\beta}) \right\| \\ &\leq o_p(n^{1/m}) \left\| \hat{m}(x_i, \hat{\beta}) \right\| \left\| \hat{\lambda}_i \right\| \end{aligned} \quad (\text{C.2})$$

where the second inequality follows from CS and KTA, Lemma D.2. Now, from Assumption 4.5 (iii), w.p.a.1

$$\begin{aligned} \mathbb{I}_i \left\| \hat{m}(x_i, \hat{\beta}) \right\| &\leq \max_{i \in \mathbb{I}_*} \|\hat{m}(x_i)\| + \left\| \hat{\beta} - \beta_0 \right\| \sum_{j=1}^n w_{ij} c(z_j) \\ &= O_p(c_n) + O_p\left(\frac{1}{n^{1/2-1/\eta}}\right) \end{aligned}$$

uniformly i . Hence,

$$\mathbb{I}_i \|\hat{r}_{1,i}\| \leq o_p(n^{1/m}) \left[\max_{i \in \mathbb{I}_*} \|\hat{m}(x_i)\| + \left\| \hat{\beta} - \beta_0 \right\| \sum_{j=1}^n w_{ij} c(z_j) \right] \mathbb{I}_i \left\| \hat{\lambda}_i \right\|. \quad (\text{C.3})$$

Let $\hat{\xi}_i = \hat{\lambda}_i / \left\| \hat{\lambda}_i \right\|$. Then,

$$\begin{aligned} 0 &= \left\| \hat{\lambda}_i \right\| \sum_{j=1}^n w_{ij} [\gamma(\hat{\mu}_i + \hat{\lambda}'_i \hat{u}_j)]^{\frac{1}{\gamma}} \hat{\xi}'_i \hat{u}_j \\ &= (\gamma\hat{\mu}_i)^{\frac{1}{\gamma}} \left\| \hat{\lambda}_i \right\| \hat{\xi}'_i \hat{m}(x_i, \hat{\beta}) + (\gamma\hat{\mu}_i)^{\frac{1}{\gamma}-1} \left\| \hat{\lambda}_i \right\|^2 (1 + o_p(1)) \hat{\xi}'_i \hat{V}(x_i, \hat{\beta}) \hat{\xi}_i \end{aligned}$$

uniformly i and j . W.p.a.1 $\hat{\xi}'_i \hat{V}(x_i, \hat{\beta}) \hat{\xi}_i$ is bounded below from KTA, Lemma B.6, by Assumption 4.5(ii). Hence, solving $\left\| \hat{\lambda}_i \right\| / (\gamma\hat{\mu}_i) = -\hat{\xi}'_i \hat{m}(x_i, \hat{\beta}) / [(1 + o_p(1)) \hat{\xi}'_i \hat{V}(x_i, \hat{\beta}) \hat{\xi}_i]$. Therefore, by KTA, Lemma D.5, and TK, Lemma C.1,

$$\begin{aligned} \mathbb{I}_i \left\| \hat{\lambda}_i \right\| / (\gamma\hat{\mu}_i) &= \left[\max_{i \in \mathbb{I}_*} \|\hat{m}(x_i)\| + \left\| \hat{\beta} - \beta_0 \right\| \sum_{j=1}^n w_{ij} c(z_j) \right] \\ &= O_p(c_n) + O_p\left(\frac{1}{n^{1/2-1/\eta}}\right), \end{aligned} \quad (\text{C.4})$$

uniformly i and j . An expansion of $\sum_{j=1}^n w_{ij} [\gamma(\hat{\mu}_i + \hat{\lambda}'_i \hat{u}_j)]^{\frac{1}{\gamma}} = 1$ yields

$$1 = (\gamma\hat{\mu}_i)^{\frac{1}{\gamma}} [1 + (1 + o_p(1)) \left\| \hat{\lambda}_i \right\| / (\gamma\hat{\mu}_i)] \hat{\xi}'_i \hat{m}(x_i, \hat{\beta}) \quad (\text{C.5})$$

uniformly i where, from (C.4), the $o_p(1)$ term is $O_p(c_n^2) + O_p(1/n^{1-2/\eta})$. From eq. (C.4), solving,

$$\begin{aligned} \mathbf{I}_i(\gamma\hat{\mu}_i) &= \mathbf{I}_i[1 - (1 + o_p(1))[\mathbf{I}_i \|\hat{\lambda}_i\| / (\gamma\hat{\mu}_i)] \hat{\xi}'_i \hat{m}(x_i, \hat{\beta})]^\gamma \\ &= \mathbf{I}_i + o_p(1) \end{aligned} \quad (\text{C.6})$$

uniformly i . Therefore, substituting eq. (C.6) into eq. (C.4),

$$\begin{aligned} \mathbf{I}_i \|\hat{\lambda}_i\| &= [\max_{i \in \mathbf{I}_*} \|\hat{m}(x_i)\| + \|\hat{\beta} - \beta_0\| \sum_{j=1}^n w_{ij} c(z_j)] \\ &= O_p(c_n) + O_p\left(\frac{1}{n^{1/2-1/\eta}}\right). \end{aligned} \quad (\text{C.7})$$

Thus, from eq. (C.3), by J,

$$\begin{aligned} \mathbf{I}_i \|\hat{r}_{1,i}\| &= o_p(n^{1/m}) [(\max_{i \in \mathbf{I}_*} \|\hat{m}(x_i)\|)^2 + \|\hat{\beta} - \beta_0\|^2 \sum_{j=1}^n w_{ij} c(z_j)^2] \\ &= o_p(n^{1/m} c_n^2) + o_p(n^{-1+1/m+2/\eta}) = o_p(1) \end{aligned} \quad (\text{C.8})$$

uniformly $i \in \mathbf{I}_*$, cf. TK, eq. (A.7). From eq. (C.6), as $\max_{i \in \mathbf{I}_*} \|\hat{V}(x_i, \hat{\beta})^{-1}\| = O_p(1)$ by TK, Lemma C.2(ii), eq. (C.1) becomes

$$\mathbf{I}_i \hat{\lambda}_i = -\mathbf{I}_i \hat{V}(x_i, \hat{\beta})^{-1} \hat{m}(x_i, \hat{\beta}) + \mathbf{I}_i \hat{r}_{2,i} \quad (\text{C.9})$$

with the asymptotic properties of the remainder term $\mathbf{I}_i \|\hat{r}_{2,i}\|$ identical to those of $\mathbf{I}_i \|\hat{r}_{1,i}\|$ given in (C.8).

Now,

$$[\gamma(\hat{\mu}_i + \hat{\lambda}'_i \hat{u}_j)]^{\frac{\gamma+1}{\gamma}} = (\gamma\hat{\mu}_i)^{\frac{\gamma+1}{\gamma}} + (\gamma+1)(\gamma\hat{\mu}_i)^{\frac{1}{\gamma}} (\hat{\lambda}'_i \hat{u}_j) + \frac{1}{2}(\gamma+1)(\gamma\hat{\mu}_i)^{\frac{1}{\gamma}-1} (\hat{\lambda}'_i \hat{u}_j)^2 + \hat{r}_{3,ij}$$

where $|\hat{r}_{3,ij}| \leq C |\hat{\lambda}'_i \hat{u}_j|^3 \leq C \|\hat{\lambda}_i\|^3 \|\hat{u}_j\|^3 \leq C \|\hat{\lambda}_i\|^3 (u_{*j})^3$. Therefore, from eq. (C.9), w.p.a.1

$$\begin{aligned} \mathbf{I}_i \sum_{j=1}^n w_{ij} [\gamma(\hat{\mu}_i + \hat{\lambda}'_i \hat{u}_j)]^{\frac{\gamma+1}{\gamma}} &= [\mathbf{I}_i(\gamma\hat{\mu}_i)]^{\frac{\gamma+1}{\gamma}} + (\gamma+1) [\mathbf{I}_i(\gamma\hat{\mu}_i)]^{\frac{1}{\gamma}} [\mathbf{I}_i \hat{\lambda}'_i \hat{m}(x_i, \hat{\beta})] \\ &\quad + \frac{1}{2}(\gamma+1) [\mathbf{I}_i(\gamma\hat{\mu}_i)]^{\frac{1}{\gamma}-1} (\mathbf{I}_i \hat{\lambda}_i)' \hat{V}(x_i, \hat{\beta}) (\mathbf{I}_i \hat{\lambda}_i) + \mathbf{I}_i \sum_{j=1}^n w_{ij} \hat{r}_{3,ij}. \end{aligned} \quad (\text{C.10})$$

As $\mathbf{I}_i(\gamma\hat{\mu}_i) = \mathbf{I}_i[1 - (1 + o_p(1))\mathbf{I}_i \hat{\lambda}'_i \hat{m}(x_i, \hat{\beta})]^\gamma$, from eq. (C.9), w.p.a.1

$$\mathbf{I}_i(\gamma\hat{\mu}_i)^{\frac{\gamma+1}{\gamma}} = \mathbf{I}_i - (1 + o_p(1))(\gamma+1) \mathbf{I}_i \hat{\lambda}'_i \hat{m}(x_i, \hat{\beta}) + \mathbf{I}_i \hat{r}_{4,i}$$

where $\mathbf{I}_i |\hat{r}_{4,i}| \leq C \left| \mathbf{I}_i \hat{\lambda}'_i \hat{m}(x_i, \hat{\beta}) \right|^2$. Note that, from (C.4), the $o_p(1)$ term is $O_p(c_n^2) + O_p(1/n^{1-2/\eta})$. From below eq. (C.2) and eq. (C.7), $\left| \mathbf{I}_i \hat{\lambda}'_i \hat{m}(x_i, \hat{\beta}) \right| \leq \left\| \mathbf{I}_i \hat{\lambda}_i \right\| \left\| \mathbf{I}_i \hat{m}(x_i, \hat{\beta}) \right\| = O_p(c_n^2) + O_p(1/n^{1-2/\eta})$

[A.15]

uniformly i . Hence, by CS and J,

$$\begin{aligned} \sum_{i=1}^n \mathbf{I}_i |\hat{r}_{4,i}| &\leq C \sum_{i=1}^n \left| \mathbf{I}_i \hat{\lambda}'_i \hat{m}(x_i, \hat{\beta}) \right|^2 \\ &\leq O_p(nc_n^4) + O_p(1/n^{1-4/\eta}). \end{aligned}$$

Therefore, substituting into eq. (C.10), from eq. (C.9),

$$\begin{aligned} \sum_{i=1}^n \mathbf{I}_i \left(\sum_{j=1}^n w_{ij} [\gamma(\hat{\mu}_i + \hat{\lambda}'_i \hat{u}_j)]^{\frac{\gamma+1}{\gamma}} - 1 \right) &= \frac{1}{2}(\gamma+1) \sum_{i=1}^n (\mathbf{I}_i \hat{\lambda}_i)' \hat{V}(x_i, \hat{\beta}) (\mathbf{I}_i \hat{\lambda}_i) \\ &\quad + \sum_{i=1}^n \mathbf{I}_i \sum_{j=1}^n w_{ij} \hat{r}_{3,ij} + O_p(nc_n^4) + O_p(1/n^{1-4/\eta}) \\ &= \frac{1}{2}(\gamma+1) \sum_{i=1}^n (-\mathbf{I}_i \hat{V}(x_i, \hat{\beta})^{-1} \hat{m}(x_i, \hat{\beta}) + \mathbf{I}_i \hat{r}_{2,i})' \\ &\quad \times \hat{V}(x_i, \hat{\beta}) (-\mathbf{I}_i \hat{V}(x_i, \hat{\beta})^{-1} \hat{m}(x_i, \hat{\beta}) + \mathbf{I}_i \hat{r}_{2,i}) \\ &\quad + \sum_{i=1}^n \mathbf{I}_i \sum_{j=1}^n w_{ij} \hat{r}_{3,ij} + O_p(nc_n^4) + O_p(1/n^{1-4/\eta}). \end{aligned}$$

Similarly to TK, eq. (A.11) and below, $\left\| \sum_{i=1}^n \mathbf{I}_i \hat{r}'_{2,i} \hat{V}(x_i, \hat{\beta}) \hat{r}_{2,i} \right\| = o_p(n^{1+2/m} c_n^4) + o_p(1/n^{1-2/m})$ and $\left| \sum_{i=1}^n \mathbf{I}_i \sum_{j=1}^n w_{ij} \hat{r}_{3,ij} \right| = O_p(nc_n^3)$. Therefore,

$$\begin{aligned} \sum_{i=1}^n \mathbf{I}_i \left(\sum_{j=1}^n w_{ij} [\gamma(\hat{\mu}_i + \hat{\lambda}'_i \hat{u}_j)]^{\frac{\gamma+1}{\gamma}} - 1 \right) &= \frac{1}{2}(\gamma+1) \hat{T}_n + o_p(n^{1+2/m} c_n^4) + o_p(1/n^{1-2/m}) \\ &\quad + O_p(nc_n^3) + O_p(1/n^{1-4/\eta}). \end{aligned}$$

■

References

- Ai, C. (1997): “A Semiparametric Maximum Likelihood Estimator”, *Econometrica*, 65, 933-964.
- 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.
- Borwein, J.M., and A.S. Lewis (1991): “Duality Relations for Entropy-Like Minimisation Problems”, *SIAM Journal of Control and Optimisation*, 29, 325-338.
- 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.
- de Jong, P. (1987): “A Central Limit Theorem for Generalized Quadratic Forms”, *Probability Theory and Related Fields*, 75, 261-277.
- 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.
- Fan, J., and I. Gijbels (1996): *Local Polynomial Modelling and Its Applications*. Chapman Hall: London.

- 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.
- Horowitz, J.L. (1998): "Bootstrap Methods for Covariance Structures," *Journal of Human Resources*, 33, 38-61.
- 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., R.H. Spady and P. Johnson (1998): "Information Theoretic Approaches to Inference in Moment Condition Models," *Econometrica*, 66, 333-357.
- Imbens, G.W. and R.H. Spady (2001): "The Performance of Empirical Likelihood and Its Generalizations," paper presented at 2001 NSF-Berkeley Econometrics Symposium on "Identification and Inference for Econometric Models".
- Judge, G. and R. Mittelhammer (2001): "Empirical Evidence Concerning the Finite Sample Performance of EL-Type Structural Equation Estimators," paper presented at NSF-Berkeley Symposium on "Identification and Inference for Econometric Models".
- Kitamura, Y. (2001): "Asymptotic Optimality of Empirical Likelihood for Testing Moment Restriction," *Econometrica*, 69, 1661-1672.
- Kitamura, Y. (2005): "Empirical Likelihood Methods in Econometrics: Theory and Practice." Invited paper presented at the Econometric Society World Congress, U.C.L., London.
- 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." Forthcoming *Econometrica*.
- 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," in *Handbook of Statistics*, Vol. 11, Maddala, G.S., C.R. Rao and H. Vinod (eds.), Amsterdam: North Holland.
- Newey, W.K. and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in Engle, R. and D. McFadden, eds., *Handbook of Econometrics*, Vol. 4, New York: North Holland.

- Newey, W.K., J.J.S. Ramalho, and R.J. Smith (2001): “Asymptotic Bias for GMM and GEL Estimators with Estimated Nuisance Parameters.” Forthcoming in *Identification and Inference in Econometric Models: Essays in Honor of Thomas J. Rothenberg*, eds. D.W.K. Andrews and J.H. Stock. Cambridge University Press: Cambridge.
- Newey, W.K., and R.J. Smith (2001): “Asymptotic Bias and Equivalence of GMM and GEL Estimators.” Working Paper No. 01/517, University of Bristol.
- 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.
- Rényi, A. (1961): “On Measures of Entropy and Estimation”, *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, 547-561.
- Robinson, P.M. (1987): “Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form,” *Econometrica*, 55, 875-891.
- Shannon, C.E., and W. Weaver (1949): *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.
- 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, 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): “Local GEL Estimation with Conditional Moment Restrictions”, working paper, Department of Economics, University of Warwick.
- Tripathi, G., and Y. Kitamura (2003): “Testing Conditional Moment Restrictions”, *Annals of Statistics*, 31, 2059-2095.