Stochastic Expansions and Moment Approximations for Three Indirect Estimators

Stelios Arvanitis and Antonis Demos
Athens University of Economics and Business

March 17, 2010

Abstract

This paper deals with properties of three indirect estimators that are known to be *(first order) asymptotically equivalent*. Specifically, we examine a) the issue of validity of the formal Edgeworth expansion of an arbitrary order. b) Given a), we are concerned with valid moment approximations and employ them to characterize the second order *bias* structure of the estimators. Our motivation resides on the fact that one of the three is reported by the relevant literature to be second order unbiased. However, this result was derived without any establishment of validity. We provide this establishment, but we are also able to massively generalize the conditions under which this second order property remains true. In this way, we essentially prove their *higher order inequivalence*. We generalize indirect estimators by introducing recursive ones, emerging from multistep optimization procedures. We are able to establish higher order unbiaseness for estimators of this sort.

KEYWORDS: Asymptotic Approximation, Second Order Bias Structure, Binding Function, Local Canonical Representation, Convex Variational Distance, Recursive Indirect Estimators, Higher order Bias.

JEL: C10, C13

1 Introduction

Indirect Inference (hereafter II), usually applied to parametric statistical models,[1] employs a (possibly) "misspecified", auxiliary model for inference on the parameter value corresponding to the true unknown measure in which the relevant sample space is equipped. The motivation is largely computational, hence the choice of the auxiliary model is primarily driven by numerical cost considerations. Despite this motivational characteristic, II gives rise to an enrichment of the theory of parametric statistical inference, due to the fact that it relies on the local inversion of functions that "bind" (possibly) different collections of probability measures defined on the same probability space.

These functions essentially describe relations between classes of random elements defined on each collection, that are typically employed for statistical estimation (e.g. moment conditions). In this respect, a collection of random elements employed to define an estimation procedure in one model, can be pulled back to another and therefore used to indirectly facilitate inference. When these collections of measures have additional structure, the resulting "binding" can be chosen so that (at least locally) it respects this structure. Consequently, the central notion of II procedures is the one of the binding function, denoted by $b(\theta)$, where θ the parameter vector to be estimated. In pure terms this constitutes a function between the measures involved in the relevant statistical models. What is usually discussed is not the function itself, but a parametric representation of it.

This paper is concerned with the *approximation* of certain finite sample properties of three indirect estimators that are known to be *(first order) asymptotically equivalent*. Specifically, for each one of them, we examine a) the validity of the formal Edgeworth expansion of its sequence of distributions, provided by the inversion of the Taylor expansion of any finite order, of the first order conditions that it satisfies. b) Given the validity, we explicitly provide conditions that establish the validity of the approximation of the first moment sequence of the estimator by the relevant sequence of inversion, and c) we explicitly provide the moment approximation of the second order expansion and use it in order to characterize the *bias* structure of the estimators up to this order. Our motivation resides in the fact that one of the three is reported by the relevant literature to be second order unbiased under a particular set of conditions. This result, which is cited bellow, was derived without any establishment of validity. We provide this establishment, but

[1] Although it can be extended into a semiparametric framework, see Dridi and Renault [8].

we also are able to massively generalize the conditions under which this second order property remains true. There are no, in the literature, analogous results for the other two estimators. Validating the expansions at any order and deriving the second order expansion for the remaining estimators, we show that the previous result does not apply in these cases. Hence we essentially derive their *higher order inequivalence.*

The expansions involved concern the so-called *delta method* of approximations of moments of estimator sequences widely used in a *formal* manner in statistics (e.g. Linton [16] and McCullagh [18]).[2] This method proceeds into deriving approximations of the analytical functional forms of extremum statistics using the implicit function theorem, and then approximating the sequence of moments by the moments of the approximations (see e.g. Sargan [21] and Phillips [20]). Hence the estimator sequence is approximated by a sequence of random elements (not necessarily defined on the same probability space), which is generally termed *stochastic expansion.* These expansions do not suffice for the approximation of distributional characteristics unless conditions that ensure some sort of *continuity* of the map that assigns to a sequence of random elements the associated sequence of probability distributions are imposed. These conditions usually work through the following mechanism: both the sequences of distributions of the estimator and the stochastic expansion sequences are proven to be (in the appropriate manner) approximated by the same sequence of Edgeworth distributions. Due to the fact that the underlying space of sequences of distributions is properly topologized, since both sequences are close to the same sequence of distributions then a topological form of the triangle inequality must hold: they must also be close.[3]

We need some further clarification on the notions that we attribute to the approximations examined. Let M and M^* denote arbitrary finite measures defined on the same measurable *topological vector* space S. Let $\mathcal{B}_C$ denote the collection of convex Borel sets of the space. The convex variational distance between these is defined as

$$\mathcal{CVD}(M, M^*) = \sup_{A \in \mathcal{B}_C} |M(A) - M^*(A)|$$

It can be easily seen that the $\mathcal{CVD}$ topologizes the set of finite measures on the space (say $\mathcal{MF}(S)$), as a pseudometrizable (hence first countable) non Hausdorff space. Consider now two arbitrary sequences (say M_n and M_n^*) of

[2]The term formal means "purely algebraic, without concern for topological matters of convergence".

[3]Note that this type of argument does not hold in general neighborhood spaces that are not topological.

the latter space that have the *same* $\mathcal{CVD}-$*limit (say* M_0*)*. We say that M_n^* provides an asymptotic approximation of order s to M_n *iff*

$$\mathcal{CVD}(M_n, M_n^*) = o\left(n^{-a}\right)$$

for some, $a = \frac{i}{2}$, $i \in \{0, 1, \ldots\}$ and $s = 2a + 1$.[4] Hence, the set of sequences of finite measures on S that $\mathcal{CVD}$ converge to M_0, say $\left((\mathcal{MF}(S))^{\mathbb{N}}, M_0\right)$ is topologized by the asymptotic approximation definition as a pseudometrizable non Hausdorff space. In this respect, the s order asymptotic approximation sequence M_n^* is simply an element of a closed ball with center M_n and an radius that depends on a.

Notice that, first, if M_n^* is a sequence of Edgeworth measures then we say that M_n has a *valid* Edgeworth expansion of order s. Remember that the Edgeworth measures are not probability measures but finite signed ones. Second, in a similar construction, we can consider the set of sequences of elements of a Euclidean space that have the same limit. Due to the fact that a Euclidean space is metric, then this set can also be topologized as a pseudometrizable non Hausdorff space if, when x_n and y_n are two such sequences that converge to x_0, we define that y_n provides an asymptotic approximation of order s to x_n *iff*

$$\|x_n - y_n\| = o\left(n^{-a}\right)$$

Again, y_n is simply an element of a closed ball with center x_n and an radius that depends on a. This can be helpful in the issue of moment approximation (of some order) of sequences of measures that are mutually asymptotic approximations.

We are essentially concerned on whether given $\mathcal{CVD}(M_n, M_n^*) = o\left(n^{-a}\right)$, it follows that $\left\|\int_S f\left(dM_n - dM_n^*\right)\right\| = o\left(n^{-a}\right)$ for a given $f \in \left(\mathbb{R}^q\right)^S$. In the case of a bounded f, the aforementioned consequence is valid. When however f is not bounded, then it generally does not hold, either because the function $\int_S f d\cdot$ on $\left((\mathcal{MF}(S))^{\mathbb{N}}, M_0\right)$ does not attain its values in $\left(\left(\mathbb{R}^q\right)^{\mathbb{N}}, x_0\right)$ (e.g. f is not integrable w.r.t. the limit distribution and/or some elements of the sequences, or some of the sequence of integrals do not converge), or in the case that the function $\int_S f d\cdot:\left((\mathcal{MF}(S))^{\mathbb{N}}, M_0\right) \to \left(\left(\mathbb{R}^q\right)^{\mathbb{N}}, x_0\right)$ is not in general *distance preserving*. This discussion essentially implies that the asymptotic approximation of distributions does not imply the asymptotic approximation of moments. We provide conditions that ensure the latter given the former in the case where $S = \mathbb{R}^q$ and $f = id_{\mathbb{R}^q}$. These conditions are reminiscent of

[4] Obviously in this set up this distance could be expressed in the dual notion of measures.

the uniform integrability ones employed in analogous circumstances, except that in this case we have to also consider the *order* of the approximation (i.e. essentially the value of a).

All three indirect estimators, considered here, essentially involve two step estimation procedures. In the first step, an estimating equation, that is part of the structure of the auxiliary model, is employed in order for the statistical information to be summarized into a statistic with values in the auxiliary parameter space. This statistic is called an *auxiliary estimator*. Under the appropriate conditions will (strongly and/or weakly) converge to the value of the binding function when evaluated at the true parameter value. This remark motivates the second step. If this function is at least locally invertible, it is inverted at the value of the auxiliary estimate in order for the indirect estimate to be computed. The auxiliary estimator is denoted in the paper by β_n whereas θ_n is the collective notation for the indirect ones, with n being the sample size.

The auxiliary estimator is defined (at least for large n) as the global minimizer of a distance function on the auxiliary parameter space. This distance function is represented by a norm, which in turn is represented by a positive definite matrix. Our set up is the outmost general, since we allow for this matrix to be stochastic and dependent on the auxiliary parameter. This matrix is possibly computed with respect to an initial estimator, a situation that mimics the issue of optimal weighting in the GMM estimation theory. We term this general framework as *stochastic weighting*.

The first indirect estimator considered here minimizes an analogous general distance function between the β_n and $b(\theta)$. It is termed GMR1 and it was proposed by Gourieroux, Monfort and Renault [13] in order for the numerical burden of the second estimator to be relaxed. The latter is termed GMR2 and it minimizes the previous distance between β_n and $E_\theta \beta_n$. This is obviously differing from the previous and is the essential reason for the second order properties of the estimator. The third estimator, called GT, was proposed by Gallant and Tauchen [10] and minimizes the norm of the *expectation* of the auxiliary estimating vector. Its motivation is obvious. In all three cases we allow for stochastic weighting in the sense described above. In most realistic cases, the expectations involved and the binding function are analytically intractable, hence approximated by simulations. It is easily seen that the simulation counterpart of the GMR2 estimator is the one involved with the maximal numerical burden among the three.

Gourieroux, Renault and Touzi [14] show that the GMR2 estimator has null, up to second order bias, since *it involves the computation of* $E_\theta \beta_n$ (called the *small sample binding function*), when i) the dimension of the structural parameter space equals the dimension of auxiliary and ii) the binding function is affine. Notice that ii) is automatically satisfied, when the auxiliary coincides with the structural model and the binding function is approximated by a consistent estimator of the auxiliary parameters. In this case the particular indirect estimator is said to perform a *bias correction* of the first step one.[5]

Notice that each of the indirect estimators, in the framework of stochastic weighting, are essentially derived from the evaluation of the inverse of a *finite sample binding function* (say $b_n(\theta, W_n, \theta_n^*)$), that depends on the weighting matrix and on an initial estimator (say θ_n^*), evaluated at the auxiliary estimator. Each of these functions generally differ across the estimators that are considered here, but under the appropriate conditions, converge uniformly on $b(\theta)$. In the special case where the involved dimensions coincide, and the weighting is non-stochastic, then $b_n(\theta, W_n, \theta_n^*) = b(\theta)$ in the case of GMR1 and GT (see lemma 2.2), while in the case of GMR2 $b_n(\theta, W_n, \theta_n^*) = E_\theta \beta_n$. Hence the stochastic weighting, essentially generalizes the structure of the functions from the inversion of which the Indirect Estimators (IE) are derived.[6]

As now that IE are derived from the computation of a local diffeomorphism (say f) at the auxiliary estimator, applying a local canonical coordinates theorem, one can always find a representation of f, termed local canonical representation, such that the binding function b is of the form $\begin{pmatrix} \theta \\ \mathbf{0}_{q-p} \end{pmatrix}$. Consequently, the appropriate parameterization of the auxiliary model is called local canonical parameterization. When this parameterization is known the resulting IE posses desirable bias properties. Specifically, we prove that in this case, and under constant weighting, the GMR2 estimator is second order unbiased. However, in most cases of interest the local canonical representation is not known. Nevertheless, by introducing a recursive multistep procedure for the GMR2 estimator, we prove that this IE can be unbiased up to given order.

We immediately provide the assumption framework needed for the definition of the examined estimators. We then provide, in section 3, assumptions sufficient for and derive the validity of the Edgeworth approximations. Given

[5] Gourieroux et al. [14] are occupied with the up to third order $\left(O\left(n^{-1}\right)\right)$ bias structure of the estimator in question. However the complexity of the third order term, does not lead to general conclusive statements. Hence we choose to examine terms up to order $O\left(n^{-\frac{1}{2}}\right)$ as in Gourieroux and Monfort [12] (chapter 4).

[6] These functions are required to be injective, at least locally. In cases where this is not true, the inversion can be performed with the use of some *measurable choice function* the existence of which resides upon the relevant framework. We do not pursue this approach here.

the results of this section, we provide, in the following one, assumptions that validate the first moment approximations and derive the approximations for $a = \frac{1}{2}$. We also discuss the bias properties of the estimators, present the local canonical form of the binding function in section 4.2, and provide multistep extensions of the GMR2 estimator that have desirable bias properties of general order (section 4.3). In section 5 we conclude. We gather all proofs in the first appendix, whereas in the second one we provide a series of useful general lemmas.

2 General Assumption Framework

We introduce our general assumption framework that facilitate the following definition of the estimators. Any other assumption will be introduced locally. The symbol $\mathcal{O}_{\varepsilon}(\theta)$ will denote the ε-ball around θ in a relevant metric space, and let $d = \max(2a + 2, 3)$.

Assumption A.1 *The results that will be later presented, lie in the premises of a well-specified, identified (differentially) parametric, and finite dimensional statistical model that is consisted of a family of probability distributions with respect to a dominating measure (say μ), defined on the measurable space $(\mathbb{R}^m, \mathcal{B}_{\mathbb{R}^m})$.*[7] *We will denote this family of distributions with $\mathcal{D}$, with a global Lipschitz parameterization, that is a (k^{th}-order) diffeomorphism (for $k \geq d$), say **par** to an open bounded subset of $\mathbb{R}^p$ for some $p \in \mathbb{N}$, which we denote by Θ.*[8] *We denote with D_0 the unknown true distribution which corresponds to the true probability measure (say P_{θ_0}) with which the underlying probability space is equipped, and with $\theta_0 =$**par**(D_0).*

Remark R.1 *Notice that, since Θ is a bounded subset of a finite dimensional Euclidean space it is also totally bounded. Further, D could be extended so as to be homeomorphic to a compact superset of Θ, say Θ^*. In this case*

[7] We could easily generalize the form of the underlying measurable space in order to retain only some desirable structures such as differentiabilty of real functions that are defined on it etc.

[8] This means that $\mathcal{D}$ (which by construction obtains the topology of variation norm) has the structure of a (of k order) differentiable manifold, that could be among others inherited by a relevant structure on the underlying measurable space, see the previous note. Since we are not interested in (almost) any geometric properties of our results, the assumption of a global parametrization is without loss of generality. It is trivial that **par** is not unique, since any other autodiffeomorphism of the same order on Θ, will produce another parametrization by composition with **par**. For further inquiries on the geometry of smooth statistical models see among others Amari and Nagaoka [1].

and in order for the differentiability properties to be retained the previous assumption could be completed with $\theta_0 \in Int(\Theta^*)$.

Let B denote a subset of $\mathbb{R}^q$ for some $q \in \mathbb{N}$ and a function $b : \Theta \to B$, which is hereafter termed as the *binding function* and we denote with D^r, the r-derivative operator that maps a function to a function that consists of the algebraic element containing all the r^{th}-order partial derivatives of the first. When A is a matrix $\|A\|$ will denote a topologically equivalent yet *submultiplicative* matrix norm, such as the Frobenius norm (i.e. $\|A\| = \sqrt{trA'A}$). Also when suprema, with respect to parameters, of derivatives are discussed these are obviously taken where the differentiated function is differentiable.

Assumption A.2 $b(\theta_0) = b(\theta)$ *iff* $\theta = \theta_0$, *and for some* $\varepsilon_1 > 0$, *the restriction* $b|_{\mathcal{O}_{\varepsilon_1}(\theta_0)} : \mathcal{O}_{\varepsilon_1}(\theta_0) \to B$ *is invertible. For some* $\varepsilon_1 \geq \varepsilon_2 > 0$, *the restriction* $b|_{\mathcal{O}_{\varepsilon_2}(\theta_0)} : \mathcal{O}_{\varepsilon_2}(\theta_0) \to B$ *is a* $k-$*diffeomorphism.*

Remark R.2 *The invertibility of the particular restriction of the binding function, implies that* θ_0 *is inferable from the knowledge of* $b(\theta_0)$ *and of the restricted binding function, a property that is a cornerstone for the concept of II, hence it is termed as* ***local indirect identification***. *Furthermore, we have that* $q \geq p$ *and that rank* $\left(\frac{\partial b}{\partial \theta'}\right) = p$, $\forall \theta \in \mathcal{O}_{\varepsilon_2}(\theta_0)$.

Assumption A.3 $b(\theta)$ *is Lipschitz on* Θ *and* $\sup_\theta \|D^r b(\theta)\| < M_r$, $\forall r = 2, \ldots, d+1$ *for* $\theta \in \mathcal{O}_{\varepsilon_3}(\theta_0)$, *for some* $\varepsilon_3 \leq \varepsilon_2$, *with* $M_r \in \mathbb{R}^+$.

We also consider the function $c : \mathbb{R}^m \times B \to \mathbb{R}^l$ for some $l \in \mathbb{N}$ such that:

Assumption A.4 p, q, l *are finite and* $p \leq q \leq l$ *and* $c(\cdot, \cdot)$ *is jointly measurable with repsect to the product algebra of* $\mathbb{R}^m \times B$, *and* $c(x, \cdot)|_{b\left(\mathcal{O}_{\varepsilon_2}(\theta_0)\right)}$ *is* $d-$*continuously differentiable on* $b\left(\mathcal{O}_{\varepsilon_2}(\theta_0)\right)$ *for* $\mu-$*almost all* $x \in \mathbb{R}^m$ *with* $k \geq d = \max(3, 2a+2)$. *Also* $\|c(x,\beta) - c(x,\beta')\| \leq u_c(x)\|\beta - \beta'\|$, $\forall \beta, \beta' \in B$ *and* $\sup_\theta E_\theta \|u_c\|^{q_0}, E_\theta \|c(x,\beta)\|^{q_0} < \infty$, *for some* $q_0 \geq \max(2a+1, 2)$ *and,* $\forall \beta \in b\left(\mathcal{O}_{\varepsilon_2}(\theta_0)\right)$, *and* $E_\theta c(x,\beta) = \mathbf{0}_{l\times 1}$, *iff* $\beta = b(\theta)$, $\forall \theta \in \mathcal{O}_{\varepsilon_2}(\theta_0)$. *Also,* $\sup_{\varphi \in \mathcal{O}_\eta(\varphi_0)} \left\|\frac{\partial}{\partial \theta'} E_\theta\left[c(x,\beta)\right]\right\|$ *and* $\sup_{\varphi \in \mathcal{O}_\eta(\varphi_0)} \left\|\frac{\partial}{\partial \theta_i \partial \theta_j} E_\theta\left[c(x,\beta)\right]\right\|$ *are bounded* $\forall i, j = 1, \ldots p$ *for some* $\eta > 0$, *where* $\varphi_0 = \left(b^{/}(\theta_0), \theta_0^{/}\right)^{/}$ *and the product topology is considered.*

Remark R.3 *The previous assumption implies the identification of* $b(\theta_0)$, *as the unique solution of* $E_{\theta_0} c(x,\beta) = \mathbf{0}_{l\times 1}$, *which along with the required* *differentiability implies that the rank* $\left(E_\theta \frac{\partial c(x,\beta)}{\partial \beta'}\right) = q$, $\forall \beta \in b\left(\mathcal{O}_{\varepsilon_2}(\theta_0)\right)$, $\forall \theta \in \mathcal{O}_{\varepsilon_2}(\theta_0)$. *Conditions of the form* $\|c(x,\beta) - c(x,\beta')\| \leq u_c(x)\|\beta - \beta'\|$, $\forall \beta, \beta' \in B$ *can be termed as global stochastic Lipschitz continuity conditions and facilitate the convergence of the auxiliary estimators to* $b(\theta_0)$.

Remark R.4 *The function* c *and the estimating equations* $E_\theta c(x, b(\theta)) = \mathbf{0}_{l\times 1}$ *can reflect part of the structure of an auxiliary model, not necessarily well specified.*

Remark R.5 *Similarly, conditions for boundeness of quantities such as* $\sup_{\theta\in\mathcal{O}_\eta(\varphi_0)} \left\|\frac{\partial}{\partial\theta'} E_\theta\left[c(x,\beta)\right]\right\|$, *holding locally on* $\Theta \times B$ *are typically used in the case of the GT estimator and can be derived from conditions like* $\lim_{n\to\infty} \sup_{\theta\in\mathcal{O}_\eta(\varphi_0)} E_\theta \left\|\sqrt{n}c_n(\beta)\right\|^2 < \infty$ *and* $\lim_{n\to\infty} \sup_{\theta\in\mathcal{O}_\eta(\theta_0)} E_\theta \left\|\sqrt{n}\overline{s}_n(\theta)\right\|^2 < \infty$ *where* $\overline{s}_n(\theta)$ *denotes the average score function. Analogously the condition for the second order derivatives would follow from the condition above and* $\lim_{n\to\infty} \sup_{\theta\in\mathcal{O}_\eta(\theta_0)} E_\theta \left\|\sqrt{n}\overline{s_n(\theta)\, s_n'(\theta)} + \overline{H}_n(\theta)\right\|^2 < \infty$ *(see also A.6 and R.10 for analogous conditions).*

Notice that due to the fact that the spaces Θ and B are separable subsets of Euclidean spaces, suprema of real random elements over these spaces are typically measurable (see van der Vaart and Wellner [24], example 1.7.5 p. 47 for φ any map from the closure to the interior which is the identity on the interior).

The following assumption concerns the weighting matrices and essentially implies that thee matrices will satisfy a L.L.N. at θ_0 or $b(\theta_0)$, and even more evaluated at points that converge to the aforementioned.

Assumption A.5 *Let* $W(x,\beta)$, $W^*(x,\theta)$ *and* $W^{**}(x,\theta)$ *be* $l \times l$, $q \times q$ *and* $l \times l$ μ*-almost surely positive definite random matrices such that d-differentiable* $\forall \beta \in b\left(\mathcal{O}_{\varepsilon_2}(\theta_0)\right)$, $\forall \theta \in \mathcal{O}_{\varepsilon_2}(\theta_0)$, *such that* $E_{\theta_0} W(x, b(\theta_0)) = W(b(\theta_0))$, $E_{\theta_0} W^*(x,\theta_0) = W^*(\theta_0)$ *and* $E_{\theta_0} W^{**}(x,\theta) = W^{**}(\theta_0)$ *are well defined positive definite matrices, and* $E_{\theta_0}\left\|W(x, b(\theta_0))\right\|^{q_0} < \infty$, $E_{\theta_0}\left\|W^*(x,\theta_0)\right\|^{q_0} < \infty$ *and* $E_{\theta_0}\left\|W^{**}(x,\theta_0)\right\|^{q_0} < \infty$ *for* q_0 *defined above.*

Analogously, in the following let $W_n(\beta)$, $W_n^*(\theta)$ and $W_n^{**}(\theta)$ denote $\frac{1}{n}\sum W(x_i,\beta)$, $\frac{1}{n}\sum W^*(x_i,\theta)$, and $\frac{1}{n}\sum W^{**}(x_i,\theta)$ respectively.

2.1 Definition of Estimators

In this section the set of estimators under examination are defined. They are all minimum distance estimators, whose existence is verified (at least asymptotically) by the previous assumption framework. In any case their existence, as well defined single valued measurable functions on the relevant sample space (say Ω^n), can be facilitated by the use of measurable choice functions.

Denote with $\mathcal{PD}(k, \mathbb{R})$ the vector space of positive definite matrices of dimension $k \times k$ (with respect to matrix and scalar multiplication). Consider the following real function from $\mathbb{R}^k \times \mathcal{PD}(k \times k)$ for $k \in \mathbb{N}$

$$(x, A) \to \left(x^{/} A x\right)^{1/2}$$

for a given matrix the previous function defines a norm on $\mathbb{R}^k$. Denote the function $(\cdot, \cdot)|_A$ with $\|\cdot\|_A$. We denote by Ω^n the sample space for sample of size n.

We next define the auxiliary estimator β_n as:

Definition D.1 *The auxiliary estimator* $\beta_n : \Omega^n \to B$ *is defined as*

$$\beta_n = \arg\min_{\beta \in B} \|c_n(\beta)\|_{W_n(\beta_n^*)}$$

Given the definition of the auxiliary estimator we define the indirect ones. We collectively denote them with θ_n, since in the following context there is not danger of confusion. The first and second of thee indirect estimators were formalized by Gourieroux et al. [13] while the third was introduced by Gallant and Tauchen [10] (see also Gourieroux and Monfort [12], chapter 4, for a summary).

The GMR1 estimator is defined as:

Definition D.2 *The GMR1 estimator* $\theta_n : \Omega^n \to B$ *is defined as*

$$\theta_n = \arg\min_{\theta \in \Theta} \|\beta_n - b(\theta)\|_{W_n^*(\theta_n^*)}$$

Under assumptions A.1 and A.3, i.e. Θ is bounded and $b(\cdot)$ is Lipschitz, B is bounded and the following lemma is trivially true.

Lemma 2.1 *Under assumptions A.1 and A.3,* $\|E_\theta \beta_n\| < \infty$

Given the above lemma, it is possible to define the GMR2 estimator as:

Definition D.3 *The GMR2 estimator* $\theta_n : \Omega^n \to B$ *is defined as*

$$\theta_n = \arg\min_{\theta \in \Theta} \|\beta_n - E_\theta \beta_n\|_{W_n^*(\theta_n^*)}$$

We denote by $E_\theta(c_n(\beta_n))$, the quantity $E_\theta(c_n(\beta))|_{\beta=\beta_n}$ for notational simplicity. Due to assumption A.4 we have that $\forall\theta\ \|E_\theta(c_n(\beta_n))\| < \infty$. Consequently, the following minimization procedure can be defined.

Definition D.4 *The GT estimator* $\theta_n : \Omega^n \to B$ *is defined as*

$$\theta_n = \arg\min_{\theta\in\Theta} \|E_\theta(c_n(\beta_n))\|_{W_n^{**}(\theta_n^*)}$$

The usual definition of the aforementioned estimator is given only when the auxiliary estimator is the MLE of the auxiliary model. The currently defined one is an obvious extension.

Remark R.6 *We implicitly assume that the* arg min *is non-empty. In case that this happens, we can compactify* Θ *as in remark R.1. On the other hand, if the* arg min *is multivalued, one can consider results such as the ones in Proposition 3.2 in Dupacova and Wets [9]. In fact, any of objective functions are jointly measurable with respect to the relevant product spaces and are almost everywhere continuous with respect to the relevant parameter spaces. Given this, any of the objective functions can be extended to random lower semi-continuous functions with proper domains the relevant spaces. Consequently, the above mentioned Proposition applies, justifying the existence of any of these estimators as a measurable selection.*

Remark R.7 *The computation of all three estimators relies on the analytical knowledge of the binding function or the engaged expectations, which are usually intractable. Due to this fact in applications approximations of these estimators are defined, in which the unknown elements are approximated by simulations.*

It is known that the three estimators are asymptotically first order equivalent (proviso a certain selection of the weighting matrix of GMR1 and GMR2 given the weighting matrix of the GT estimator). However, in the special case where $p = q$ we have the following lemma (see appendix for a proof).

Lemma 2.2 *When GMR1 and GT are consistent and* $p = q = l$*, with probability* $1 - o(n^{-a})$

GMR1=GT

Notice that the previous lemma makes sense for large enough n, due to the possibility of non-empty boundaries, and/or non existence of either or both of the estimators.

Remark R.8 *Notice that in this framework and in analogy to the particular relationship between the GMR1 and the GT estimators, we could also define a variant of the latter (it would be homologous to the GMR2 estimator, hence could be termed as GT2 estimator), as the solution of* $c_n(E_\theta(\beta_n)) = \mathbf{0}_p$. *Obviously, since* $c_n(\beta_n) = \mathbf{0}_p$ *by construction, then GMR2=GT2. This provides another characterization of the distinction between the GMR1 and GMR2 estimators in this particular set up. The two estimators are different because* $c_n(E_\theta(\cdot))$ *and* $E_\theta c_n((\cdot))$ *have different roots and therefore* ***their distinction lies in non commutativity****. This observation gives rise to the next lemma. Furthermore, the GT2 estimator could also be generalized with the introduction of differences in the relevant dimensions, stochastic weighting etc. In this respect it would not generally coincide with the GMR2 estimator hence should be addressed as a distinct case of an indirect estimator, with which we are not concerned in the present paper.*

Lemma 2.3 *When* $p = q = l$ *and* $c(x_i, \beta) = f(x_i) - E_\beta f(x_i) = f(x_i) - g(\beta)$ *then:*

1. the GMR1 estimator is essentially a GMM estimator.

2. If g is linear then GMR1=GMR2.

Remark R.9 *1. would be valid even if* $\beta_n = r \circ g^{-1} \circ \frac{1}{n} f(\omega_i)$ *for* r *a bijection. Hence the GMR1 can be a GMM estimator even in cases that the auxiliary is an appropriate transformation of a GMM estimator.*

3 Validity of Edgeworth Approximations

In this section we expand the assumption framework, in order to validate the Edgeworth approximations and using this we derive the validity. Recall that every estimator considered is an extremum one, and the criterion from which it emerges is at least locally differentiable. Accommodating these facts we employ the following steps to prove the validity. First, we prove that the estimators satisfy the first order conditions with probability $1 - o(n^{-a})$. Then a justified use of the mean value theorem proves $o(n^{-a})$ asymptotic tightness of $\sqrt{n}$ transformation of the estimators. Third, due to the first step a local approximation of the $\sqrt{n}$ transformation is obtained by a Taylor expansion of the first order conditions and using the second step it is proven that the relevant remainder is bounded by an $o(n^{-a})$ real sequence with probability $1 - o(n^{-a})$. Taking also into account corollary AC.1 we get that if valid, the $\sqrt{n}$ transformation and the approximation have the same Edgeworth expansion. Finally, the validity is established from the validity of the relevant expansion of the aforementioned approximation.

This methodology coincides with the one in Andrews [2] and is essentially based on local differentiability, lemma AL.6 and Bhattacharya and Ghosh [4] which provide a theorem of invariance of validity of Edgeworth approximations with respect to locally differentiable functions. Notice also that lemma AL.6 enables the extension of the results in non differentiable case, but this will not be pursued here.

Assumptions Specific to the Validity of the Edgeworth Approximations

Let $f(x,\beta)$ denote the vector that contains stacked all the distinct components of $c(x,\beta)$, $W(x,\beta)$, $W^{*}(x,\theta)$ and $W^{**}(x,\theta)$ as well as their derivatives up to the order $d=\max(3,2a+2)$.

Assumption A.6 $\sup_{\theta\in\mathcal{O}_{\varepsilon_4}(\theta_0)}\left\|D^{r}E_{\theta}\beta_{n}\right\|<M_{r}^{*}$*, for* $0<\varepsilon_4\leq\varepsilon_2$*, for* $r=2,\ldots,d+1$*, and* $M_{r}^{*}>0$*.*

Remark R.10 *Assumption A.6 along with Assumption A.3 imply that for* $r=2,\ldots,d+1$*,* $\sup_{\theta\in\mathcal{O}_{\min(\varepsilon_3,\varepsilon_4)}(\theta_0)}\left\|D^{r}\left(E_{\theta}\beta_{n}-b(\theta)\right)\right\|<M_{r}+M_{r}^{*}$*, which in turn means that* $D^{r-1}\left(E_{\theta}\beta_{n}-b(\theta)\right)$ *are uniformly Lipschitz on* $\mathcal{O}_{\min(\varepsilon_3,\varepsilon_4)}(\theta_0)$*, and therefore uniformly equicontinuous on the same ball. This implies the commutativity of the limit with respect to* n *and the derivative operator (of order* $r-1$*) uniformly over* $\mathcal{O}_{\min(\varepsilon_3,\varepsilon_4)}(\theta_0)$*. Due to Assumption A.1 for* $k\geq d+1$*, this assumption is verified via conditions of the form* $\sup_{\theta\in\mathcal{O}_{\min(\varepsilon_3,\varepsilon_4)}(\theta_0)}E_{\theta}\left\|\sqrt{n}\left(\beta_{n}-b(\theta)\right)\right\|^{2}=O(1)$ *and* $\sup_{\theta\in\mathcal{O}_{\min(\varepsilon_3,\varepsilon_4)}(\theta_0)}E_{\theta}\left\|\sqrt{n}\overline{l}_{n}(\theta)\right\|^{2}=O(1)$ *where* $\overline{l}_{n}(\theta)$ *depends on derivatives of the (well defined in our setting) average likelihood function. For example for* $r=2$*, we have that* $\overline{l}_{n}(\theta)=\overline{s_{n}(\theta)s_{n}^{\prime}(\theta)}+\overline{H}_{n}(\theta)$*.*

Assumption A.7 $E_{\theta}\left\|f(x,\beta)\right\|^{q_1}<\infty$*,* $\forall\theta\in\mathcal{O}_{\varepsilon_2}(\theta_0)$*,* $\forall\beta\in b\left(\mathcal{O}_{\varepsilon_2}(\theta_0)\right)$ *for* $q_1=2a+3$*,* $\left\|f(x,\beta)-f(x,\beta^{\prime})\right\|\leq\kappa_{\gamma}\left\|\beta-\beta^{\prime}\right\|$*,* $\forall\beta\in b\left(\mathcal{O}_{\varepsilon_2}(\theta_0)\right)$*,* μ*-almost surely for an almost surely positive random variable* κ_{γ}*, with* $E_{\theta}\kappa_{\gamma}^{q_1}<\infty$*,* $\forall\theta\in\mathcal{O}_{\varepsilon_2}(\theta_0)$*.*[9]

This condition that could be termed as local stochastic Lipschitz continuity condition facilitate the Edgeworth approximations of the relevant sequences of random elements.

[9]Notice the local nature of the moment existence conditions here and in assumption A.4. These are stronger that the relevant conditions of Andrews [2], and facilitate mainly the case of the GT estimator.

Assumption A.8 *The Weak Dependence assumption and the Cramer type of condition of Andrews [2] or Goetze and Hipp [11] hold for the sequence* $\{f(x_n, b(\theta_0))\}_n$ *and the sequence of characteristic functions of* $\frac{1}{n}\sum f(x_i, b(\theta_0))$ *respectively.*

Remark R.11 *The last two assumptions guarantee that the (unknown) sequence of distributions of the sequence of random elements* $\sqrt{n}\left[\frac{1}{n}\sum f(x_i, b(\theta_0)) - E_{\theta_0}\frac{1}{n}\sum f(x_i, b(\theta_0))\right]$ *can be approximated by a sequence of Edgeworth distributions of order of error* $o(n^{-a})$ *(see Andrews [2]). Notice that the Cramer condition on the conditional characteristic function of* $\frac{1}{n}\sum f(x_i, b(\theta_0))$ *could be implied through controlling the order of magnitude of tail moments of the relevant partial sum.*

Assumption A.9 . *The relevant sequences of distributions of the initial estimators,* β_n^* *and* θ_n^* *can be approximated by a sequence of Edgeworth distributions with an* $o(n^{-a})$ *error.*

This will be trivially satisfied when β_n^* is defined via c and the relevant weighting matrix is independent of β and deterministic. The analogous argument applies for θ_n^*. We now present the results on the validity of Edgeworth approximations for any a for any of the four estimators defined above. We begin with the auxiliary estimator.

Auxiliary Estimator

We can prove the following lemma concerning the auxiliary estimator, that is essentially a direct application of the relevant results in Andrews [2].

Lemma 3.1 *Under assumptions A.1, A.4, A.5, and A.7-A.9 there exists an Edgeworth distribution* $\mathcal{EDG}_a(\bullet)$ *such that*

$$\sup_{A\in\mathcal{B}_C}\left|P_{\theta_0}\left(\sqrt{n}\left(\beta_n - b(\theta_0)\right)\in A\right) - \mathcal{EDG}_a(A)\right| = o\left(n^{-a}\right).$$

Let us now proceed to the validity of the Indirect Estimators

Indirect Estimators

We next present in more details the analogous results for the IE. We proceed in two steps. In the first one we prove the $o(n^{-a})$-consistency and $o(n^{-a})$-tightness and in the sequel we prove the existence of the Edgeworth expansion.

Lemma 3.2 *i) Under the assumptions of lemma 3.1 and under assumption A.2 we have that*

$$P_{\theta_0}\left(\|\theta_n-\theta_0\|>C_2\frac{\ln^{1/2}n}{n^{1/2}}\right)=o\left(n^{-a}\right)\ \textit{for some } C_2>0$$

where θ_n is the GT estimator. ii) If additionally assumption A.3 applies, then θ_n is the GMR1 estimator. iii) If additionally assumption A.6 applies, then θ_n is the GMR2 one.

Existence of Edgeworth Expansions of Indirect Estimators

Lemma 3.3 *Under the assumptions of lemmas 3.2 and assumptions A.8-A.9 the GMR1, and GT estimators admit valid Edgeworth expansions of order $s=2a+1$. Furthermore, if the auxiliary estimator has a valid Edgeworth expansion of order $s=2a+2$, then the GMR2 admits a valid expansion of order $s=2a+1$.*

4 Validity of 1st Moment Expansions

Having established the validity of Edgeworth expansions in every case of the examined estimators, we are concerned with the approximation of their first moment sequences with a view towards the approximation of their bias structure. We know from section 1 that the validity of the former does not imply the validity of the latter. We provide a general lemma, which utilizes the Edgeworth expansions, along with further assumptions that validate the required approximations. These are integrability assumptions that involve rate of convergence, and are presented immediately along with remarks that comment on their applicability.

In the following if A is a measurable set, we denote with $P_n(A)=P\left(\sqrt{n}\left(\theta_n-\theta_0\right)\in A\right)$ where θ_n is any of the examined estimators (auxiliary or indirect) and Q_n a sequence of distributions such that $\mathcal{CVD}\left(P_n,Q_n\right)=o\left(n^{-a}\right)$.

Assumption A.10

$$\exists\epsilon>0:n^{a+\frac{1}{2}}P\left(\sqrt{n}\left(\theta_n-\theta_0\right)\in\sqrt{n}\left(\Theta-\theta_0\right)\backslash\mathcal{O}_{K(\ln n)^{\epsilon}}(0)\right)=o\left(1\right),$$

Remark R.12 *The above assumption is valid when $\sqrt{n}\left(\theta_n-\theta_0\right)$ has a valid Edgeworth expansion of order $s=2a+2$ (see Magdalinos [17], Lemma 2).*

Assumption A.11

$$n^a \int_{\mathbb{R}^q \setminus \mathcal{O}_{K(\ln n)^\epsilon}(0)} \|x\| \, |dQ_n| = o(1)$$

Remark R.13 *In fact if Q_n is the Edgeworth distribution we have that $A = n^a \int_{\mathbb{R}^q \setminus \mathcal{O}_{K(\ln n)^\epsilon}(0)} \|x\| \, |dQ_n| = n^a \int_{\mathbb{R}^q \setminus \mathcal{O}_{K(\ln n)^\epsilon}(0)} \|f_n(z)\| \, d\Phi + o(1)$ where Φ is the multivariate standard normal cumulative distribution function, and as $f_n(z)$ is a polynomial in z we get: $A - o(1) \leq n^a \int_{\mathbb{R}^q \setminus \mathcal{O}_{K(\ln n)^\epsilon}(0)} \left\| \sum_{i=0}^{2a} n^{-\frac{i}{2}} f_i(z) \right\| d\Phi \leq \sum_{i=0}^{2a} n^{\frac{2a-i}{2}} \int_{\mathbb{R}^q \setminus \mathcal{O}_{K(\ln n)^\epsilon}(0)} \|f_i(z)\| \, d\Phi$ where $f_i(z)$ appropriate polynomials in z. Now $n^{\frac{2a-i}{2}} \int_{\mathbb{R}^q \setminus \mathcal{O}_{K(\ln n)^\epsilon}(0)} \|f_i(z)\| \, d\Phi \leq C n^{\frac{2a-i}{2}} \int_{\mathbb{R}^q \setminus \mathcal{O}_{K(\ln n)^\epsilon}(0)} \|z\|^{2\lambda_i} \, d\Phi = C n^{\frac{2a-i}{2}} \int_{\mathbb{R}^q \setminus \mathcal{O}_{K(\ln n)^\epsilon}(0)} \left(\sum_{j=1}^q z_j^2 \right)^{\lambda_i} d\Phi$. Now the l^{th} term in the expansion of the λ_i^{th} power will be of the form: $\prod_{j=1}^{q} z_j^{k_{j,l}}$, where $\sum_{j=1}^q k_{j,l} = 2\lambda_i$.*

Hence, $A - o(1) \leq C n^{\frac{2a-i}{2}} \sum_{l=1}^{q^{\lambda_i}} \int_{\mathbb{R}^q \setminus \mathcal{O}_{K(\ln n)^\epsilon}(0)} \prod_{j=1}^{q} z_j^{k_{j,l}} d\Phi$

$= C n^{\frac{2a-i}{2}} (2\pi)^{-\frac{q}{2}} \sum_{l=1}^{q^{\lambda_i}} \prod_{j=1}^{q} \int_{\mathbb{R} \setminus (-K(\ln n)^\epsilon, K(\ln n)^\epsilon)} z_j^{k_{j,l}} \exp\left(-\frac{z_j^2}{2}\right) dz_j$

$= C n^{\frac{2a-i}{2}} \left(\frac{\pi}{2}\right)^{-\frac{q}{2}} \sum_{l=1}^{q^{\lambda_i}} \prod_{j=1}^{q} \int_{K(\ln n)^\epsilon}^{\infty} z_j^{k_{j,l}} \exp\left(-\frac{z_j^2}{2}\right) dz_j$ as $k_{j,l}$ is even. Now by a change of variables we get that $A - o(1)$

$\leq C n^{\frac{2a-i}{2}} \left(\frac{\pi}{2}\right)^{-\frac{q}{2}} \sum_{l=1}^{q^{\lambda_i}} \prod_{j=1}^{q} 2^{\frac{k_{j,l}-1}{2}} \int_{\frac{K^2(\ln n)^{2\epsilon}}{2}}^{\infty} t^{\frac{k_{j,l}+1}{2}-1} \exp(-t) \, dt$

$= C n^{\frac{2a-i}{2}} \left(\frac{\pi}{2}\right)^{-\frac{q}{2}} \sum_{l=1}^{q^{\lambda_i}} \prod_{j=1}^{q} 2^{\frac{k_{j,l}-1}{2}} \Gamma\left(\frac{k_{j,l}-1}{2}, \frac{K^2(\ln n)^{2\epsilon}}{2}\right)$ where $\Gamma(\bullet, \bullet)$ is the incomplete Gamma function (see e.g. Gradshteyn and Ryzhik [15] formula 8.350). For $\ln n \to \infty$ we have that

$\Gamma\left(\frac{k_{j,l}-1}{2}, \frac{K^2(\ln n)^{2\epsilon}}{2}\right) = \left(\frac{K^2(\ln n)^{2\epsilon}}{2}\right)^{\frac{k_{j,l}-3}{2}} \exp\left(-\frac{K^2(\ln n)^{2\epsilon}}{2}\right) \left[1 + O\left(\left(\frac{K^2(\ln n)^{2\epsilon}}{2}\right)^{-1}\right)\right]$

$\leq \left(\frac{K^2(\ln n)^{2\epsilon}}{2}\right)^{\frac{k_{j,l}-3}{2}} \exp\left(-\frac{K^2(\ln n)^{2\epsilon}}{2}\right)$ (see e.g. Gradshteyn and Ryzhik [15] formula 8.357). Hence

$A \leq C(\pi)^{-\frac{q}{2}} 2^{\frac{3q}{2}} \sum_{l=1}^{q^{\lambda_i}} \prod_{j=1}^{q} (\ln n)^{\epsilon(k_{j,l}-3)} K^{k_{j,l}-3} \exp\left(\frac{(2a-i)\ln n - K^2(\ln n)^{2\epsilon}}{2}\right) + o(1).$

Now for $\epsilon > \frac{1}{2}$, and $K > 0$ we have that

$C\left(\pi\right)^{-\frac{q}{2}}2^{\frac{3q}{2}}\sum_{l=1}^{q^{\lambda_i}}\prod_{j=1}^{q}\left(\ln n\right)^{\epsilon\left(k_{j,l}-3\right)}K^{k_{j,l}-3}\exp\left(\frac{(2a-i)\ln n-(\ln n)^{2\epsilon}}{2}\right)\rightarrow 0$ *as* $n\rightarrow\infty$. *Hence assumption A.11 applies in this case.*

Lemma 4.1 *Given the assumptions A.10 and A.11 above then*

$$n^a\left\|\int_{\mathbb{R}^q}x\left(dP_n-dQ_n\right)\right\|=o\left(1\right).$$

Remark R.14 *Due to the two previous remarks and Lemma 2 in Magdalinos [17], it suffices that* $\sqrt{n}\left(\theta-\theta_0\right)$ *has a valid Edgeworth expansion of order* $s=2a+2$, *since in this case we can choose* $\epsilon>\frac{1}{2}$ *and* $K\geq\sqrt{2a+1}$ *for the above Lemma to be valid.*

Hence we have an analytical procedure that justify the following results in our assumption framework.

4.1 Valid 2nd order Bias approximation for the Indirect estimators

In this section, given the previous results, we are concerned with the bias structure of second order for each of the examined estimators. In order to facilitate the presentation, we make the following definition.

Definition D.5 *Let* $\{x_n\}$ *and* $\{y_n\}$ *denote two sequence of random elements with values in an normed space. We denote the relation* $x_n\underset{a}{\sim}y_n$ *when* $\left\|E\left(x_n-y_n\right)\right\|=o\left(n^{-a}\right)$.

Remark R.15 *Due to the positive definiteness of the norm and the triangle inequality* $\underset{a}{\sim}$ *is an equivalence relation on the set of sequences of random elements whose first moments converge to the same limit.*

We are ready to employ the previous results for the case of $a=\frac{1}{2}$. We essentially invert the Taylor expansion of the first order condition that with high probability satisfies each one of the estimators considered, and are able to ignore the remainders due to the results of the previous paragraphs. We have that $\sup_{A\in\mathcal{B}_C}\left|\mathcal{EDG}\left(A\right)-\Phi\left(A_n\right)\right|=o\left(n^{-a}\right)$ for suitable choice of the sequence $\{A_n\}$ emerging from a bijective correspondence $A\rightarrow A_n$. Hence $\sup_{A\in\mathcal{B}_C}\left|P\left(x_n\in A\right)-P\left(z\in A_n\right)\right|=o\left(n^{-a}\right)$ where x_n denotes the sequence of random elements that we wish to approximate in the relevant sense, and z denotes a standard normal random vector. Then, due to the fact that $P\left(z\in A_n\right)=P\left(\left(g_n\left(z\right)+o\left(n^{-a}\right)\right)\in A\right)=P\left(g_n\left(z\right)\in A\right)+o\left(n^{-a}\right)$ for a suitable choice of a polynomial in z function sequence and the smoothness of Φ

(see Magdalinos [17] or footnote 11 for the definition of smoothness of a distribution; this is implied by analytical smoothness in the case where a density exists), we have that $\sup_{A\in\mathcal{B}_C}|P(x_n\in A)-P(g_n(z)\in A)|$. We then employ lemma 4.1 to obtain the needed results on the mean approximations. Notice also that if there exists a $q_n(z)$ such that $g_n(z)=q_n(z)+o(n^{-a})$, if $x_n \underset{a}{\sim} g_n(z)$, then $x_n \underset{a}{\sim} q_n(z)$, in the light of remark R.15, something that will be needed in the case of GMR2. We also present in the Appendix of General Lemmas a lemma concerning approximations of inverse matrices that will be useful in what follows. Finally, the following assumption concerns the initial estimators, either β_n^* or θ_n^*, and is in the same spirit of assumption assumption A.9.

Assumption A.12 *Any initial estimator has an analogous first moment approximation with the one that it defines.*

Auxiliary Estimators

We begin with the auxiliary estimator β_n. The next lemma summarizes the results.

Lemma 4.2 *If $\sqrt{n}(\beta_n-b(\theta_0))$ has a valid Edgeworth expansion of third order*

$$\sqrt{n}(\beta_n-b_0)\underset{1/2}{\sim} k_1+\frac{k_2}{\sqrt{n}}$$

where

$$k_1=-Qc_\beta^{/}(b_0)W_0c(z,b_0)$$

and

$$\begin{aligned} k_2 &= -Qc_\beta^{/}(b_0)W_0c^*(z,b_0)-QAk_1-\frac{1}{2}Qc_\beta^{/}(b_0)W_0\left\{k_1^{/}c_{\beta,\beta'}(b_0)_j k_1\right\}_{j=1,\ldots,l} \\ &\quad -Q\left[c_\beta^{/}(b_0)w(z,b_0)+c_\beta(z,b_0)W_0\right]c(z,b_0) \\ &\quad -Q\left[c_\beta^{/}(b_0)\left\{W_{\beta^{/}}(b_0)_{rj}k_1^*\right\}_{r,i=1,\ldots,l}+\left\{k_1^{/}c_{\beta,\beta'}(b_0)_j\right\}_{j=1,\ldots,l}W_0\right]\left[c(z,b_0)+c_\beta(b_0)k_1\right] \end{aligned}$$

where $b_0=b(\theta_0)$, $W_0=W(b_0)$, $Q=\left[E_{\theta_0}\frac{\partial c^{/}(x_1,b_0)}{\partial\beta}W_0E_{\theta_0}\frac{\partial c(x_1,b_0)}{\partial\beta^{/}}\right]^{-1}$, $Sym[B]=\frac{1}{2}\left(B+B^{/}\right)$, $A=2Sym\left[\begin{array}{c} E_{\theta_0}\frac{\partial c^{/}(x_1,b_0)}{\partial\beta}W_0c_\beta^{/}(z,b_0) \\ +\frac{1}{2}E_{\theta_0}\frac{\partial c^{/}(x_1,b_0)}{\partial\beta}w(z,b_0)E_{\theta_0}\frac{\partial c(x_1,b_0)}{\partial\beta^{/}}\end{array}\right]$, k_1^* is the relevant term of the analogous expansion of the initial auxiliary estimator β_n^* by assumption A.12, and $z\sim N(0,\Sigma)$ where Σ depends on the problem at hand.

Remark R.16 *It is easy to see that when $l = q$ the results do not depend on the weighting matrix as expected.*

Remark R.17 *$E_{\theta_0} k_1$ is null as this term corresponds to the normal component of the estimators which are asymptotically first order unbiased. Also under relevant integrability conditions that are easily derived in the spirit of lemma 4.1, $E_{\theta_0} k_2$ will depend on the first order asymptotic variance ,on the non linearity of c with respect to β, on the properties of the weighting matrix and the initial auxiliary estimator as well as on the relation between l and q (see Newey and Smith [19]).*

Indirect Estimators

We proceed to state the main results concerning the expansions of the three indirect estimators. These reveal a quite different behavior of GMR2 from the other two, due to the fact that the computation of the particular estimator is based upon the term $E_\theta \beta_n$. For this subsection we denote $b_0 = b(\theta_0)$, $b_{0,j}$ is the j^{th} element of b_0, $W_0^* = W^*(\theta_0)$, $W_{0,rj}^*$ is the (r, j) element of W_0^*, $W_0^{**} = W^{**}(\theta_0)$ and $W_{0,rj}^{**}$ is the (r, j) element of W_0^{**}.

GMR1 Estimator We begin with the GMR1 estimator. The results reveal aspects of the previous remark. The estimator is generally second order biased due to the relation between p and q, the general non linearity of the binding function, the behavior of the weighting matrix and through this of the initial estimator θ_n^*.

Lemma 4.3 *If $\sqrt{n}(\theta_n - \theta_0)$ has a valid Edgeworth expansion of third order, then*

$$\sqrt{n}(\theta_n - \theta_0) \underset{1/2}{\sim} q_1 + \frac{q_2}{\sqrt{n}}$$

where

$$q_1 = \Gamma \frac{\partial b_0^{/}}{\partial \theta} W_0^* k_1,$$

$$\begin{aligned} q_2 &= \Gamma \frac{\partial b_0^{/}}{\partial \theta} w^*(z, \theta_0) \left[Id_{l \times l} - \frac{\partial b_0}{\partial \theta^{/}} \Gamma \frac{\partial b_0^{/}}{\partial \theta} W_0^* \right] k_1 \\ &+ \Gamma \frac{\partial b_0^{/}}{\partial \theta} W_0^* \left[k_2 - \frac{1}{2} \left\{ q_1^{/} \frac{\partial^2 b_{0,j}}{\partial \theta \partial \theta'} q_1 \right\}_{j=1,\ldots,l} \right] \\ &+ \Gamma \left[\frac{\partial b_0^{/}}{\partial \theta} \left\{ \frac{\partial W_{0,rj}^*}{\partial \theta} q_1^* \right\}_{r,i=1,\ldots,l} + \left\{ q_1^{/} \frac{\partial^2 b_{0,j}}{\partial \theta \partial \theta'} \right\}_{j=1,\ldots,q} W_0^* \right] A k_1, \end{aligned}$$

$\Gamma = \left(\frac{\partial b_0^{/}}{\partial \theta} W_0^* \frac{\partial b_0}{\partial \theta^{/}}\right)^{-1}$, $A = Id_{q\times q} - \frac{\partial b_0^{/}}{\partial \theta}\Gamma \frac{\partial b_0}{\partial \theta^{/}} W_0^*$, q_1^* *is the relevant term of the analogous expansion of the initial estimator,* θ_n^**, due to assumption A.12, and* k_1 *and* k_2 *are given in 4.2.*

Notice that, when $p = q$, we have that $\frac{\partial b_0}{\partial \theta^{/}} \left(\frac{\partial b_0^{/}}{\partial \theta} W_0^* \frac{\partial b_0}{\partial \theta'}\right)^{-1} \frac{\partial b_0^{/}}{\partial \theta} W_0^* = Id_{q\times q}$. Consequently, the following corollary is trivial.

Corollary 1 *When* $p = q$ *we obtain*

$$q_1 = \left(\frac{\partial b_0}{\partial \theta^{/}}\right)^{-1} k_1$$

$$q_2 = \left(\frac{\partial b_0}{\partial \theta^{/}}\right)^{-1} k_2 - \frac{1}{2}\left(\frac{\partial b_0}{\partial \theta^{/}}\right)^{-1} \left\{ q_1^{\prime} \frac{\partial^2 b_{0,j}}{\partial \theta \partial \theta'} q_1 \right\}_{j=1,\ldots,l}$$

From the corollary it is evident that in the case where $b(\theta)$ is affine and $p = q$ the estimator essentially retains the structure of the auxiliary one. Note that a trivial case in which this holds, is when β_n is a consistent estimator of θ_0. More complex cases in which this is possible are stated below.

GMR2 Estimator We continue with the case of the GMR2 estimator. Although the caveat met before, i.e. the existence of non trivial terms in the expansion due to non linearities, the expansion contains the term $-E_{\theta_0} k_2$ something that is not present in the other two, a fact that is attributed to the computation of $E_\theta \beta_n$. Although this result that it is known from the work of Gourieroux et al. [14] and Gourieroux and Monfort [12] in the case of equality of dimensions, is significantly generalized here. What is also generalized in the next subsection is the scope of the representations of the binding functions that ensure (under appropriate conditions) that the particular estimator is second order unbiased **due to the aforementioned term**.

The next preliminary expansion result concerns the approximation of derivatives of $E_\theta \beta_n$. It is based on the fcat that under assumption A.6 (see remark R.10, as well) we have that
$\left\| D^r \left[E_\theta \left(\beta_n - b(\theta) - \frac{k_1(\theta)}{\sqrt{n}} \right) \right] |_{\theta=\theta_0} \right\| ???? = \left\| D^r \left[E_\theta \left(\beta_n - b(\theta) \right) \right] |_{\theta=\theta_0} \right\|$
$\leq M \left\| E_{\theta_0} (\beta_n - b_0) \right\| = o(1)$, $r = 1, 2$.

Lemma 4.4

$$\left\| \frac{\partial}{\partial \theta^{/}} (E_\theta \beta_n) |_{\theta=\theta_0} - \frac{\partial b_0}{\partial \theta^{/}} \right\| = o(1)$$

$$\left\| \frac{\partial^2 (E_\theta \beta_n)_j}{\partial\theta\partial\theta^{/}} |_{\theta=\theta_0} - \frac{\partial^2 b_{0,j}}{\partial\theta\partial\theta^{/}} \right\| = o(1),\ j = 1, \ldots, q$$

We are now ready to state the expansion.

Lemma 4.5 *If $\sqrt{n}(\theta_n - \theta_0)$ has a valid Edgeworth expansion of third order, then*

$$\sqrt{n}(\theta_n - \theta_0) \underset{1/2}{\sim} q_1 + \frac{q_2}{\sqrt{n}}$$

where

$$q_1 = \Gamma \frac{\partial b_0^{/}}{\partial\theta} k_1,$$

$$\begin{aligned} q_2 &= \Gamma \frac{\partial b_0^{/}}{\partial\theta} w^*(z,\theta_0) k_1 - \Gamma w^*(z,\theta_0) \Gamma \frac{\partial b_0^{/}}{\partial\theta} k_1 + \Gamma \left\{ \frac{\partial^2 b_{0,j}}{\partial\theta\partial\theta^{/}} q_1 \right\}_{j=1,\ldots,q} W_0^* A k_1 \\ &+ \Gamma \frac{\partial b_0^{/}}{\partial\theta} W_0^* (k_2 - E_{\theta_0} k_2) - \frac{1}{2} \Gamma \frac{\partial b_0^{/}}{\partial\theta} W_0^* \left\{ q_1^{/} \frac{\partial^2 b_{0,j}}{\partial\theta\partial\theta^{/}} q_1 \right\}_{j=1,\ldots,q} \\ &+ \Gamma \frac{\partial b_0^{/}}{\partial\theta} \left\{ \frac{\partial W_{0,rj}^*}{\partial\theta^{/}} q_1^* \right\}_{r,i=1,\ldots,q} \left[Id_{q\times q} - \frac{\partial b_0}{\partial\theta^{/}} \Gamma \frac{\partial b_0^{/}}{\partial\theta} W_0^* \right] k_1, \end{aligned}$$

$\Gamma = \left(\frac{\partial b_0^{/}}{\partial\theta} W_0^* \frac{\partial b_0}{\partial\theta^{/}} \right)^{-1}$, $A = Id_{q\times q} - \frac{\partial b_0}{\partial\theta^{/}} \Gamma \frac{\partial b_0^{/}}{\partial\theta} W_0^*$, *$q_1^*$ is the relevant term of the initial estimator, θ_n^*, due to assumption A.12, and k_1 and k_2 are given in 4.2.*

Remark R.18 *As expected GMR1 and GTMR2 are first order equivalent as their q_1 terms coincide.*

Remark R.19 *The term $\Gamma \frac{\partial b_0^{/}}{\partial\theta} W_0^* (k_2 - E_{\theta_0} k_2)$ is obtained due to the presence of $E_\theta \beta_n$ in the definition of the estimator and not of $b(\theta)$ or something similar as in the cases of GMR1 and GT estimators.*

The following two corollaries are trivial.

Corollary 2 *When $p = q$ we obtain*

$$q_1 = \left(\frac{\partial b_0}{\partial\theta^{/}} \right)^{-1} k_1$$

$$q_2 = \left(\frac{\partial b_0}{\partial\theta^{/}} \right)^{-1} (k_2 - E_{\theta_0} k_2) - \frac{1}{2} \left(\frac{\partial b_0}{\partial\theta^{/}} \right)^{-1} \left\{ q_1^{/} \frac{\partial^2 b_{0,j}}{\partial\theta\partial\theta'} q_1 \right\}_{j=1,\ldots,l}$$

Corollary 3 *If in addition to the provisions of the previous corollary $b(\theta)$ is affine $E_{\theta_0} q_2 = \mathbf{0}_p$.*

Remark R.20 *In this particular case, the estimator is obviously **second order unbiased** a property that is not shared with its other two counterparts. This result is already known for the case where β_n is a consistent estimator of θ_0, whence the GMR2 obviously performs a second order bias correction. If in addition $E_\theta \beta_n$ is linear, then the estimator is totally unbiased (see Gourieroux et al. [14]).*

The particular analysis on the properties of the present estimator provided by the relevant literature restricts to the case of $p = q$. We extend it in the most general setup and provide a geometric characterization of the binding function, in section 4.2, that sheds light to the circumstances under which this is linear, thereby extending massively the scope of the last result.

GT Estimator We conclude the presentation of the expansions with the last case of the GT estimator. The expansion is more involved since it is obtained from the second order Taylor expansion of the first order conditions that the estimator satisfies with high probability for large enough n, around $(\theta_0, b(\theta_0))$. We shall need the following assumption:

Assumption A.13 *Integration with respect to the measures involved in the statistical model and derivation with respect to θ and β are commutative.*

Remark R.21 *This assumption can be established upon the existence of random elements such that the dominated convergence theorem applies for the elements involved in the integration and derivation procedures (see for example Davidson [6], theorem 9.31).*

Lemma 4.6 *Under assumption A.13 and if $\sqrt{n}(\theta_n - \theta_0)$ has a valid Edgeworth expansion of third order, then*

$$\sqrt{n}(\theta_n - \theta_0) \underset{1/2}{\sim} q_1 + \frac{q_2}{\sqrt{n}}$$

where

$$q_1 = J\Gamma W_0^{**} E_\beta^c k_1,$$

$$
\begin{aligned}
q_2 &= J\Gamma W_0^{**} E_\beta^c k_2 + J\Gamma \left[\frac{\partial W_{0,rj}^{**}}{\partial \theta^/} q_1^*\right]_{r,j=1,\ldots,l} E_\beta^c \left(k_1 - \frac{\partial b_0}{\partial \theta^/} q_1\right) \\
&+ J\left[\Gamma w^{**}(z,\theta_0) + \left\{\frac{\partial b_0^/}{\partial \theta} E_{\beta,\beta}^{c,j} k_1\right\}_{j=1,\ldots,l} W_0^{**}\right] E_\beta^c A k_1 \\
&- J\left\{\left(\frac{\partial b_0^/}{\partial \theta} E_{\beta,\beta}^{c,j} \frac{\partial b_0}{\partial \theta^/} - \left\{\frac{\partial b_0^/}{\partial \theta_r \partial \theta} E_{\beta,\beta}^{c,j}\right\}_{r=1,\ldots p}\right) q_1\right\}_{j=1,\ldots,l} W_0^{**} E_\beta^c \left(k_1 - \frac{\partial b_0}{\partial \theta^/} q_1\right) \\
&+ \frac{1}{2} J\Gamma W_0^{**} \left\{ \begin{array}{c} q_1^/ \frac{\partial b_0^/}{\partial \theta} E_{\beta,\beta}^{c,j} \left(\frac{\partial b_0}{\partial \theta^/} q_1 - k_1\right) - q_1^/ \left[E_{\theta_0} \frac{\partial c_n(b)_j}{\partial \beta'} \frac{\partial^2 b_0}{\partial \theta_r \partial \theta'}\right]_{r=1,\ldots p} q_1 \\ + \left(k_1^/ - q_1^/ \frac{\partial b_0^/}{\partial \theta}\right) E_{\beta,\beta}^{c,j} k_1 \end{array} \right\}_{j=1,\ldots,l},
\end{aligned}
$$

$E_{\theta_0} \frac{\partial c_n(b_0)}{\partial \beta^/} = E_\beta^c$, $E_{\theta_0} \frac{\partial^2 c_n(b_0)_j}{\partial \beta \partial \beta'} = E_{\beta,\beta}^{c,j}$, $J = \left(\frac{\partial b_0^/}{\partial \theta} \left(E_\beta^c\right)^/ W_0^{**} E_\beta^c \frac{\partial b_0}{\partial \theta^/}\right)^{-1}$, $A = Id_{q\times q} - \frac{\partial b_0}{\partial \theta^/} J \frac{\partial b_0^/}{\partial \theta} \left(E_\beta^c\right)^/ W_0^{**} E_\beta^c$, $\Gamma = \frac{\partial b_0^/}{\partial \theta} \left(E_\beta^c\right)^/$, q_1^* *is the relevant term of the initial estimator,* θ_n^**, due to assumption A.12, and* k_1 *and* k_2 *are given in 4.2.*

Remark R.22 *Again it is evident that the structure of the second order terms depends on the relevant structure of the auxiliary estimator, on non linearities of the auxiliary first order conditions, on the stochastic weighting and on the relation between* l*,* q *and* p*. This estimating procedure does not produce the term* $E_{\theta_0} k_2$ *as is also the case for the GMR1 counterpart.*

We obtain easily the following corollary that confirms the already known first order relationship between the three estimators.

Corollary 4 *GT estimator* $\underset{0}{\sim}$ *(GMR1 estimator* $\underset{0}{\sim}$ *GMR2 estimator) iff the weighting matrix for the GMR1 and the GMR2 estimators is chosen as*

$$
W^*(x_i, \theta_0) = E_{\theta_0} \frac{\partial c_n(b(\theta_0))'}{\partial \beta} W^{**}(x_i, \theta_0) E_{\theta_0} \frac{\partial c_n(b(\theta_0))}{\partial \beta'}
$$

for a given $W^{**}(x_i, \theta_0)$ *for the GT estimator.*

In the special case of equality among the involved dimensions, i.e. $p = q = l$, we obtain the following corollary, which is proven with the help of lemma AL.3 in the Appendix.

Corollary 5 *When $p = q = l$ we obtain*

$$q_1 = \left(\frac{\partial b(\theta_0)}{\partial \theta'} \right)^{-1} k_1$$

and

$$q_2 = \left(\frac{\partial b(\theta_0)}{\partial \theta'} \right)^{-1} k_2 - \frac{1}{2} \left(\frac{\partial b(\theta_0)}{\partial \theta'} \right)^{-1} \left[q_1 \frac{\partial^2 b_j(\theta_0)}{\partial \theta \partial \theta'} q_1' \right]_{j=1,\dots,l}$$

Remark R.23 *This corollary is in accordance with lemma 2.2. It shows that **neither the GT estimator is second order unbiased** under the appropriate framework.*

4.2 Local Canonical Representation of the Binding Function

In this paragraph we assume without loss of generality that Θ and B are open. By assumption A.1 the underlying statistical model has the structure of a C^k-differentiable manifold of dimension p. This manifold is **globally** diffeomorphic to Θ. Assumption A.4 enables the possibility that $c(x, \beta)$ lies on a particular bundle (Hilbert bundle, see among others Amari and Nagaoka [1]) over an auxiliary statistical model that analogously has the structure of a C^k-differentiable manifold of dimension q, **globally** diffeomorphic to B, topologized again by the total variation norm. The function $b(\theta)$ that is the crucial element of the inferential procedures, described above, is essentially a parametric representation of an underlying function (say f) between the **manifolds**, which when composed with the aforementioned diffeomorphisms gives $b(\theta)$. That is, using the notation of assumption A.1, if the auxiliary statistical manifold is denoted by $\mathcal{D}^*$ and the relevant diffeomorphism to B is $\mathbf{par}^*$, then $b = \mathbf{par}^* \circ f \circ \mathbf{par}^{-1}$. The function f shares by construction many properties with its relevant representation. That is there is a open neighborhood of P_Ω say $\mathcal{O}_{P_\Omega}$, such that f is a diffeomorphism onto $f(\mathcal{O}_{P_\Omega})$. It is easy to see that $b(\theta)$ is simply a manifestation of this property which extends to any other representation of f. That is, if Θ' is an open bounded subset of $\mathbb{R}^p$ diffeomorphic to $\mathcal{O}_{P_\Omega}$ by $\mathbf{par}_*$, and B' is an open bounded subset of $\mathbb{R}^q$ diffeomorphic to B by $\mathbf{par}^*_*$ then the relevant representation $b^* : \Theta' \to B'$ restricted as $b'|_{\mathcal{O}_{P_\Omega}} = \mathbf{par}^*_* \circ f|_{\mathcal{O}_{P_\Omega}} \circ \mathbf{par}_*^{-1}$ is a diffeomorphism. Furthermore, by theorem 10.2 of Spivak [23] (p. 44) if $p \leq q$, there always exists an open bounded subset of $\mathbb{R}^q$, say B'' diffeomorphic to $\mathcal{D}^*$ by $\mathbf{par}^*_{**}$ (hence diffeomorphic to B by (say) g), such that the representation $b^{**} : \Theta \to B''$ restricts as $b^{**}|_{\mathbf{par}^{-1}(\mathcal{O}_{P_\Omega})} = \mathbf{par}^*_{**} \circ f|_{\mathcal{O}_{P_\Omega}} \circ \mathbf{par}^{-1} = \left(\theta_1, \theta_2, \dots, \theta_p, \underbrace{0, \dots, 0}_{q-p} \right)$. This representation is called canonical immersion around P_Ω. Hence, following the proof of theorem 10.2 of Spivak [23] and noting that the target of the constructed coordinate system of $\mathcal{D}^*$ that proves the theorem is diffeomorphic to the one of the initial coordinate system on the same manifold, the next lemma is easily proven.

Lemma 4.7 *There exists an open bounded subset of* $\mathbb{R}^q$*, say* B''*, and a diffeomorphism* $g : B \to B''$ *such that* $b^{**}|_{\mathbf{par}^{-1}(\mathcal{O}_{P_\Omega})} : \mathcal{O}_{\varepsilon_2}(\theta_0) \to B''$ *is given by* $b^{**}(\theta) = \begin{pmatrix} \theta \\ \mathbf{0}_{q-p} \end{pmatrix}$ $\forall \theta \in \mathbf{par}^{-1}(\mathcal{O}_{P_\Omega})$.

Remark R.24 *Given* Θ*,* B ***can always be chosen*** *so that the binding function* b *is of the form* $\begin{pmatrix} \theta \\ \mathbf{0}_{q-p} \end{pmatrix}$ *at least in a small enough neighborhood of* θ_0*. We call this canonical representation of the binding function around* θ_0 *and, from this point until the end of the present subsection, we denote it by* $b(\theta)$*.*[10] *It is easily seen that when* $b(\theta)$ *is on the relevant form, the aforementioned expansions simplify in some extend. We explore some interesting cases. In every one of these we assume that* $W^*(x,\theta_0) = W^* = \begin{pmatrix} W_1 & W_3 \\ W_3' & W_2 \end{pmatrix}$ *where* W_1 *is* $p \times p$*,* W_2 *is* $(q-p) \times (q-p)$*, and* W_3 *is* $p \times (q-p)$ *and they are non stochastic independent of* θ*.*

Let us consider first the expansion of the GMR1 estimator. Noting first that $\frac{\partial b(\theta_0)}{\partial \theta'} = \begin{pmatrix} Id_{p\times p} \\ \mathbf{0}_{(q-p)\times p} \end{pmatrix}$ and $\frac{\partial b^2(\theta_0)_j}{\partial \theta \partial \theta'} = \mathbf{0}_{p\times p}$, $\forall j = 1, \ldots, q$, directly substituting in the results of lemma 4.3 we trivially get the following corollary.

Corollary 6 *Consider lemma 4.3, suppose that* $b(\theta)$ *is in local canonical form and* $W^*(x,\theta_0) = W^* = \begin{pmatrix} W_1 & W_3 \\ W_3' & W_2 \end{pmatrix}$ *then*

$$q_1 = \begin{pmatrix} Id_{p\times p} & W_1^{-1} W_3 \end{pmatrix} k_1$$

and

$$q_2 = \begin{pmatrix} Id_{p\times p} & W_1^{-1} W_3 \end{pmatrix} k_2$$

Remark R.25 *It is evident that* $\min_{W_3} \|E_{\theta_0} q_2\| = \left\| \begin{pmatrix} (E_{\theta_0} k_2)_1 \\ \vdots \\ (E_{\theta_0} k_2)_p \end{pmatrix} \right\|$ *for* $W_3 = \mathbf{0}_{p\times q-p}$ *where* $(E_{\theta_0} k_2)_i$ *denotes the* i^{th} *element of the particular vector.*

[10] This abuse of notation can not create any problem of confusion until the end of the current subsection. Later on and where needed we will distinguish the notations explicitely.

The analogous results for the GT estimator are not considered here due to the fact that they constitute an easy exercise without providing any new information. The second and final case concerns the GMR2 estimator. Again, direct substitutions on the results of lemma 4.5 and taking into account $\frac{\partial b(\theta_0)}{\partial \theta'} = \begin{pmatrix} Id_{p\times p} \\ \mathbf{0}_{q-p\times p} \end{pmatrix}$, and $\frac{\partial b^2(\theta_0)_j}{\partial\theta\partial\theta'} = \mathbf{0}_{p\times p}$, $\forall j = 1,\dots,q.$, we get the obvious corollary.

Corollary 7 *Consider lemma 4.5, suppose that $b(\theta)$ is in local canonical form and $W^*(x,\theta_0) = W^* = \begin{pmatrix} W_1 & W_3 \\ W_3' & W_2 \end{pmatrix}$ then*

$$q_1 = \begin{pmatrix} Id_{p\times p} & W_1^{-1}W_3 \end{pmatrix} k_1$$

and

$$q_2 = \begin{pmatrix} Id_{p\times p} & W_1^{-1}W_3 \end{pmatrix} (k_2 - E_{\theta_0}k_2)$$

Remark R.26 *The GMR2 estimator is second order unbiased even in cases where $q > p$, when there is non stochastic weighting given that the binding function is in local canonical representation. This is a new result. First it extends the relevant result of the aforementioned literature to allow for cases of differing dimensions, as long as the Hessian matrices of the binding function vanish and the weighting is deterministic. Second, since the binding function can always be in local canonical form, there always exists a parameterization of c so that the previous statement holds. This says that* ***given an admissible auxiliary statistical model, there always exists an auxiliary parameterization such that the previous result is valid****, proviso the relevant weighting structure. Hence this result massively generalizes the one in the relevant literature.*

Let us continue with an example. In this, lemma 2.2 holds for any n due to global invertibility of the corresponding binding functions and the absence of boundaries.

Example Consider the case in which the true underlying distribution is described by the following MA(1) specification

$$x_t = u_t + \theta_0 u_{t-1}, \qquad t = ..., -1, 0, 1, ..., \quad u_t \overset{iid}{\backsim} N(0,1)$$

for *some* $\theta_0 \in (-1,1)$, while the auxiliary model is consisted of all the joint distributions represented by the following parametric AR(1) model

$$x_t = \beta x_{t-1} + \varepsilon_t, \qquad t = ..., -1, 0, 1, ..., \quad \varepsilon_t \overset{iid}{\backsim} N(0,1)$$

where $\beta \in \left(-\frac{1}{2}, \frac{1}{2}\right)$. Let β_n be the conditional maximum likelihood estimator for the previous model, i.e. $\beta_n = \frac{\sum_{i=2}^{n} x_i x_{i-1}}{\sum_{i=2}^{n} x_{i-1}^2}$, which is easily seen that converges in probability to $b(\theta_0) = \frac{\theta_0}{1+\theta_0^2}$. Hence in this particular case $p = q = l = 1$, $c(x_i, \beta) = x_i x_{i-1} - \beta x_{i-1}^2$, and $b : (-1, 1) \to \left(-\frac{1}{2}, \frac{1}{2}\right)$ is globally invertible. We obtain from Demos and Kyriakopoulou [7]

$$k_1 = \frac{\left(\theta_0^2 + 4\theta_0^4 + \theta_0^6 + \theta_0^8 + 1\right)^2}{1 - \theta_0^2} z$$

$$k_2 = -\left(\theta_0^4 + 2\theta_0^3 - 2\theta_0^2 + 2\theta_0 + 1\right) \frac{\theta_0^2 + \theta_0 + 1}{\left(\theta_0^2 + 1\right)^3} z^2$$

In the case of the GMR1 estimator equal to GT estimator which is $\theta_n = \frac{1 - \sqrt{-4\beta_n^2 + 1}}{2\beta_n}$ we obtain from corollary 1

$$q_1 = \frac{\left(1 + \theta_0^2\right)^2 \left(\theta_0^2 + 4\theta_0^4 + \theta_0^6 + \theta_0^8 + 1\right)^2}{\left(1 - \theta_0^2\right)^2} z$$

$$\begin{aligned} q_2 = & -\frac{\left(\theta_0^4 + 2\theta_0^3 - 2\theta_0^2 + 2\theta_0 + 1\right)\left(\theta_0^2 + \theta_0 + 1\right)}{1 - \theta_0^4} z^2 \\ & -\frac{\theta_0\left(\theta_0^2 - 3\right)\left(\theta_0^2 + 4\theta_0^4 + \theta_0^6 + \theta_0^8 + 1\right)^4}{\left(1 - \theta_0^4\right)\left(1 - \theta_0^2\right)^2} z^2 \end{aligned}$$

Notice that when $\theta_0 = 0$, then $q_1 = z$, and $q_2 = z^2$. Finally, for the GMR2 estimator we obtain from corollary 2 that

$$q_1 = \frac{\left(1 + \theta_0^2\right)^2 \left(\theta_0^2 + 4\theta_0^4 + \theta_0^6 + \theta_0^8 + 1\right)^2}{\left(1 - \theta_0^2\right)^2} z$$

$$\begin{aligned} q_2 = & -\frac{\left(\theta_0^4 + 2\theta_0^3 - 2\theta_0^2 + 2\theta_0 + 1\right)\left(\theta_0^2 + \theta_0 + 1\right)}{1 - \theta_0^4} \left(z^2 - 1\right) \\ & -\frac{\theta_0\left(\theta_0^2 - 3\right)\left(\theta_0^2 + 4\theta_0^4 + \theta_0^6 + \theta_0^8 + 1\right)^4}{\left(1 - \theta_0^4\right)\left(1 - \theta_0^2\right)^2} z^2 \end{aligned}$$

which implies that the estimator is unbiased at $\theta_0 = 0$ but not locally unbiased (see bellow). Now, for the issue of the local canonical form of the binding function, we obtain that the local parametrization of the AR(1) model arises from the re-parametrization given by $\beta^* = \frac{1-\sqrt{1-4\beta^2}}{2\beta}$, and in this case $b^*(\theta) = \theta$, for any θ. Notice that a consistent auxiliary estimator for $b^*(\theta_0) = \theta_0$ is $\beta_n^* = \frac{1-\sqrt{1-4\beta_n^2}}{2\beta_n}$, and the GMR2 estimator derived by this is second order unbiased by lemma 7.

4.3 GMR2 Recursion

In this section we are concerned with the generalization of the previous properties of the GMR2 estimator to arbitrary order. First we make the distinction between several notions of unbiaseness of a given order. An estimator (say θ_n) admitting a moment expansion (say $g\left(z, \frac{1}{\sqrt{n}}, \theta_0\right)$ for g a relevant function) such as the aforementioned, will be termed s^{th}-order unbiased at θ_0, if and only if $\sqrt{n}\left(\theta_n - \theta_0\right) \underset{\frac{(s-1)}{2}}{\sim} g\left(z, \frac{1}{\sqrt{n}}, \theta_0\right)$ with $E\left(g\left(z, \frac{1}{\sqrt{n}}, \theta_0\right)\right) = 0$. Analogously it will be termed s^{th}-order unbiased locally around θ_0, if the relevant expansion is valid, and $\sqrt{n}\left(\theta_n - \theta\right) \underset{\frac{(s-1)}{2}}{\sim} g\left(z, \frac{1}{\sqrt{n}}, \theta\right)$ with $E\left(g\left(z, \frac{1}{\sqrt{n}}, \theta\right)\right) = 0$ in an *open ball* with center θ_0. Finally, it will be termed s^{th}-order unbiased if the relevant expansion is valid in every neighborhood of θ_0, and $\sqrt{n}\left(\theta_n - \theta\right) \underset{\frac{(s-1)}{2}}{\sim} g\left(z, \frac{1}{\sqrt{n}}, \theta\right)$ with $E\left(g\left(z, \frac{1}{\sqrt{n}}, \theta\right)\right) = 0$ everywhere. Notice that up to the previous section we were essentially concerned with the first notion.

Now, the set up enabling lemma 4.7, concerning the local canonical representation of the binding function $b(\theta)$, implies that if cofinitely $E_\theta \beta_n$ *is a local diffeomorphism*, there exists a *sequence* of local auxiliary parameterizations, for which $E_\theta(\beta_n)$ are in canonical form in a neighborhood of θ_0. In this case the GMR2 estimator is unbiased, i.e. if $\forall \theta \in \mathcal{O}_\varepsilon(\theta_0)$ we have that $b_n(\theta) = E_\theta \beta_n^* = \begin{pmatrix} \theta \\ 0_{q-p} \end{pmatrix}$, the GMR2 is given by $\theta_n = b_n^{-1} \circ \beta_n^*$ and we have that $E_{\theta_0}\theta_n = E_{\theta_0}\left(b_n^{-1} \circ \beta_n^*\right) = b_n^{-1} \circ E_{\theta_0}\left(\beta_n^*\right) = b_n^{-1} \circ b_n(\theta_0) = \theta_0$. Consequently, a natural question arises whether it is possible to retrieve this sequence. This question is out of the scope of the present paper.

For an indirect answer to the aforementioned question, we define recursive indirect estimation procedures as follows. Let $\theta_n^{(0}$ denote either the GT or the GMR1 estimator.

Definition D.6 *Let $r \in \mathbb{N}$, the recursive $r-$GMR2 estimator $(\theta_n^{(r})$ is defined in the following steps:*

1. $\theta_n^{(1} = \arg\min_\theta \left\| \theta_n^{(0} - E_\theta \theta_n^{(0} \right\|$,

2. *for* $r > 1$ $\theta_n^{(r} = \arg\min_\theta \left\| \theta_n^{(r-1} - E_\theta \theta_n^{(r-1} \right\|$.

Remark R.27 *In the case where* $r = 1$ *we essentially obtain equivalent results to the ones of the canonical representation paragraph, due to the fact that this procedure imitates the expression of the binding function in local canonical form. Hence the case of* $r = 1$ *can be perceived as "practically" equivalent to the procedure described in the previous section. Furthermore, when* $p = q$*, then this equivalence is actually an equality.*

In order to establish the validity of the results to be presented, we need to strengthen in some sense assumptions A.3 and A.6.

Assumption A.14 $E(k_i(\theta, z))$ *are* d*-differentiable at* θ_0 *and* $n^a \left\| D^r \left(E_\theta \beta_n - b(\theta) - \sum_{i=1}^{2a+1} \frac{1}{n^{i/2}} E(k_i(\theta, z)) \right) \right\| |_{\theta=\theta_0} = o(1)$, $r = 1, \ldots, d$.

Remark R.28 *The assumption above is satisfied if* $E_\theta \beta_n = b(\theta) + \sum_{i=1}^{\infty} \frac{1}{n^{i/2}} E(k_i(\theta, z))$, $\forall \theta \in \mathcal{O}_{\varepsilon_5}(\theta_0)$, *for some* $\varepsilon_5 > 0$, $\sum_{i=1}^{\infty} \| D^r E(k_i(\theta, z)) |_{\theta=\theta_0} \| < M_r^{**}$, *for* $M_r^{**} > 0$, *since in this case we have that* $n^a \left\| E_\theta \beta_n - b(\theta) - \sum_{i=1}^{2a+1} \frac{1}{n^{i/2}} E(k_i(\theta, z)) \right\| = \left\| \sum_{i=2a+2}^{\infty} \frac{1}{n^{i/2-a}} E(k_i(\theta, z)) \right\|$ *and therefore* $n^a \left\| D^r \left(E_\theta \beta_n - b(\theta) - \sum_{i=1}^{2a+1} \frac{1}{n^{i/2}} E(k_i(\theta, z)) \right) \right\| |_{\theta=\theta_0} = \left\| \sum_{i=2a+2}^{\infty} \frac{1}{n^{i/2-a}} D^r E(k_i(\theta, z)) \right\| \leq \sum_{i=2a+2}^{\infty} \frac{1}{n^{i/2-a}} \| D^r E(k_i(\theta, z)) \| = o(1)$. *Notice that* $E_\theta \beta_n = b(\theta) + \sum_{i=1}^{\infty} \frac{1}{n^{i/2}} E(k_i(\theta, z))$ *will follow if the assumptions depending on* a *are strengthened in order to hold for any* a*, due to the fact that* θ_0 *is arbitrary, while the derivative summability condition will follow from relevant arguments concerning the derivation of series.*

Now, we can prove the following proposition. Notice that the validity of the approximations rely on the relevant results addressed in the previous sections and the previous assumption, hence we do not explicitly describe them.

Proposition 8 *With the above notation, let lemma 4.5 or lemma 4.6 hold* ***locally*** *around* θ_0*, then the* $r-$*GMR2 estimator, is of order* $2r + 1$ *unbiased at* θ_0*.*

Remark R.29 *Consider again the case where* $r = 1$*. Then* $1-$*GMR2 is actually third order unbiased at* θ_0 *hence the previous results are essentially expanded if* $\theta_n^{(0}$ *has a local moment approximation.*

Remark R.30 *Proposition 8 essentially holds* ***locally*** *at θ_0 due to the properties of open balls as basic sets of neighborhoods (see also the example below).*

It is worth mentioning that the recursive GMR2 procedure is a generalization of iterated bootstrap. To elaborate on this, consider the case that we have the GMR1 estimator of θ. Bootstrapping this estimator is equivalent to 1-step GMR2 estimation (1-step in the spirit of Andrews [2]) on GMR1 (see Gourieroux et al. [14] section 1.5). Bootstrapping the bootstrapped GMR1 is equivalent to 1-step GMR2 on 1-step GMR2 on GMR1 etc. Consequently, the iterated bootstrap estimator is a recursive 1-step GMR2, on every recursion.

Let us now return to our example.

Example (continued) Now, from the local canonical form of the binding function, section 4.2, we obtained that the local parametrization as the reparametrization given by $\beta^* = \frac{1-\sqrt{1-4\beta^2}}{2\beta}$. The particular reparameterization and the employment of GMR2 on it, coincides (see remark R.27) with the defined $1-$GMR2. The analogous expansion of the auxiliary estimator (or equivalently of $\theta_n^{(0)}$) coincides with the one of the GMR1 presented above. For the bias corrector GMR2 (or equivalently $\theta_n^{(1)}$) we have that

$$q_1 = \frac{\left(1+\theta_0^2\right)^2\left(\theta_0^2+4\theta_0^4+\theta_0^6+\theta_0^8+1\right)^2}{\left(1-\theta_0^2\right)^2} z$$

$$\begin{aligned} q_2 \quad = \quad & -\frac{\left(\theta_0^4+2\theta_0^3-2\theta_0^2+2\theta_0+1\right)\left(\theta_0^2+\theta_0+1\right)}{1-\theta_0^4}\left(z^2-1\right) \\ & -\frac{\theta_0\left(\theta_0^2-3\right)\left(\theta_0^2+4\theta_0^4+\theta_0^6+\theta_0^8+1\right)^4}{\left(1-\theta_0^4\right)\left(1-\theta_0^2\right)^2}\left(z^2-1\right) \end{aligned}$$

establishing the second order unbiaseness. Notice that due to the global (hence local) nature of the moment approximation of Demos and Kyriakopoulou [7], proposition 8 holds globally, establishing that $\sqrt{n}\left(\theta_n^{(1}-\theta\right) \underset{1}{\sim} q_1 + \frac{q_2}{\sqrt{n}}$, which is also in accordance with the third order approximation actually employed in Demos and Kyriakopoulou [7].

5 Conclusions

In this section we first provide a brief review of our results. First, we provide conditions that ensure the validity of the formal Edgeworth approximation of the auxiliary and the three IE for any finite order. The aforementioned validation was previously unattained by the relevant literature. Second, given the validity, we provide integrability conditions that validate moment approximations of the aforementioned estimators. These conditions validate the partial results of the relevant literature. Third, we provide a general definition of estimators as the GT one, even when the auxiliary criterion is not of the likelihood type. Note that this type of estimators are eligible to more general definitions. Fourth, we provide new results on the issue of second order properties of the three indirect estimators, i.e. the expansions of GMR1 and GT estimators are new and reveal a higher order asymptotic inequivalence with the GMR2. Fifth, we massively generalize the GMR2 expansion. We are able to generalize the conditions under which the GMR2 is second order unbiased (at θ_0) even in this set up. Sixth, we characterize the fact that due to the notion of the local canonical form of the binding function, there always exists a parameterization of the auxiliary model, under which the GMR2 is second order unbiased under constant weighting. Finally, in response to the issue of higher order bias correction, we define indirect estimators that emerge from multistep optimization procedures. Strengthening the previous results, with a view towards local validity of the relevant moment approximations, we are able to provide recursive indirect estimators that are locally unbiased at any given order.

An application of the Edgeworth approximations could lay in the derivation of properties of indirect testing procedures. Furthermore, the extension of the previous results in a semiparametric setup or in non-standard cases, i.e. when $b(\theta_0)$ is in the boundary of B even if θ_0 is in the interior of Θ, could be of interest. Additionally, the determination of invariant parts of the expansions with respect to reparameterizations could be very fruitful. An application of the results in Andrews [2] on the iterated bootstrap could make possible a detailed examination of the relation between this estimator and the recursive GMR2 one. Our results, could also be extended in the more general case, where the auxilliary parametric space is naturally or arbitrarily restricted with equality and/or inequality restrictions (see Calzolari, Fiorentini and Sentana [5]). We leave all these questions for future work.

References

[1] Amari Sh.-Ich. and Nagaoka H., "Methods of Information Geometry", Trans. of Mathematical Monographs, AMS, Ox. Un. Press, 2000.

[2] Andrews D.W.K. (2002), "Equivalence of the Higher-order Asymptotic Efficiency of k-step and Extremum Statistics," *Econometric Theory*, 18, pp. 1040–1085.

[3] Arvanitis S. and A. Demos (2010) "Stochastic Expansions and Moment Approximations for three Indirect Estimators" mimeo AUEB, http://www.aueb.gr/users/demos/WorkingPapers/bias-ii.pdf

[4] Bhattacharya R.N. , and J.K. Ghosh (1978),"On the validity of the formal Edgeworth expansion", *The Annals of Statistics,* Vol. 6, No. 2, pp. 434-451.

[5] Calzolari, C., Fiorentini G. and E. Sentana (2004), Constrained Indirect Inference Estimation, *Review of Economic Studies* **71***, pp.* 945-973*.*

[6] Davidson J. "Stochastic Limit Theory" Oxford University Press, 1994.

[7] Demos A. and Kyriakopoulou D., "Edgeworth expansions for the MLE and MM estimators of an MA(1) model", *mimeo*, AUEB, 2008.

[8] Dridi R., and E. Renault,"Semi-parametric Indirect Inference", mimeo, LSE and CREST-INSEE and Universite de Paris IX Dauphine, 2000.

[9] Dupacova, J. and R. Wets (1988), "Asymptotic behaviour of Statistical Estimators and of Optimal solutions of Stochastic Optimization Problems", *The Annals of Statistics,* Vol. 16, No. 4, pp. 1517-1549.

[10] Gallant A. Ronald, and Tauchen G.E. (1996), "Which Moments to Match," *Econometric Theory*, Vol. 12, pp. 657-681.

[11] Goetze F., and C. Hipp (1983), "Asymptotic expansions for sums of weakly dependent random vectors", *Probability Theory and Related Fields*, 64, pp. 211-239.

[12] Gourieroux C.,and A. Monfort, "*Simulation-based Econometric Methods*", CORE Lectures, Ox. Un. Press, 1996.

[13] Gourieroux C., A. Monfort, and E. Renault (1993), "Indirect Inference", *Journal of Applied Econometrics*, Volume 8 Issue 1, pp. 85-118.

[14] Gourieroux C., E. Renault and N. Touzi, 2000, "Calibration by simulation for small sample bias correction", in "*Simulation-Based Inference in Econometrics: Methods and Applications*", Cambridge University Press, Cambridge.

[15] Gradshteyn I. S., and Ryzhik I. M., "Table of Integrals, Series, and Products" A. Jeffrey Editor, Academic Press 1994.

[16] Linton, O. 1997. An asymptotic expansion in the GARCH(1,1) model. *Econometric Theory* 13, 558-581.

[17] Magdalinos M.A. (1992), "Stochastic Expansions and Asymptotic Approximations", *Econometric Theory*, Vol. 8, No. 3, pp. 343-367.

[18] McCullagh P. (1987), "*Tensor Methods in Statistics*" Chapman and Hall, London, New York.

[19] Newey, W.K., and R.J. Smith (2001), "*Asymptotic Bias and Equivalence of GMM and GEL Estimators*", mimeo, Department of Economics, MIT.

[20] Phillips P.C.B. (1980). Finite sample theory and the distribution of alternative estimators of the marginal propensity to consume. *Review of Economic Studies* 47, 183-224.

[21] Sargan, J.D. 1976. Econometric estimators and the Edgeworth approximation. *Econometrica* 44, 421-448.

[22] Smith A.A. (1993), "Estimating Nonlinear Time-series Models using Simulated Vector Autoregressions", *Journal of Applied Econometrics,* Volume 8 Issue 1, pp. 63-84.

[23] Spivak Michael, (1999) "*A Comprehensive Introduction to DIFFERENTIAL GEOMETRY*", Vol. 1, Publish or Perish, Inc, Texas.

[24] van der Vaart, A.W. and J.A. Wellner, (2000) "Weak Convergence and Empirical Processes" Springer.

Appendices

A Proofs of Lemmas and Corollaries.

Proof of Lemma 2.2. When $p = q = l$ due to consistency, the GT estimator satisfies with probability $1 - o(n^{-a})$

$$E_{\theta_n} c_n(\beta_n) = \mathbf{0}_p$$

yet from assumption A.4 we have that

$$E_{\theta_n} c_n(\beta) = \mathbf{0}_p \text{ iff } \beta = b(\theta_n)$$

hence the estimator equivalently satisfies

$$\beta_n - b(\theta_n) = \mathbf{0}_p$$

which defines the GMR 1 estimator in these special circumstances. ■

Proof of Lemma 2.3. In the first case we have that $\beta_n = g^{-1} \circ \frac{1}{n} f(x_i)$, $b(\theta) = g^{-1} \circ E_\theta f(x_i) = g^{-1} \circ m(\theta)$, GMR1$= m^{-1} \circ g \circ \beta_n = m^{-1} \circ \frac{1}{n} f(x_i)$. For the second case, if g is linear then $E_\theta \beta_n = g^{-1} \circ E_\theta \frac{1}{n} f(x_i) = g^{-1} \circ m(\theta) = b(\theta)$, and the result follows. ■

Proof of Lemma 3.1. Notice that assumptions 1-4 in Andrews [2] correspond to assumptions A.1, A.4, A.5, and A.7-A.9. The result follows from Lemmas 5 and 9 of Andrews [2]. ■

Proof of Lemma 3.2. i) (**GT**) We denote by $E_\theta(c_n(\beta_n))$, the quantity $E_\theta(c_n(\beta))|_{\beta=\beta_n}$. Let, in the definition of θ_n (D.4), W_n^{**} denote the weighting matrix $W_n^{**}(\theta_n^*)$. By Lemma AL.1 we have that $P_{\theta_0}(\|W_n^{**} - W_0^{**}\| > \varepsilon) = o(n^{-a})$, for $\varepsilon > 0$ and $W_0^{**} = W^{**}(\theta_0)$, and it follows that for $K > 0$ we have:

$$P_{\theta_0}(\|W_n^{**}\| > K) = o\left(n^{-a}\right) \quad (1)$$

where $K \geq \|W^{**}\| + \varepsilon$. Notice that this result is true for $W_n^*(\theta_n^*)$, the matrix employed in either GMR1 or GMR2. Further, due to lemma 3.1 and by assumption A.4 we have

$$P_{\theta_0}\left(\sup_\theta \|E_\theta c_n(\beta_n) - E_\theta c_n(b_0)\| > \varepsilon\right) = o\left(n^{-a}\right) \quad for\ \varepsilon > 0, \quad (2)$$

where $b_0 = b(\theta_0)$ Consequently, the consistency of θ_n follows from Lemma 5 of Andrews [2]. Hence θ_n is in the interior of Θ and $\frac{\partial J_n(\theta_n)}{\partial\theta} = 0$ with probability $1 - o(n^{-a})$. It follows that element by element mean value expansions of $\frac{\partial J_n(\theta_n)}{\partial\theta}$ around θ_0 and rearrangement gives: $\theta_n - \theta_0 = -\left(\frac{\partial^2 J_n(\theta_n^+)}{\partial\theta\partial\theta'}\right)^{-1} \frac{\partial J_n(\theta_0)}{\partial\theta}$

with probability $1-o(n^{-a})$, where θ_n^+ lies between θ_n and θ_0 and may be different across rows. Hence it suffices to show that there are C^* and K^+ positive reals such that

$$P_{\theta_0}\left(\left\|\frac{\partial}{\partial\theta}J_n(\theta_0)\right\| > C^*\frac{\ln^{1/2} n}{n^{1/2}}\right) = o\left(n^{-a}\right) \tag{3}$$

and

$$P\left(\left\|\left(\frac{\partial^2}{\partial\theta\partial\theta^{/}}J_n\left(\theta_n^+\right)\right)^{-1}\right\| > K^+\right) = o\left(n^{-a}\right). \tag{4}$$

We have that $\left\|\frac{\partial J_n(\theta_0)}{\partial\theta}\right\| = \left\|2\frac{\partial E_{\theta_0}\left[c^{/}(\beta_n)\right]}{\partial\theta}W_n^{**}E_{\theta_0}\left[c\left(\beta_n\right)\right]\right\|$. First, notice that by equation (1) there is $K>0$ such that $P_{\theta_0}\left(\|W_n^{**}\|>K\right)=o\left(n^{-a}\right)$. Second, as $E_{\theta_0}\left[c\left(b_0\right)\right]=0$, due to assumption A.4 and by lemma 3.1 $P_{\theta_0}\left(\|E_{\theta_0}\left[c\left(\beta_n\right)\right]\| > C\frac{\ln^{1/2} n}{n^{1/2}}\right) \leq$ $P_{\theta_0}\left(\|\beta_n - b_0\| > \frac{C}{E_{\theta_0}\|u_c(x_i)\|}\frac{\ln^{1/2} n}{n^{1/2}}\right) = o\left(n^{-a}\right)$. Finally, by assumption A.4 $\forall\varepsilon>0$ we have that
$P_{\theta_0}\left(\left\|\frac{\partial E_{\theta_0}\left[c^{/}(\beta_n)\right]}{\partial\theta} - \frac{\partial E_{\theta_0}\left[c^{/}(b_0)\right]}{\partial\theta}\right\| > \varepsilon\right) \leq P_{\theta_0}\left(M^*\left\|E_{\theta_0}\left[c\left(\beta_n\right)\right] - E_{\theta_0}\left[c\left(b_0\right)\right]\right\| > \varepsilon\right) =$ $o\left(n^{-a}\right)$. Hence, there is $K^{**}>0$ such that $P_{\theta_0}\left(\left\|\frac{\partial E_{\theta_0}\left[c^{/}(\beta_n)\right]}{\partial\theta}\right\| > K^{**}\right) =$ $o\left(n^{-a}\right)$. Applying now lemma AL.2 the proof of equation (3) is complete.
Now to prove equation (4) notice, first, that $P_{\theta_0}\left(\left\|2\frac{\partial E_{\theta_n^+}\left[c^{/}(\beta_n)\right]}{\partial\theta}W_n^{**}\frac{\partial E_{\theta_n^+}\left[c(\beta_n)\right]}{\partial\theta^{/}}\right\| > \frac{K^+}{2}\right)$

$\leq P_{\theta_0}\left(\left\|\frac{\partial E_{\theta_n^+}\left[c^{/}(\beta_n)\right]}{\partial\theta}\right\|^2 \|W_n^{**}\| > \frac{K^+}{4}\right)$

$\leq P_{\theta_0}\left(\|W_n^{**}\|>K^*\right) + P_{\theta_0}\left(\left\|\frac{\partial E_{\theta_n^+}\left[c^{/}(\beta_n)\right]}{\partial\theta}\right\|^2 \|W_n^{**}\| > \frac{K^*}{4} \cap \|W_n^{**}\| \leq K^*\right)$

$\leq o\left(n^{-a}\right) + P_{\theta_0}\left(\left\|\frac{\partial E_{\theta_n^+}\left[c^{/}(\beta_n)\right]}{\partial\theta}\right\|^2 \|W_n^{**}\| > \frac{K^*}{4} \cap \|W_n^{**}\| \leq K^*\right)$

$= o\left(n^{-a}\right) + P_{\theta_0}\left(\left\|\frac{\partial E_{\theta_n^+}\left[c^{/}(\beta_n)\right]}{\partial\theta}\right\| > K^{++}\right)$ where $K^{++} \geq \sqrt{\frac{K^+}{4K^*}}$.

Now denoting that for $\varphi_n = \left(\beta_n^{/}, \left(\theta_n^+\right)^{/}\right)^{/}$ and $\varphi_0 = \left(b_0^{/}, \theta_0^{/}\right)^{/}$ we have that
$P_{\theta_0}\left(\left\|\frac{\partial}{\partial\theta}E_{\theta_n^+}\left[c^{/}\left(\beta_n\right)\right]\right\| > K^{++}\right)$
$\leq \quad P_{\theta_0}\left(\sup_{\theta\in\mathcal{O}_\eta(\varphi_0)}\left\|\frac{\partial E_{\theta_n^+}\left[c^{/}(\beta_n)\right]}{\partial\theta}\right\| > K^{++}\right) \quad + \quad P_{\theta_0}\left(\varphi_n \notin \mathcal{O}_\eta\left(\varphi_0\right)\right)$
$\leq \quad P_{\theta_0}\left(\left\{\beta_n \notin \mathcal{O}_\eta\left(b\left(\theta_0\right)\right)\right\} \cup \left\{\theta_n^+ \notin \mathcal{O}_\eta\left(\theta_0\right)\right\}\right)$

$\leq P_{\theta_0}(\beta_n \notin \mathcal{O}_\eta(b(\theta_0))) + P_{\theta_0}(\theta_n \notin \mathcal{O}_\eta(\theta_0)) = o(n^{-a})$ where K^{++} can be chosen as $K^{++} \geq \max\left(\sqrt{\frac{K^+}{4K^*}}, \Lambda\right)$ where Λ is an upper bound of $\sup_{\theta \in \mathcal{O}_\eta(\varphi_0)} \left\| \frac{\partial E_\theta\left[c^{/}(\beta)\right]}{\partial \theta} \right\|$ (by assumption A.4). Hence $\frac{\partial E_{\theta_n^+}\left[c^{/}(\beta_n)\right]}{\partial \theta} W_n^{**} \frac{\partial E_{\theta_n^+}[c(\beta_n)]}{\partial \theta^{/}} \rightarrow \frac{\partial E_{\theta_0}\left[c^{/}(b_0)\right]}{\partial \theta} W_0^{**} \frac{\partial E_{\theta_0}[c(b_0)]}{\partial \theta^{/}}$ with probability $o(n^{-a})$ and $\frac{\partial E_{\theta_0}\left[c^{/}(b_0)\right]}{\partial \theta} W_0^{**} \frac{\partial E_{\theta_0}[c(b_0)]}{\partial \theta^{/}}$ non-singular. This follows from the fact that $\forall \theta \in \Theta$ we have that $E_\theta[c(b(\theta))] = 0$ and by the Implicit Function Theorem we have that $\frac{\partial E_\theta\left[c^{/}(b(\theta))\right]}{\partial \theta} = -\frac{\partial b^{/}(\theta)}{\partial \theta} \frac{\partial E_\theta\left[c^{/}(b(\theta))\right]}{\partial \beta}$. Now $\frac{\partial b^{/}(\theta)}{\partial \theta}$ is a $p \times q$ matrix and $rank\left(\frac{\partial b^{/}(\theta)}{\partial \theta}\right) = p$, by assumption A.2 above, whereas $\frac{\partial E_\theta\left[c^{/}(b(\theta))\right]}{\partial \beta}$ is an $q \times l$ matrix with $rank\left(\frac{\partial E_\theta\left[c^{/}(b(\theta))\right]}{\partial \beta}\right) = q$, by assumption A.4 above, and it follows that $rank\left(\frac{\partial E_\theta\left[c^{/}(b(\theta))\right]}{\partial \theta}\right) = p$. It follows that as W_0^{**} is non-singular, by assumption A.5 above, $rank\left(\frac{\partial E_{\theta_0}\left[c^{/}(b_0)\right]}{\partial \theta} W_0^{**} \frac{\partial E_{\theta_0}[c(b_0)]}{\partial \theta^{/}}\right) = p$.

Further we have to prove that $P_{\theta_0}\left(\left\| 2\left\{\frac{\partial^2 E_{\theta_n^+}\left[c^{/}(\beta_n)\right]}{\partial \theta_i \partial \theta_j} W_n^{**} E_{\theta_n^+}[c(\beta_n)]\right\}_{i,j=1,\ldots,p}\right\| > \frac{K^+}{2}\right) = o(n^{-a})$. In fact we can prove that for $\varepsilon > 0$ we have that

$P_{\theta_0}\left(\left|\frac{\partial^2 E_{\theta_n^+}\left[c^{/}(\beta_n)\right]}{\partial \theta_i \partial \theta_j} W_n^{**} E_{\theta_n^+}[c(\beta_n)]\right| > \varepsilon\right) = o(n^{-a})$. But

$$P_{\theta_0}\left(\left|\frac{\partial^2 E_{\theta_n^+}\left[c^{/}(\beta_n)\right]}{\partial \theta_i \partial \theta_j} W_n^{**} E_{\theta_n^+}[c(\beta_n)]\right| > \varepsilon\right)$$

$$\leq P_{\theta_0}\left(\left\|\frac{\partial^2 E_{\theta_n^+}\left[c^{/}(\beta_n)\right]}{\partial \theta_i \partial \theta_j}\right\| \left\|W_n^{**}\right\| \left\|E_{\theta_n^+}[c(\beta_n)]\right\| > \varepsilon\right)$$

$$\leq P_{\theta_0}(\|W_n^{**}\| > K) + P_{\theta_0}\left(\left\|\frac{\partial^2 E_{\theta_n^+}\left[c^{/}(\beta_n)\right]}{\partial \theta_i \partial \theta_j}\right\| \left\|W_n^{**}\right\| \left\|E_{\theta_n^+}[c(\beta_n)]\right\| > \varepsilon \cap \|W_n^{**}\| \leq K\right)$$

$$= o(n^{-a}) + P_{\theta_0}\left(\left\|\frac{\partial^2 E_{\theta_n^+}\left[c^{/}(\beta_n)\right]}{\partial \theta_i \partial \theta_j}\right\| \left\|W_n^{**}\right\| \left\|E_{\theta_n^+}[c(\beta_n)]\right\| > \varepsilon \cap \|W_n^{**}\| \leq K\right)$$

$$\leq o(n^{-a}) + P_{\theta_0}\left(\left\|\frac{\partial^2 E_{\theta_n^+}\left[c^{/}(\beta_n)\right]}{\partial \theta_i \partial \theta_j}\right\| \left\|E_{\theta_n^+}[c(\beta_n)]\right\| > \varepsilon^*\right) \text{ where } \varepsilon^* = \frac{\varepsilon}{K}$$

To prove that $P_{\theta_0}\left(\left\|\frac{\partial^2 E_{\theta_n^+}\left[c^{/}(\beta_n)\right]}{\partial \theta_i \partial \theta_j}\right\| \left\|E_{\theta_n^+}[c(\beta_n)]\right\| > \varepsilon^*\right) = o(n^{-a})$ it suffices to prove that for $\Lambda^+ > 0$ $P_{\theta_0}\left(\left\|\frac{\partial^2 E_{\theta_n^+}\left[c^{/}(\beta_n)\right]}{\partial \theta_i \partial \theta_j}\right\| > \Lambda^+\right) = o(n^{-a})$ and for $\varepsilon^{**} > 0$ $P_{\theta_0}\left(\left\|E_{\theta_n^+}[c(\beta_n)]\right\| > \varepsilon^{**}\right) = o(n^{-a})$. For the second order derivatives we have that $\forall i, j = 1, \ldots, p$ $P_{\theta_0}\left(\left\|\frac{\partial^2 E_{\theta_n^+}[c(\beta_n)]}{\partial \theta_i \partial \theta_j}\right\| > \Lambda^+\right)$

$\leq P_{\theta_0}\left(\sup_{\theta\in\mathcal{O}_\eta(\varphi_0)}\left\|\frac{\partial^2 E_{\theta_n^+}[c(\beta_n)]}{\partial\theta_i\partial\theta_j}\right\| > \Lambda^+\right) + P_{\theta_0}\left(\varphi_n \notin \mathcal{O}_\eta(\varphi_0)\right)$
$\leq P_{\theta_0}\left(\{\beta_n \notin \mathcal{O}_\eta(b(\theta_0))\} \cup \{\theta_n^+ \notin \mathcal{O}_\eta(\theta_0)\}\right)$
$\leq P_{\theta_0}\left(\beta_n \notin \mathcal{O}_\eta(b(\theta_0))\right) + P_{\theta_0}\left(\theta_n \notin \mathcal{O}_\eta(\theta_0)\right) = o\left(n^{-a}\right)$ where again Λ^+ can be chosen as an upper bound of $\sup_{\theta\in\mathcal{O}_\eta(\varphi_0)}\left\|\frac{\partial^2 E_\theta[c(\beta)]}{\partial\theta_i\partial\theta_j}\right\|$ which exists due to assumption A.4. Further, for $\varepsilon^{**} \leq \varepsilon_2$ $P_{\theta_0}\left(\left\|E_{\theta_n^+}\left[c\left(\beta_n\right)\right]\right\| > \varepsilon^{**}\right)$ as $E_{\theta_n^+}\left[c\left(\beta_n\right)\right] \to E_{\theta_0}\left[c\left(b_0\right)\right] = 0$ as $E_\theta\left[c\left(b\left(\theta\right)\right)\right] = 0$ $\forall\theta \in \mathcal{O}_{\varepsilon_2}\left(\theta_0\right)$ due to continuous mapping. It follows that $\frac{\partial^2 E_{\theta_n^+}\left[c^{/}\left(\beta_n\right)\right]}{\partial\theta_i\partial\theta_j} W_n^{**} E_{\theta_n^+}\left[c\left(\beta_n\right)\right] \to 0$ with probability $o\left(n^{-a}\right)$.
Hence $\frac{\partial^2}{\partial\theta\partial\theta^{/}} J_n\left(\theta_n^+\right) \to 2\frac{\partial E_{\theta_0}\left[c^{/}(b_0)\right]}{\partial\theta} W_0^{**} \frac{\partial E_{\theta_0}[c(b_0)]}{\partial\theta^{/}}$, a non-singular matrix, with probability $o\left(n^{-a}\right)$. It follows that $\left(\frac{\partial^2 J_n\left(\theta_n^+\right)}{\partial\theta\partial\theta^{/}}\right)^{-1} \to \left(2\frac{\partial E_{\theta_0}\left[c^{/}(b_0)\right]}{\partial\theta} W_0^{**} \frac{\partial E_{\theta_0}[c(b_0)]}{\partial\theta^{/}}\right)^{-1}$ with probability $o\left(n^{-a}\right)$ and, the proof of equation (4) is complete. Consequently, the result follows by Lemma 5 of Andrews [2].
ii) (**GMR1**) Notice that $p\lim\left(\beta_n - b\left(\theta\right)\right) = b_0 - b\left(\theta\right)$. Hence $\forall\theta \in \Theta$ and $\forall\varepsilon > 0$
$P_{\theta_0}\left(\sup_{\theta\in\Theta}\left\|\beta_n - b\left(\theta\right) - \left[b_0 - b\left(\theta\right)\right]\right\| > \varepsilon\right) = P_{\theta_0}\left(\left\|\beta_n - b_0\right\| > \varepsilon\right) = o\left(n^{-a}\right)$ from Lemma 3.1 and consequently, the consistency of θ_n follows from Lemma 5 of Andrews [2].
Hence θ_n is in the interior of Θ and $\frac{\partial J_n(\theta_n)}{\partial\theta} = 0$ with probability $1 - o\left(n^{-a}\right)$, where $J_n\left(\theta\right) = \left(\beta_n - b\left(\theta\right)\right)^{/} W_n^*\left(\beta_n - b\left(\theta\right)\right)$, and $W_n^* = W_n^*\left(\theta_n^*\right)$. With the same logic as in i) it suffices to show that

$$P_{\theta_0}\left(\left\|\frac{\partial}{\partial\theta} J_n\left(\theta_0\right)\right\| > C^* \frac{\ln^{1/2} n}{n^{1/2}}\right) = o\left(n^{-a}\right) \quad (5)$$

and

$$P\left(\left\|\left(\frac{\partial^2}{\partial\theta\partial\theta^{/}} J_n\left(\theta_n^+\right)\right)^{-1}\right\| > K\right) = o\left(n^{-a}\right). \quad (6)$$

and apply Lemma 5 of Andrews [2]. For equation (5) notice that $\frac{\partial J_n(\theta_0)}{\partial\theta} = -2\frac{\partial b_0^{/}}{\partial\theta} W_n^*\left(\beta_n - b_0\right)$ and consequently

$$P_{\theta_0}\left(\left\|\frac{\partial}{\partial\theta} J_n\left(\theta_0\right)\right\| > C^* \frac{\ln^{1/2} n}{n^{1/2}}\right) \leq P_{\theta_0}\left(\left\|W_n^*\right\|\left\|\beta_n - b_0\right\| > C\frac{\ln^{1/2} n}{n^{1/2}}\right)$$

where $C = \frac{C^*}{2\left\|\frac{\partial b_0^{/}}{\partial\theta}\right\|}$ by the submultiplicative property of the norm and by assumption A.3. Now we have that for $K^* > 0$ we have $P_{\theta_0}\left(\left\|W_n^*\right\| > K^*\right) =$

$o\left(n^{-a}\right)$, which is true from Lemma AL.1 (see also equation (1)), and $P_{\theta_0}\left(\|\beta_n - b_0\| > C\frac{\ln^{1/2} n}{n^{1/2}}\right)$ by Lemma 3.1 and the result follows by lemma AL.2.

Further, $\frac{\partial^2 J_n\left(\theta_n^+\right)}{\partial\theta\partial\theta^{/}} = 2\frac{\partial b\left(\theta_n^+\right)^{/}}{\partial\theta}W_n^*\frac{\partial b\left(\theta_n^+\right)}{\partial\theta^{/}} - 2\left[\frac{\partial^2 b\left(\theta_n^+\right)^{/}}{\partial\theta_i\partial\theta_j}W_n^*\left(\beta_n - b\left(\theta_n^+\right)\right)\right]_{i,j=1,\ldots,p}$.

It suffices to show that for $K > 0$ we have that $P_{\theta_0}\left(\left\|\frac{\partial^2 J_n\left(\theta_n^+\right)}{\partial\theta\partial\theta^{/}}\right\| > K\right) = o\left(n^{-a}\right)$. But

$$P_{\theta_0}\left(\left\|\frac{\partial^2 J_n\left(\theta_n^+\right)}{\partial\theta\partial\theta^{/}}\right\| > K\right) \leq P_{\theta_0}\left(\left\|2\frac{\partial b^{/}\left(\theta_n^+\right)}{\partial\theta}W_n^*\frac{\partial b\left(\theta_n^+\right)}{\partial\theta^{/}}\right\| > \frac{K}{2}\right) + P_{\theta_0}\left(\left\|2\left[\frac{\partial^2 b^{/}\left(\theta_n^+\right)}{\partial\theta_i\partial\theta_j}W_n^*\left(\beta_n - b\left(\theta_n^+\right)\right)\right]_{i,j=1,\ldots,p}\right\| > \frac{K}{2}\right) \tag{7}$$

Now we have from above that $P_{\theta_0}\left(\|W_n^*\| > K^*\right) = o\left(n^{-a}\right)$. Consequently, we have

$$P_{\theta_0}\left(\left\|2\frac{\partial b^{/}\left(\theta_n^+\right)}{\partial\theta}W_n^*\frac{\partial b\left(\theta_n^+\right)}{\partial\theta^{/}}\right\| > \frac{K}{2}\right) \leq o\left(n^{-a}\right) + P_{\theta_0}\left(\left\|\frac{\partial b^{/}\left(\theta_n^+\right)}{\partial\theta}\right\|^2 > K^{**}\right)$$

where $K^{**} = \frac{K}{4K^*}$. Further, for $\varepsilon^* > 0$ and due to consistency of θ_n^+ we have that $P_{\theta_0}\left(\left\|\frac{\partial b^{/}\left(\theta_n^+\right)}{\partial\theta}\right\|^2 > K^*\right) \leq o\left(n^{-a}\right) + P_{\theta_0}\left(\left\|\frac{\partial b^{/}\left(\theta_n^+\right)}{\partial\theta}\right\|^2 > K^* \cap \left\|\theta_n^+ - \theta_0\right\| \leq \varepsilon^*\right)$

Now as $\left\|\theta_n^+ - \theta_0\right\| \leq \varepsilon^*$ and choosing $\varepsilon^* \leq \varepsilon_4$ we have that $\theta_n^+ \in \mathcal{O}_{\varepsilon_4}\left(\theta_0\right)$ with probability $1 - o\left(n^{-a}\right)$, due to assumption A.6. Hence by choosing $K^* \geq \max\left(M_1^*, \frac{K}{4K^*}\right)$ we have that $P_{\theta_0}\left(\left\|\frac{\partial}{\partial\theta}b^{/}\left(\theta_n^+\right)\right\|^2 > K^* \cap \left\|\theta_n^+ - \theta_0\right\| \leq \varepsilon^*\right) = o\left(n^{-a}\right)$. Consequently

$$P_{\theta_0}\left(\left\|2\frac{\partial b^{/}\left(\theta_n^+\right)}{\partial\theta}W_n^*\frac{\partial b\left(\theta_n^+\right)}{\partial\theta^{/}}\right\| > \frac{K}{2}\right) = o\left(n^{-a}\right) \tag{8}$$

Now for any $\varepsilon > 0$ we have that $P\left(\left\|\beta_n - b\left(\theta_n^+\right)\right\| > \varepsilon\right) \leq P\left(\|\beta_n - b_0\| > \frac{\varepsilon}{2}\right) + P\left(\left\|b_0 - b\left(\theta_n^+\right)\right\| > \frac{\varepsilon}{2}\right) = o\left(n^{-a}\right)$ as the first probability is $o\left(n^{-a}\right)$ due to Lemma 3.1, and the second is also $o\left(n^{-a}\right)$ due to assumption A.2 and the consistency of θ_n^+. Further by assumption A.2, for $\varepsilon^{**} > 0$, $P_{\theta_0}\left(\left\|\frac{\partial^2 b\left(\theta_n^+\right)}{\partial\theta_i\partial\theta_j} - \frac{\partial^2 b_0}{\partial\theta_i\partial\theta_j}\right\| > \varepsilon^{**}\right) = o\left(n^{-a}\right)$ for any $i, j = 1, \ldots, p$. Consequently, $\exists K > 0$ such that

$$P_{\theta_0}\left(\left\|\left[\frac{\partial^2 b^{/}\left(\theta_n^+\right)}{\partial\theta_i\partial\theta_j}W_n^*\left(\beta_n - b\left(\theta_n^+\right)\right)\right]_{i,j=1,\ldots,p}\right\| > \frac{K}{2}\right) = o\left(n^{-a}\right). \tag{9}$$

Hence by equations (8) and (9) we have that the probability in equation (7) in $o(n^{-a})$. Consequently, equation (6) is true and the proof is complete by applying Lemma 5 of Andrews [2].

iii) (**GMR2**) Notice that $p\lim(\beta_n - E_\theta\beta_n) = b_0 - b(\theta)$. Hence $\forall\theta \in \Theta$ and $\forall\varepsilon > 0$

$P_{\theta_0}(\sup_{\theta\in\Theta}\|\beta_n - E_\theta\beta_n - [b_0 - b(\theta)]\| > \varepsilon)$

$\leq P_{\theta_0}(\|\beta_n - b_0\| + \sup_{\theta\in\Theta}\|E_\theta\beta_n - b(\theta)\| > \varepsilon)$.

Now we know from Lemma 3.1 above that $P_{\theta_0}\left(\|\beta_n - b_0\| > \frac{\varepsilon}{2}\right) = o(n^{-a})$. Hence it suffices to prove that for

$$\forall\varepsilon > 0, \ \ \exists\, n^* \in \mathbb{N} : \sup_{\theta\in\Theta}\|E_\theta\beta_n - b(\theta)\| < \varepsilon, \ \forall n > n^*.$$

For this we need to prove that first, $\|E_\theta\beta_n - b(\theta)\| \to 0$, pointwise on a dense subset of Θ, and second $\|E_\theta\beta_n - b(\theta)\|$ is asymptotically uniformly equicontinuous (due to Arzella-Ascoli Theorem). For the first one notice that $P_{\theta_0}(\|\beta_n - b(\theta_0)\| > \varepsilon) = o(n^{-a})$ and θ_0 is arbitrary. Hence, $P_\theta(\|\beta_n - b(\theta)\| > \varepsilon) = o(n^{-a})$ for any $\theta \in \Theta$. Furthermore, as B is bounded the series $\beta_n - b(\theta)$ is uniformly integrable, and as $\|E_\theta\beta_n - b(\theta)\| \leq E_\theta\|\beta_n - b(\theta)\|$ we get $\|E_\theta\beta_n - b(\theta)\| \to 0$, i.e. $\|E_\theta\beta_n - b(\theta)\| = o(1)$

For the second it suffices to prove that $E_\theta\beta_n - b(\theta)$ is uniformly Lipschitz. But $\|[E_\theta\beta_n - b(\theta)] - [E_{\theta^*}\beta_n - b(\theta^*)]\| \leq \|E_\theta\beta_n - E_{\theta^*}\beta_n\| + \|b(\theta) - b(\theta^*)\|$, and $\|b(\theta) - b(\theta^*)\| \leq k\|\theta - \theta^*\|$ by assumption A.3. Further, $\|E_\theta\beta_n - E_{\theta^*}\beta_n\| = \|[E_\theta\beta_n - b(\theta)] - [E_{\theta^*}\beta_n - b(\theta)]\|$

$= \left\|\int_{\mathbb{R}^n}(\beta_n - b(\theta))\,dP_\theta - \int_{\mathbb{R}^q}(\beta_n - b(\theta))\,dP_{\theta^*}\right\|$

$\leq q\max_{i=1,\ldots,q}\int_{\mathbb{R}^n}|(\beta_n - b(\theta))_i|\,|dP_\theta - dP_{\theta^*}|$

$\leq qM_1\int_{\mathbb{R}^n}|dP_\theta - dP_{\theta^*}| = qM_1 TVD(P_\theta, P_{\theta^*}) \leq qM_1 C\|\theta - \theta^*\|$ where M_1 is the diameter of B, and $TVD(P_\theta, P_{\theta^*})$ is the Total Variation Distance between the two measures and the last inequality follows from the smoothness of the parametrization of the statistical model (assumption A.1).[11] Hence $\|[E_\theta\beta_n - b(\theta)] - [E_{\theta^*}\beta_n - b(\theta^*)]\| \leq [k + qM_1C]\|\theta - \theta^*\|$ and consequently, $P_{\theta_0}(\sup_{\theta\in\Theta}\|\beta_n - E_\theta\beta_n - [b(\theta_0) - b(\theta)]\| > \varepsilon) = o(n^{-a})$, which proves the $o(n^{-a})$ consistency of the GMR2 estimator. Now, as $J_n(\theta) = (\beta_n - E_\theta\beta_n)^{/} W_n^*(\beta_n - E_\theta\beta_n)$, where $W_n^* = W^*(\theta_n^*)$, with the same logic as in i) and ii), it suffices equations (5) and (6) apply to the GMR2 and follow Lemma 5 of Andrews [2].

Now $\left.\frac{\partial J_n(\theta)}{\partial\theta}\right|_{\theta=\theta_0} = \left.-2\frac{\partial E_\theta\beta_n^{/}}{\partial\theta}W_n^*(\beta_n - E_\theta\beta_n)\right|_{\theta=\theta_0}$ and into account that, for $a > 0$ we have that $\|E_{\theta_0}\beta_n - b_0\| \leq E_{\theta_0}\|\beta_n - b_0\|$

$\leq M_1 P_{\theta_0}\left(\|\beta_n - b_0\| > C_3\frac{\ln^{1/2}n}{n^{1/2}}\right) + C_3\frac{\ln^{1/2}n}{n^{1/2}}P_{\theta_0}\left(\|\beta_n - b_0\| \leq C_3\frac{\ln^{1/2}n}{n^{1/2}}\right) = M_1 o(n^{-a}) +$

[11] Recall that a distribution Ψ is smooth *iff* for every set A, $\delta > 0$, and $A^\delta = \{x \in S : \min_{y\in A}\|x - y\| < \delta\}$, $\left|\Psi(A^\delta) - \Psi(A)\right| = o(\delta)$, A collection of distributions is called smooth if every member of it is smooth.

$C_3\frac{\ln^{1/2}n}{n^{1/2}}\left(1-o\left(n^{-a}\right)\right)=o\left(n^{-a}\right)+C_3\frac{\ln^{1/2}n}{n^{1/2}}=O\left(\frac{\ln^{1/2}n}{n^{1/2}}\right)$, we have that
$P_{\theta_0}\left(\|\beta_n-E_{\theta_0}\beta_n\|>C_1\frac{\ln^{1/2}n}{n^{1/2}}\right)$
$\leq P_{\theta_0}\left(\|\beta_n-b_0\|+\|E_{\theta_0}\beta_n-b_0\|>C_1\frac{\ln^{1/2}n}{n^{1/2}}\right)$
$\leq P_{\theta_0}\left(\|\beta_n-b_0\|+o\left(\frac{\ln^{1/2}n}{n^{1/2}}\right)>C_1\frac{\ln^{1/2}n}{n^{1/2}}\right)\leq P_{\theta_0}\left(\|\beta_n-b_0\|>C_4\frac{\ln^{1/2}n}{n^{1/2}}\right)=$ $o\left(n^{-a}\right)$ (see lemma 3.1). For $a=0$ we have that the GMR2 is asymptotically equivalent to GMR1 (Gourieroux et al. [13]). Hence for $a\geq 0$ we have that $P_{\theta_0}\left(\|\beta_n-E_{\theta_0}\beta_n\|>C_1\frac{\ln^{1/2}n}{n^{1/2}}\right)=o\left(n^{-a}\right)$. Further, due to assumption A.6 equation (1) and lemma AL.2 it follows that equation (5) holds.

Now notice that $\frac{\partial^2 J_n\left(\theta_n^+\right)}{\partial\theta\partial\theta^{/}}=2\frac{\partial E_{\theta_n^+}\beta_n^{/}}{\partial\theta}W_n^*\frac{\partial E_{\theta_n^+}\beta_n}{\partial\theta^{/}}-2\left[\frac{\partial E_{\theta_n^+}\beta_n^{/}}{\partial\theta_i\partial\theta_j}W_n^*\left(\beta_n-E_{\theta_n^+}\beta_n\right)\right]_{i,j=1,\ldots,p}$ and $2\frac{\partial E_{\theta_n^+}\beta_n^{/}}{\partial\theta}W_n^*\frac{\partial E_{\theta_n^+}\beta_n}{\partial\theta^{/}}-2\left[\frac{\partial E_{\theta_n^+}\beta_n^{/}}{\partial\theta_i\partial\theta_j}W_n^*\left(\beta_n-E_{\theta_n^+}\beta_n\right)\right]_{i,j=1,\ldots,p}\rightarrow 2\frac{\partial E_{\theta_0}\beta_n^{/}}{\partial\theta}W_0^*\frac{\partial E_{\theta_0}\beta_n}{\partial\theta^{/}}$ with probability $1-o\left(n^{-a}\right)$, where $rank\left(\frac{\partial E_{\theta_0}\beta_n^{/}}{\partial\theta}W_0^*\frac{\partial E_{\theta_0}\beta_n}{\partial\theta^{/}}\right)=p$. To prove this employ the consistency of θ_n^+, the assumption A.6 and lemma AL.2. This proves equation (6) and completes the proof. ■

Proof of Lemma 3.3. i) For GMR1 we apply lemma AL.5 where θ_n =GMR1, $\varphi_n=\left(\beta_n^{/},\left(\theta_n^*\right)^{/}\right)$ and the application is justified by the fact that provision 1 holds due to 3.1, 3.2, and A.9, 2 follows from A.3, A.5, A.7 and A.9 and 3 follows from lemma 5 of Andrews [2] and A.9. Let $S_n=\left(f_n^{/},\beta_n^{/}-b^{/}\left(\theta_0\right)\right)^{/}$, where $f_n=\frac{1}{n}\sum_{i=1}^n f\left(x_i,b\left(\theta_0\right),\theta_0\right)$ and f is defined in A.7, and $S=\left(Ef_n^{/},\mathbf{0}_{1\times q}\right)^{/}$. By remark R.11 and lemma 3.1 $\sqrt{n}\left(S_n-S\right)$ has an Edgeworth expansion of order $s=2a+1$. Hence $\pi^*\left(R_n^*\right)=G\left(S_n\right)$ where $G\left(.\right)$ smooth. and $G\left(S\right)=0$ and from Bhattacharya and Ghosh [4] $\sqrt{n}G\left(S_n\right)$ has an Edgeworth expansion of the same order.

ii) For GT the proof is analogous to (i) apart from the fact that 3.2 has to be evoked instead of 3.2. The only thing different is J_n which obeys the provisions of AL.5 additionally due to assumption A.4.

iii) For GMR2 we apply again lemma AL.5 where θ_n =GMR2, $\varphi_n=\begin{pmatrix}\beta_n\\ \theta_n^*\end{pmatrix}$ and the application is justified by the fact that that provision 1 holds due to 3.1, 3.2, and A.9, 2 follows from A.6, A.5, A.7 and A.9 and 3 follows from lemma 5 of Andrews [2] and A.9. Notice that in this case R_n^* is expanded by $\begin{pmatrix} D^1E_{\theta_0}\beta_n^{/}, & \ldots & D^{d-1}E_{\theta_0}\beta_n^{/}\end{pmatrix}^{/}$. Now define
$S_n^*=\begin{pmatrix} f_n^{/}, & \beta_n^{/}-E_{\theta_0}\beta_n^{/}, & D^1E_{\theta_0}\beta_n^{/}, & \ldots & D^{d-1}E_{\theta_0}\beta_n^{/}\end{pmatrix}^{/}$ then $\sqrt{n}\left(S_n^*-E_{\theta_0}S_n^*\right)=$

$\sqrt{n}\begin{pmatrix} f_n^{/}-E_{\theta_0}f_n^{/}, & \beta_n^{/}-E_{\theta_0}\beta_n^{/}, & 0 & \ldots & 0 \end{pmatrix}^{/}$ has an Edgeworth expansion of order $s = 2a+1$. This is justified by assumptions A.7 and A.8 for $\sqrt{n}\left(f_n^{/}-E_{\theta_0}f_n^{/}\right)^{/}$ and by lemma AL.6 for $\sqrt{n}\left(\beta_n - E_{\theta_0}\beta_n\right)$ which is valid if $\sqrt{n}\left(\beta_n - b\left(\theta_0\right)\right)$ has a valid Edgeworth expansion of order $s = 2a+2$ (due to Lemma 4.1 below and remarks R.12 and R.13). $S^* = \begin{pmatrix} S^{/} & 0 & \ldots & 0 \end{pmatrix}^{/}$ and $\pi^*\left(R_n^*\right) = G\left(S_n^*\right)$ where $G\left(S^*\right) = 0$. Hence again due to the analogous result of Bhattacharya and Ghosh [4] $\sqrt{n}G\left(S_n^*\right)$ has an Edgeworth expansion of the same order. ■

Proof of Lemma 4.1. Assume now that $\sup_{A\in\mathcal{B}_C}\left|P_n(A) - Q_n(A)\right| = O\left(n^{-a-\eta}\right)$, where $\mathcal{B}_C$ denote the collection of convex Borel sets of $\mathbb{R}^q$ and $\eta > 0$. Now $n^a\left\|\int_{\mathbb{R}^q} x\left(dP_n - dQ_n\right)\right\|$

$\leq n^a\left\|\int_{B(0,K(\ln n)^\epsilon)} x\left(dP_n - dQ_n\right)\right\| + n^a\left\|\int_{\mathbb{R}^q\backslash B(0,K(\ln n)^\epsilon)} xdP_n\right\| + n^a\left\|\int_{\mathbb{R}^q\backslash B(0,K(\ln n)^\epsilon)} xdQ_n\right\|$

$\leq n^a K\left(\ln n\right)^\epsilon \int_{B(0,K(\ln n)^\epsilon)}\left|dP_n - dQ_n\right| + n^a\int_{\mathbb{R}^q\backslash B(0,K(\ln n)^\epsilon)}\|x\|\left(dP_n + \left|dQ_n\right|\right)$

$\leq K\left(\ln n\right)^\epsilon \sup_{A\in\mathcal{B}_C} n^a\left|P_n(A) - Q_n(A)\right| + n^a\int_{\mathbb{R}^q\backslash B(0,K(\ln n)^\epsilon)}\|x\|\left(dP_n + \left|dQ_n\right|\right)$

Let P_n be the distribution of $\sqrt{n}\left(\theta_n - \theta_0\right)$. Then $n^a\int_{\mathbb{R}^q\backslash B(0,K(\ln n)^\epsilon)}\|x\|\,dP_n =$ $n^a\int_{[\mathbb{R}^q\backslash B(0,K(\ln n)^\epsilon)]\cap\sqrt{n}(\Theta-\theta_0)}\|x\|\,dP_n + n^a\int_{[\mathbb{R}^q\backslash B(0,K(\ln n)^\epsilon)]\cap\left(\sqrt{n}(\Theta-\theta_0)\right)^C}\|x\|\,dP_n =$

$= n^a\int_{[\mathbb{R}^q\backslash B(0,K(\ln n)^\epsilon)]\cap\sqrt{n}(\Theta-\theta_0)}\|x\|\,dP_n$ as the support of P_n is $\sqrt{n}\left(\Theta - \theta_0\right)$. $n^a\int_{[\mathbb{R}^q\backslash B(0,K(\ln n)^\epsilon)]\cap\sqrt{n}(\Theta-\theta_0)}\|x\|\,dP_n = n^a\int_{\sqrt{n}(\Theta-\theta_0)\backslash B(0,K(\ln n)^\epsilon)}\|x\|\,dP_n$ for n large enough.

Hence $n^a\int_{\mathbb{R}^q\backslash B(0,K(\ln n)^\epsilon)}\|x\|\,dP_n \leq n^{a+\frac{1}{2}}\rho\int_{\sqrt{n}(\Theta-\theta_0)\backslash B(0,K(\ln n)^\epsilon)} dP_n$ where ρ is such that $\mathcal{O}_\rho(0) \supseteq \Theta - \theta_0$ and ρ exists as Θ is bounded by assumption. Hence $n^a\left\|\int_{\mathbb{R}^q} x\left(dP_n - dQ_n\right)\right\| \leq K\left(\ln n\right)^\epsilon \sup_{A\in\mathcal{B}_C} n^a\left|P_n(A) - Q_n(A)\right|$ $+ n^{a+\frac{1}{2}}\rho P\left(\sqrt{n}\left(\theta_n - \theta_0\right) \in \sqrt{n}\left(\Theta - \theta_0\right)\backslash B\left(0, K\left(\ln n\right)^\epsilon\right)\right) + n^a\int_{\mathbb{R}^q\backslash B(0,K(\ln n)^\epsilon)}\|x\|\left|dQ_n\right|$.

As $\sup_{A\in\mathcal{B}_C} n^a\left|P_n(A) - Q_n(A)\right| = O\left(n^{-\eta}\right)$ for $\eta > 0$, we have that $K\left(\ln n\right)^\epsilon \sup_{A\in\mathcal{B}_C} n^a\left|P_n(A) - Q_n(A)\right| = o(1)$ and the result follows due to assumptions A.10 and A.11 above. ■

Proof of Lemma 4.2. To conserve space we shall only sketch the proof (for an analytic one see Arvanitis and Demos [3]). As $\beta_n = \arg\min_{\beta\in B} c_n^{/}(\beta) W_n c_n(\beta)$, where $W_n = W_n\left(\beta_n^*\right)$, it follows that $\frac{\partial}{\partial\beta}c_n^{/}\left(\beta_n\right) W_n\sqrt{n}c_n\left(\beta_n\right) = \mathbf{0}_q$. Expanding $c_n\left(\beta_n\right)$ and W_n around b_0 and keeping the relevant terms we get and expression for $\sqrt{n}\left(\beta_n - b_0\right)$. Employing now the moment approximations for the analogous terms of $\frac{1}{n}\sum_{i=1}^n\left(f\left(x_i, b_0, \theta_0\right) - E\left(f\left(x_i, b_0, \theta_0\right)\right)\right)$, due to remark R.11 and lemma 4.1 and holding terms up to the relevant order, $\frac{\partial c_n^{/}(b_0)}{\partial\beta} = c_\beta^{/}\left(b_0\right) + \frac{1}{\sqrt{n}}c_\beta^{/}\left(z, b_0\right)$, $W_n\left(b_0\right) = W_0 + \frac{1}{\sqrt{n}}w\left(z, b_0\right)$, and $\frac{1}{\sqrt{n}}\sum c\left(x_i, b_0\right) =$ $c\left(z, b_0\right) + \frac{1}{\sqrt{n}}c^*\left(z, b_0\right)$ where $z \sim N\left(0, \Sigma\right)$, the elements of $c_\beta\left(z, b_0\right)$, $w\left(z, b_0\right)$ and $c\left(z, b_0\right)$ are finite polynomials in z with $O(1)$ coefficients and $E_{\theta_0}c\left(z, b_0\right) =$

$E\left(c^{*}\left(z,b_0\right)\right)=0$, we get that $\left[\frac{\partial}{\partial\beta}\frac{1}{n}\sum c'\left(x_i,b_0\right)\frac{1}{n}\sum W\left(x_i,b_0\right)\frac{\partial}{\partial\beta^{/}}\frac{1}{n}\sum c\left(x_i,b_0\right)\right]^{-1}=$ $Q^{-1}-\frac{1}{\sqrt{n}}Q^{-1}AQ^{-1}+o\left(n^{-\frac{1}{2}}\right)$ due to lemma AL.4 where Q, and A are given in the declaration of the lemma (see Corollary 1 Magdalinos [17], as well). Further, with an $o\left(n^{-\frac{1}{2}}\right)$ error we have

$\frac{\partial c_n^{/}(b(\theta_0))}{\partial\beta}W_n\left(b_0\right)=c_\beta^{/}\left(b_0\right)W_0+\frac{1}{\sqrt{n}}\left[c_\beta^{/}\left(b_0\right)w\left(z,b_0\right)+c_\beta^{/}\left(z,b_0\right)W_0\right]$,

$\frac{\partial^2}{\partial\beta\partial\beta^{/}}\frac{1}{n}\sum c_j\left(x_i,b_0\right)=c_{\beta,\beta'}\left(b_0\right)+\frac{1}{\sqrt{n}}c_{\beta,\beta'}\left(z,b_0\right)$ and $\frac{\partial}{\partial\beta^{/}}\frac{1}{n}\sum W_{rj}\left(x_i,b_0\right)=W_{\beta^{/}}\left(b_0\right)_{rj}+\frac{1}{\sqrt{n}}w_{\beta^{/}}\left(z,b_0\right)_{rj}$ due to remark R.11 and lemma 4.1.

Hence, employing $\sqrt{n}\left(\beta_n^{*}-b_0\right)\underset{1/2}{\sim}k_1^{*}+\frac{k_2^{*}}{\sqrt{n}}$ (see assumption A.9) and collecting terms we get the result. ■

Proof of Lemma 4.3. Utilizing assumption A.12 we have that θ_n satisfies the following equation:

$\frac{\partial b^{/}(\theta_n)}{\partial\theta}W_n^{*}\left(\theta_n^{*}\right)\sqrt{n}\left(\beta_n-b\left(\theta_n\right)\right)=\mathbf{0}_p$. Expanding $\frac{\partial b^{/}(\theta_n)}{\partial\theta}$, $W_n^{*}\left(\theta_n^{*}\right)$ and $\sqrt{n}\left(\beta_n-b\left(\theta_n\right)\right)$ around θ_0 keeping the relevant terms we can solve for $\sqrt{n}\left(\theta_n-\theta_0\right)$. Employing now the moment approximations for the analogous terms of $\frac{1}{n}\sum_{i=1}^{n}\left(f\left(x_i,b_0,\theta_0\right)-E\left(f\left(x_i,b_0,\theta_0\right)\right)\right)$, due to remark R.11, lemma 4.1 and due to lemma AL.4 we get:

$$\left(\frac{\partial b_0^{/}}{\partial\theta}\frac{1}{n}\sum W^{*}\left(x_i,\theta_0\right)\frac{\partial b_0}{\partial\theta^{/}}\right)^{-1}=\Gamma-\frac{1}{\sqrt{n}}\Gamma\frac{\partial b_0^{/}}{\partial\theta}w^{*}\left(z,\theta_0\right)\frac{\partial b_0}{\partial\theta^{/}}\Gamma$$

where Γ is given in the declaration of the lemma. Further

$$\left(\frac{\partial b_0^{/}}{\partial\theta}\frac{1}{n}\sum W^{*}\left(x_i,\theta_0\right)\frac{\partial b_0}{\partial\theta^{/}}\right)^{-1}\frac{\partial b_0^{/}}{\partial\theta}\frac{1}{n}\sum W^{*}\left(x_i,\theta_0\right)$$

$$=\Gamma\frac{\partial b_0^{/}}{\partial\theta}W_0^{*}+\frac{1}{\sqrt{n}}\Gamma\frac{\partial b_0^{/}}{\partial\theta}w^{*}\left(z,\theta_0\right)-\frac{1}{\sqrt{n}}\Gamma\frac{\partial b_0^{/}}{\partial\theta}w^{*}\left(z,\theta_0\right)\frac{\partial b_0}{\partial\theta^{/}}\Gamma\frac{\partial b_0^{/}}{\partial\theta}W_0^{*}$$

It follows that substituting in the expression of $\sqrt{n}\left(\theta_n-\theta_0\right)$ the above terms and collecting terms we get the result. ■

Proff of Lemma 4.5. Employing again the procedure as in the relevant proofs before and utilizing assumption A.12, remark R.11, lemmas 4.1, AL.4 and 4.4 we have that θ_n satisfies the condition: $\frac{\partial E_{\theta_n}\left(\beta_n^{/}\right)}{\partial\theta}W_n^{*}\left(\theta_n^{*}\right)\sqrt{n}\left(\beta_n-E_{\theta_n}\beta_n\right)=\mathbf{0}_p$. Expanding $\frac{\partial E_{\theta_n}\left(\beta_n^{/}\right)}{\partial\theta}$, $W_n^{*}\left(\theta_n^{*}\right)$ and $\sqrt{n}\left(\beta_n-E_{\theta_n}\beta_n\right)$ around θ_0 and $b\left(\theta_0\right)$, and keeping the relevant terms we can solve for $\sqrt{n}\left(\theta_n-\theta_0\right)$. Now taking into account that first, $\frac{\partial E_{\theta_0}\beta_n^{/}}{\partial\theta}=\frac{\partial b_0^{/}}{\partial\theta}+o\left(1\right)=\frac{\partial b_0^{/}}{\partial\theta}+B_n$ with $\left\|B_n\right\|=o\left(1\right)$, second, $W_n^{*}\left(\theta_0\right)=W_0^{*}+\frac{1}{\sqrt{n}}w^{*}\left(z,\theta_0\right)$, where $z\sim N\left(0,\Sigma\right)$ and the elements of $w^{*}\left(z,\theta_0\right)$ are $O\left(1\right)$ finite polynomials in z, and third,

$$\left[\left(\frac{\partial b_0^{/}}{\partial\theta}+B_n\right)\left(W_0^{*}+\frac{1}{\sqrt{n}}w^{*}\left(z,\theta_0\right)\right)\left(\frac{\partial b_0}{\partial\theta^{/}}+B_n^{/}\right)\right]^{-1}$$

$= \Gamma + K_n - \frac{1}{\sqrt{n}} \Gamma w^* (z, \theta_0) \Gamma$ where $\|K_n\| = o(1)$, $\Gamma = \left(\frac{\partial b_0^/}{\partial \theta} W_0^* \frac{\partial b_0}{\partial \theta^/} \right)^{-1}$. Noticing now that $\left\| \left[\Gamma B_n + K_n \left(\frac{\partial b_0^/}{\partial \theta} + B_n \right) \right] W_0^* \right\| = o(1)$ we get the result. ■

Proof of Lemma 4.6. Let $s_n(\theta)$ and $H_n(\theta)$ denote the gradient (score) and the Hessian of the loglikelihood function of $\mathcal{D}$ respectively. First, notice that for all θ in $\mathcal{O}_{\varepsilon_2}(\theta_0)$ the identity $E_\theta c_n(b(\theta)) = \mathbf{0}_l$ is well defined by assumptions A.3, A.4. Due to assumption A.13 and taking derivatives with respect to $\theta^/$ $\frac{\partial E_\theta c_n(b(\theta))}{\partial \theta^/} = \mathbf{0}_{l \times p}$ and it follows that

$$E_\theta \frac{\partial c_n(b(\theta))}{\partial \beta^/} \frac{\partial b(\theta)}{\partial \theta^/} = -E_\theta c_n(b(\theta)) s_n^/(\theta). \tag{10}$$

Also, since $\frac{\partial E_\theta c_n(\beta)}{\partial \beta^/} = E_\theta \frac{\partial c_n(\beta)}{\partial \beta^/}$ (assumption A.13), we have

$$E_\theta \frac{\partial^2 c_n(b(\theta))_j}{\partial \beta \partial \theta^/} \frac{\partial b(\theta)}{\partial \theta^/} = E_\theta \frac{\partial^2 c_n(b(\theta))_j}{\partial \beta \partial \beta^/} \frac{\partial b(\theta)}{\partial \theta^/} = -E_\theta \frac{\partial c_n(b(\theta))_j}{\partial \beta} s_n^/(\theta). \tag{11}$$

Then $\frac{\partial^2 E_\theta c_n(b(\theta))_j}{\partial \theta \partial \theta^/} = E_\theta \frac{\partial^2 c_n(b(\theta))_j}{\partial \theta \partial \theta'} + E_\theta \frac{\partial c_n(b(\theta))_j}{\partial \theta} s_n^/(\theta) + E_\theta s_n(\theta) \frac{\partial c_n(b(\theta))_j}{\partial \theta^/} + E_\theta c_n(b(\theta))_j H_n(\theta) + E_\theta c_n(b(\theta))_j s_n(\theta) s_n^/(\theta)$

$= -\frac{\partial \beta^/(\theta)}{\partial \theta} \left(E_\theta \frac{\partial^2 c_n(b(\theta))_j}{\partial \beta \partial \beta^/} \right) \frac{\partial b(\theta)}{\partial \theta^/} + \left[E_\theta \frac{\partial c_n(b(\theta))_j}{\partial \beta^/} \frac{\partial^2 \beta(\theta)}{\partial \theta \partial \theta_r} \right]_{r=1,\ldots,p} + E_\theta c_n(b(\theta))_j H_n(\theta) +$ $E_\theta c_n(b(\theta))_j s_n(\theta) s_n^/(\theta)$ by equation (11) and since $\frac{\partial^2 E_\theta c_n(b(\theta))_j}{\partial \theta \partial \theta^/} = \mathbf{0}_{p \times p}$ we get

$$\begin{aligned} & \frac{\partial \beta^/(\theta)}{\partial \theta} \left(E_\theta \frac{\partial^2 c_n(b(\theta))_j}{\partial \beta \partial \beta^/} \right) \frac{\partial \beta(\theta)}{\partial \theta^/} - \left\{ E_\theta \frac{\partial c_n(b(\theta))_j}{\partial \beta'} \frac{\partial^2 b(\theta)}{\partial \theta \partial \theta_r} \right\}_{r=1,\ldots,p} \\ = \; & E_\theta c_n(b(\theta))_j H_n(\theta) + E_\theta c_n(b(\theta))_j s_n(\theta) s_n^/(\theta). \end{aligned} \tag{12}$$

Now utilizing assumption A.12, remark R.11, lemmas 4.1, AL.4, equations (10), (11), (12), expanding $\sqrt{n} E_\theta(c_n(\beta))$ around θ_0 and b_0, taking into account that $W_n^{**}(\theta_n^*) = W_n^{**}(\theta_0) + \frac{1}{\sqrt{n}} \left[\frac{\partial}{\partial \theta^/} W_n^{**}(\theta_0)_{rj} \sqrt{n}(\theta_n^* - \theta_0) \right]_{r,j=1,\ldots,l}$ and $\sqrt{n}(\beta_n - b_0) \underset{1/2}{\sim} k_1 + \frac{1}{\sqrt{n}} k_2$ we get an expression which involves the quantities $s_n(\theta_0)$, $H_n(\theta_0)$, and $W_n^{**}(\theta_0)$. Employing equations (11), (12) and $W_n^{**}(\theta_0) = W^{**}(\theta_0) + \frac{1}{\sqrt{n}} w^{**}(z, \theta_0)$, where $z \sim N(0, \Sigma)$, we are able to solve for $\sqrt{n}(\theta_n - \theta_0)$ and the result follows. ■

Proof of Corollary 5. By direct substitution we obtain that $q_1 = \left(\frac{\partial b(\theta_0)}{\partial \theta'} \right)^{-1} k_1$, and $q_2 =$

$\left(\frac{\partial b(\theta_0)}{\partial \theta'} \right)^{-1} k_2 - \frac{1}{2} \left(\frac{\partial b(\theta_0)}{\partial \theta'} \right)^{-1} \left(E_{\theta_0} \frac{\partial c_n(b(\theta_0))}{\partial \beta'} \right)^{-1} \left[tr q_1 q_1^{/\cdot} \left[E_{\theta_0} \frac{\partial c_n(b(\theta_0))_j}{\partial \beta'} \frac{\partial^2 b(\theta_0)}{\partial \theta_r \partial \theta'} \right]_{r=1,\ldots p} \right]_{j=1,\ldots,l}$

$= \left(\frac{\partial b(\theta_0)}{\partial \theta'}\right)^{-1} k_2 - \frac{1}{2}\left(\frac{\partial b(\theta_0)}{\partial \theta'}\right)^{-1}\left(E_{\theta_0}\frac{\partial c_n(b(\theta_0))}{\partial \beta'}\right)^{-1}\left[\sum_r E_{\theta_0}\frac{\partial c_n(b(\theta_0))_j}{\partial \beta'} trq_1 q_1^{/}\left[\frac{\partial^2 b_r(\theta_0)}{\partial\theta\partial\theta'}\right]_{r=1,...p}\right]_{j=1,...,l},$

and by lemma AL.3 we have $q_2 = \left(\frac{\partial b(\theta_0)}{\partial \theta'}\right)^{-1} k_2 - \frac{1}{2\sqrt{n}}\left(\frac{\partial b(\theta_0)}{\partial \theta'}\right)^{-1}\left[trq_1 q_1^{/}\frac{\partial^2 b_j(\theta_0)}{\partial\theta\partial\theta'}\right]_{j=1,...,l}.$

■

Proof of Proposition 8. We prove the proposition by induction (detailed proof can be found in Arvanitis and Demos [3]). Let the i^{th} element of a vector x be denoted by x_i. Then assume that, in the assumed neighborhood of θ_0, the proposition is true for $r = h$, i.e. assume that, for $i = 1, ..p$ we have:

$\sqrt{n}\left(\theta_n^{(h} - \theta_0\right)_i = (k_1)_i + \frac{1}{n^{\frac{1}{2}}}(k_2)_i + ... + \frac{1}{n^{\frac{2h+3}{2}}}(k_{2h+4})_i + o\left(n^{-\frac{2h+3}{2}}\right),$

with

$$E_\theta (k_1)_i = E_\theta (k_2)_i = .. = E_\theta (k_{2h+1})_i = 0, \qquad (13)$$

i.e. $\theta_n^{(h}$ is $O\left(n^{-\frac{2h+1}{2}}\right)\ \theta_0 - unbiased$. Now for any $\theta \in \mathcal{O}_\varepsilon(\theta_0)$, and any $i, j, m, l, r = 1, ...p$, we have that

$\left(E_\theta \theta_n^{(h}\right)_i = \theta_i + \frac{1}{n^{\frac{2h+2}{2}}}E_\theta (k_{2h+2})_i + \frac{1}{n^{\frac{2h+3}{2}}}E_\theta (k_{2h+3})_i + \frac{1}{n^{\frac{2h+4}{2}}}E_\theta (k_{2h+4})_i + o\left(n^{-\frac{2h+4}{2}}\right),$

$\frac{\partial\left(E_\theta\theta_n^{(h}\right)_i}{\partial\theta_j} = \delta_{ij} + \frac{1}{n^{\frac{2h+2}{2}}}\frac{\partial E_\theta (k_{2h+2})_i}{\partial\theta_j} + \frac{1}{n^{\frac{2h+3}{2}}}\frac{\partial E_\theta (k_{2h+3})_i}{\partial\theta_j} + \frac{1}{n^{\frac{2h+4}{2}}}\frac{\partial E_\theta (k_{2h+4})_i}{\partial\theta_j} + o\left(n^{-\frac{2h+4}{2}}\right)$

and all higher order derivatives are of $O\left(n^{-\frac{2h+2}{2}}\right)$ (the same applies for $\frac{\partial\left(E_\theta\theta_n^{(h}\right)_i}{\partial\theta_j}$, for $i \neq j$).

Then the $h+1^{st}-step$ GMR2 estimator is defined as $\theta_n^{(h+1} = \arg\min_\theta \left(\theta_n^{(h} - E_\theta\theta_n^{(h}\right)^2$. Hence we have that $\theta_n^{(h} - E_{\theta_n^{(h+1}}\theta_n^{(h} = 0$. Taking into account the previous equation, expanding $E_{\theta_n^{(h+1}}\theta_n^{(h}$ around θ_0, and noticing that $\left(\frac{\partial\left(E_\theta\theta_n^{(h}\right)_i}{\partial\theta_i}\right)^{-1} = 1 - \frac{1}{n^{\frac{2h+2}{2}}}\frac{\partial E(k_{h+2})_i}{\partial\theta_j} - \frac{1}{n^{\frac{2h+3}{2}}}\frac{\partial E_\theta(k_{h+3})_i}{\partial\theta_j} - \frac{1}{n^{\frac{2h+4}{2}}}\frac{\partial E_\theta(k_{h+4})_i}{\partial\theta_j} + o\left(n^{-\frac{2h+4}{2}}\right)$ we get $\sqrt{n}\left(\theta_n^{(h+1} - \theta_0\right)_i$ as a function of terms that have zero expectation (due to equation (13)), terms that involve the product $(k_1)_i (k_1)_j$ and have non-zero expectation, and an error of $o\left(n^{-\frac{2h+3}{2}}\right)$. It follows that

$E_{\theta_0}\left(\theta_n^{(h+1}\right)_i = (\theta_0)_i - \frac{1}{2}\frac{1}{n^{\frac{2h+4}{2}}}\sum_{j=1}^{p}\sum_{m=1}^{p}\frac{\partial^2 E_{\theta_0}(k_{2h+2})_i}{\partial\theta_m\partial\theta_j}E_{\theta_0}\left[(k_1)_j (k_1)_m\right] + o\left(n^{-\frac{2h+4}{2}}\right),$

which establishes the proposition due to the fact that i is arbitrary. ■

Proof of Lemma 2.2. When $p = q = l$ due to consistency, the GT estimator satisfies with probability $1 - o(n^{-a})$ $E_{\theta_n} c_n(\beta_n) = \mathbf{0}_p$. Yet from assumption A.4 we have that $E_{\theta_n} c_n(\beta) = \mathbf{0}_p$ iff $\beta = b(\theta_n)$. Hence the estimator equivalently satisfies $\beta_n - b(\theta_n) = \mathbf{0}_p$, which defines the GMR1 estimator in these special circumstances. ■

Proof of Lemma 2.3. In the first case we have that $\beta_n = g^{-1} \circ \frac{1}{n} f(x_i)$, $b(\theta) = g^{-1} \circ E_\theta f(x_i) = g^{-1} \circ m(\theta)$, GMR1$= m^{-1} \circ g \circ \beta_n = m^{-1} \circ \frac{1}{n} f(x_i)$. For the second case, if g is linear then $E_\theta \beta_n = g^{-1} \circ E_\theta \frac{1}{n} f(x_i) = g^{-1} \circ m(\theta) = b(\theta)$, and the result follows. ■

B Proofs of General Lemmas

The following are a collection of helpful lemmas that are frequently referenced in the proofs of the main results. The first concerns weighting matrices and initial estimators in general, hence it is directly connected to assumptions A.5 and A.7.

Lemma AL.1 *Suppose that* $W_n(\omega, \theta_n^*)$, $W(\theta_0)$, θ_n^* *are defined as in assumptions A.5 and A.9, then*

$$P_\Omega\left(\|W_n(\omega, \theta_n^*) - W(\theta_0)\| > \varepsilon\right) = o\left(n^{-a}\right), \quad \forall \varepsilon > 0$$

and

$$P_\Omega\left(\left\|D^r W_n(\omega, \theta)|_{\theta=\theta_n^*} - D^r W(\theta)|_{\theta=\theta_0}\right\| > \varepsilon\right) = o\left(n^{-a}\right), \ \forall \varepsilon > 0, \ \textit{and} \ r < d+1.$$

Proof. Under assumptions A.5 and A.9, Lemmas 3 and 5 of [2], and due the triangle inequality we have that $P_\Omega\left(\|W_n(\omega, \theta_n^*) - W(\theta_0)\| > \varepsilon\right)$
$\leq P_\Omega\left(\|W_n(\omega, \theta_0) - W(\theta_0)\| > \frac{\varepsilon}{2}\right) + P_\Omega\left(\|W_n(\omega, \theta_n^*) - W_n(\omega, \theta_0)\| > \frac{\varepsilon}{2}\right)$
$\leq o(n^{-a}) + P_\Omega\left(u_n\|\theta_n^* - \theta_0\| > \frac{\varepsilon}{2}\right) = o(n^{-a})$ and
$P_\Omega\left(\left\|D^r W_n(\omega, \theta)|_{\theta=\theta_n^*} - D^r W(\theta)|_{\theta=\theta_0}\right\| > \varepsilon\right)$
$\leq P_\Omega\left(\left\|D^r W_n(\omega, \theta)|_{\theta=\theta_n^*} - D^r W_n(\omega, \theta)|_{\theta=\theta_0}\right\| > \frac{\varepsilon}{2}\right)$
$+ P_\Omega\left(\left\|D^r W_n(\omega, \theta)|_{\theta=\theta_0} - D^r W(\theta)|_{\theta=\theta_0}\right\| > \frac{\varepsilon}{2}\right)$
$\leq P_\Omega\left(u_{Dn}\|\theta_n^* - \theta_0\| > \frac{\varepsilon}{2}\right) + o(n^{-a}) = o(n^{-a})$. ■

In the following lemma $\omega(n)$ is a deterministic function of n. In most cases $\omega(n) = \frac{\ln^{\frac{1}{2}} n}{n^{\frac{1}{2}}}$, but $\omega(n) = 1$ is also allowed.

Lemma AL.2 *Let* $P_\Omega(\|a_n\| > K) = o(n^{-a})$, *for some* $K > 0$ *and* $P_\Omega(\|\beta_n\| > C\omega(n)) = o(n^{-a})$, *for some* $C > 0$. *Then*

$$P_\Omega(\|a_n\beta_n\| > C^*\omega(n)) = o\left(n^{-a}\right), \textit{ for some } C^* > 0$$

Proof. By the submultiplicativity property $P_\Omega(\|a_n\beta_n\| > C^*\omega(n))$
$\leq P_\Omega(\{\|a_n\|\|\beta_n\| > C^*\omega(n)\} \cap \{\|a_n\| > K\})$
$+ P_\Omega(\{\|a_n\|\|\beta_n\| > C^*\omega(n)\} \cap \{\|a_n\| \leq K\})$
$\leq P_\Omega(\|a_n\| > K) + P_\Omega(K\|\beta_n\| > C^*\omega(n)) = o(n^{-a})$ for $C^* \geq CK$. ■

Lemma AL.3 *For $A, M \in \mathcal{M}^{q\times q}$, and M invertible*

$$\left\{tr\left(A\frac{\partial^2 b(\theta_0)_i}{\partial\theta\partial\theta'}\right)\right\}_{i=1,\ldots,q} = M^{-1}\left\{\sum_{j=1}^{q} M_{i,j}tr\left(A\frac{\partial^2 b(\theta_0)_j}{\partial\theta\partial\theta'}\right)\right\}_{i=1,\ldots,q}$$

Proof. Let $u = \left[tr\left(A\frac{\partial^2 b_i(\theta_0)}{\partial\theta\partial\theta'}\right)\right]_{i=1,\ldots,q}$, and $v = M^{-1}\left[tr\left(A\frac{\partial^2 b_j(\theta_0)}{\partial\theta\partial\theta'}\right)\right]_{i=1,\ldots,q}$. Then $v_i = \sum_{j=1}^{q}\sum_{m=1}^{q} M^{i,m}M_{m,j}tr\left(A\frac{\partial^2 b_j(\theta_0)}{\partial\theta\partial\theta'}\right) = \sum_{j=1}^{q}\delta_{i,j}tr\left(A\frac{\partial^2 b_j(\theta_0)}{\partial\theta\partial\theta'}\right) = tr\left(A\frac{\partial^2 b_i(\theta_0)}{\partial\theta\partial\theta'}\right) = u_i$, $\forall i = 1,\ldots q$. ■

Lemma AL.4 *Let X and $Y_i(z)$ be square matrices, with X being non-singular and $Y_i(z)$ has elements of finite degree polynomials in z, and $z \sim N(0,\Sigma)$. Then*

$$\left(X + \sum_{i=1}^{2a}\frac{1}{n^{\frac{i}{2}}}Y_i(z)\right)^{-1} = X^{-1} + \sum_{i=1}^{2a}\frac{1}{n^{\frac{i}{2}}}K_i(z) + R_n(z)$$

where $R_n(z)$ is such that $P(\|R_n(z)\| > \gamma_n) = o(n^{-a})$ and $\gamma_n = o(n^{-a})$.

Proof. For $n \geq n^*$ we have that $\|R_n(z)\| \leq \frac{1}{n^{a+\frac{1}{2}}}\|R(z)\|$ where the elements of $R(z)$ are finite polynomials of z. Then it suffices to find $c > 0$ and $\varepsilon > 0$ such that $n^a P(\|R_n(z)\| > cn^{-a-\varepsilon}) = o(1)$. However, notice that $n^a P(\|R_n(z)\| > cn^{-a-\varepsilon}) \leq n^a P\left(\frac{1}{n^{a+\frac{1}{2}}}\|R(z)\| > cn^{-a-\varepsilon}\right) \leq n^{a-\frac{k}{2}+k\varepsilon}m$ where $\frac{E\|R(z)\|^k}{c^k} = m$ and any $k \in \mathbb{N}$, due to the Markov inequality and the normality of z. Hence we need $a - \frac{k}{2} + k\varepsilon < 0 \Rightarrow \varepsilon < \frac{1}{2} - \frac{a}{k}$ and $\varepsilon > 0$. This is satisfied for any $k > 2a$. ■

Let us denote as θ_n any of the examined (auxiliary or indirect) estimators. We denote with φ_n either β_n^* or $\begin{pmatrix}\beta_n\\ \theta_n^*\end{pmatrix}$ as these are defined in the section concerning the definition of the examined estimators. We finally denote with J_n any of the criteria that are involved in the aforementioned definitions, i.e. $J_n(\theta,\varphi) = Q_n^{/}(\theta,\varphi)W(\varphi)Q_n(\theta,\varphi)$, and J is its probability limit. Our next lemma concerns the derivation of the validity of the Edgeworth expansion in all examined cases. It essentially determines that the local approximation of $\sqrt{n}(\theta_n - \theta_0)$ obtained by the inversion of a polynomial approximation of the first order conditions, has an error that is not greater that any $o(n^{-a})$-real sequence with probability $1-o(n^{-a})$. This result, along with the provisions of corollary AC.1 that follows, establish that these two sequences have the same Edgeworth expansions if any one of them has a valid Edgeworth expansion.

Lemma AL.5 *If 1.* $P_{\theta_0}\left(\left\|n^{\frac{1}{2}}(\theta_n-\theta_0)\right\|>C_\theta \ln^{1/2} n\right)=o\left(n^{-a}\right)$*,* $P_{\theta_0}\left(\left\|n^{\frac{1}{2}}(\varphi_n-\varphi_0)\right\|>C_\varphi \ln^{1/2} n\right)=o\left(n^{-a}\right)$ *and* $P_{\theta_0}\left(\left\|n^{\frac{1}{2}}Q_n(\theta_0,\varphi_0)\right\|>C_Q \ln^{1/2} n\right)$ *for* $C_\theta, C_\varphi, C_Q>0$*,*

2. $\frac{\partial J_n(\theta,\varphi)}{\partial\theta}$ *is differentiable of order* $d-1$ *in a neighborhood of* (θ_0,φ_0) *and the* $d-1$ *order derivative is Lipschitz in this neighborhood (or in a subset of it) the Lipschitz coefficient is bounded with probability* $1-o\left(n^{-a}\right)$*, and* $\frac{\partial J^2(\theta_0,\varphi_0)}{\partial\theta\partial\theta'}$ *is positive definite,*

3. $P_{\theta_0}\left(\left\|n^{\frac{1}{2}}(\varphi_n-\varphi_0)-n^{\frac{1}{2}}\pi(R_n)\right\|>\omega_n^*\right)=o\left(n^{-a}\right)$ *with* π*,*R_n*, and* ω_n^* *analogous to the relevant quantities of the present lemma (see below) that are derived in an analogous manner with a potentially different* J_n*,*

then there exists a smooth function $\pi^*:\mathbb{R}^m\to\mathbb{R}^p$*, that is independent of* n *such that*

$$P_{\theta_0}\left(\left\|n^{\frac{1}{2}}(\theta_n-\theta_0)-n^{\frac{1}{2}}\pi^*(R_n^*)\right\|>\omega_n\right)=o\left(n^{-a}\right)$$

where R_n^* *is the sequence of random elements with values on* $\mathbb{R}^m$*, with components the distinct components of* $\frac{\partial J_n(\theta_0,\varphi_0)}{\partial\theta}$*, and* $\left\{D^{j_1,j_2}\left(\frac{\partial J_n(\theta,\varphi)}{\partial\theta}\right)|_{(\theta=\theta_0,\varphi=\varphi_0)}\right\}_{j_1+j_2=i,i=1,..,d-1}$*, where* $D^{j_1,j_2}\left(\frac{\partial J_n(\theta,\varphi)}{\partial\theta}\right)=D_\varphi^{j_2}\circ D_\theta^{j_1}\left(\frac{\partial J_n(\theta,\varphi)}{\partial\theta}\right)$*,* $m=\dim(R_n^*)$ *and* $\omega_n=o\left(n^{-a}\right)$ *deterministic.*

Proof. Denoting $\overline{\theta_n}=\theta_n-\theta_0$, $\overline{\varphi_n}=\varphi_n-\varphi_0$, a $(d-1)$-Taylor expansion about (θ_0,φ_0) on the conditions $\frac{\partial J_n(\theta_n,\varphi_n)}{\partial\theta}=\mathbf{0}_p$ would obtain

$\frac{\partial J_n(\theta_0,\varphi_0)}{\partial\theta}+\sum_{\substack{i=1\\ j_1+j_2=i\\ j_1,j_2\geq 0}}^{d-1}\frac{1}{i!}\binom{i}{j_1}D^{j_1,j_2}\left(\frac{\partial J_n((\theta_0,\varphi_0))}{\partial\theta}\right)\begin{pmatrix}\overline{\theta_n}^{j_1}\\ \overline{\varphi_n}^{j_2}\end{pmatrix}+r_n^*=\mathbf{0}_p$ where $\begin{pmatrix}\underbrace{\overline{\theta_n},\cdots,\overline{\theta_n}}_{j_1\text{ times}}\\ \underbrace{\overline{\varphi_n},\cdots,\overline{\varphi_n}}_{j_2\text{ times}}\end{pmatrix}=\begin{pmatrix}\overline{\theta_n}^{j_1}\\ \overline{\varphi_n}^{j_2}\end{pmatrix}$ and the remainder is

$r_n^*=\sum_{j_1=0}^{d-1}\frac{1}{(d-1)!}\binom{d-1}{j_1}D^{j_1,d-1-j_1}\left(\frac{\partial J_n\left(\theta_n^+,\varphi_n^+\right)}{\partial\theta}-\frac{\partial J_n((\theta_0,\varphi_0))}{\partial\theta}\right)\begin{pmatrix}\overline{\theta_n}^{j_1}\\ \overline{\varphi_n}^{j_2}\end{pmatrix}$, where each $D^{i_1,i_2}\left(\frac{\partial J_n(\theta_0,\varphi_0)}{\partial\theta}\right)$ is an i_1+i_2 tensor defined on $\underbrace{\mathbb{R}^p\otimes\cdots\otimes\mathbb{R}^p}_{i_1\text{ times}}\otimes\underbrace{\mathbb{R}^q\otimes\cdots\otimes\mathbb{R}^q}_{i_2\text{ times}}$ with values in $\mathbb{R}^p$, with coefficients the i_1^{th} derivatives of $\frac{\partial J_n(\theta,\theta)}{\partial\theta}$ with respect to the components of the initial θ and the i_2^{th} derivatives of $\frac{\partial J_n(\theta,\varphi)}{\partial\theta}$ with respect to the components of the final φ at (θ_0,φ_0), and θ_n^+, φ_n^+ between θ_n and θ_0, and φ_n and φ_0, respectively. Hence the previous can be rewritten as (Andrews [2], Lemma 8) $v^*(\theta_n-\theta_0,\varphi_n-\varphi_0,R_n^*+\epsilon_n^*)=\mathbf{0}_p$ where

$v:\mathbb{R}^p\times\mathbb{R}^q\times\mathbb{R}^m\to\mathbb{R}^m$ is smooth and $\epsilon_n^*=\begin{pmatrix} r_n^* \\ \mathbf{0}_{m-p}\end{pmatrix}$. If we denote with R^* the probability limit of R_n^*, and with R the probability limit of R_n, then it is trivial that $v\left(\mathbf{0}_p,\mathbf{0}_q,R^*\right)=\mathbf{0}_p$, and that $\frac{\partial v(z,x,y)}{\partial z'}|_{(\mathbf{0}_p,\mathbf{0}_q,R^*)}=\frac{\partial J^2(\theta_0,\varphi_0)}{\partial\theta\partial\theta'}$ which is positive definite by 2. Hence the implicit function theorem applies and dictates that $\exists U_{\mathbf{0}_p}\subseteq\mathbb{R}^p$ an open neighborhood of $\mathbb{R}^p$, $\exists V_{(\mathbf{0}_q,\mathbf{R})}\subseteq\mathbb{R}^q\times\mathbb{R}^m$ an open neighborhood of $(\mathbf{0}_q,R^*)$, and a unique smooth function $\pi^*:V_{(\mathbf{0}_q,\mathbf{R}^*)}\to U_{\mathbf{0}_p}$ such that $v\left(\boldsymbol{\pi}^*\left(x,y\right),x,y\right)=\mathbf{0}_p\ \forall\left(x,y\right)\in V_{(\mathbf{0}_q,\mathbf{R}^*)}$. Given that $P_{\theta_0}\left(\theta_n-\theta_0\in U_{\mathbf{0}_p}\right)=1-o\left(n^{-a}\right)$, $P_{\theta_0}\left(\varphi_n-\varphi_0\in U_{\mathbf{0}_q}\right)=1-o\left(n^{-a}\right)$ $P_{\theta_0}\left(R_n^*+\epsilon_n^*-R^*\in U_{\mathbf{0}_m}\right)=1-o\left(n^{-a}\right)$, $\varphi_n-\varphi_0=\pi\left(R_n\right)+\epsilon_n$, $P_{\theta_0}\left(\sqrt{n}\left\|\epsilon_n\right\|>\omega_n^*\right)=o\left(n^{-a}\right)$, $\omega_n^*=o\left(n^{-a}\right)$ and $\pi\left(.\right)$ is smooth, we obtain that for large enough n
$\theta_n-\theta_0=\pi^*\left(\varphi_n-\varphi_0,R_n^*+\epsilon_n^*\right)=\pi^*\left(\pi\left(R_n\right)+\epsilon_n,R_n^*+\epsilon_n^*\right)$
with $\pi^*\left(\mathbf{0}_q,R^*\right)=\mathbf{0}_p$. Consequently, we have to find $C>0$ such that for $\omega_n=Cn^{\frac{1}{2}-\frac{d}{2}}\ln^{\frac{d}{2}}n=o\left(n^{-a}\right)$ we have

$$P_{\theta_0}\left(n^{\frac{1}{2}}\left\|\pi^*\left(\pi\left(R_n\right)+\epsilon_n,R_n^*+\epsilon_n^*\right)-\pi^*\left(\pi\left(R_n\right),R_n^*\right)\right\|>\omega_n\right)=o\left(n^{-a}\right)$$

and

$$P\left(\sqrt{n}\left\|D^{j_1,j_2}\left(\frac{\partial Q_n^{/}\left(\theta_0,\varphi_0\right)}{\partial\theta}W\left(\varphi_0\right)\right)\begin{pmatrix}\overline{\theta_n}^{j_1}\\ \overline{\varphi_n}^{j_2}\end{pmatrix}Q_n\left(\theta_0,\varphi_0\right)\right\|>\omega_n\right)=o\left(n^{-a}\right).$$

Following the proof of Andrews [2], Lemma 8, we conclude that for $C>K\left(C_\varphi\right)^{j_2}\left(C_\theta\right)^{j_1}C_Q$, where C_φ, C_θ, and C_Q are as in the declaration of the lemma, and K is an asymptotic bound in probability of the Lipschitz coefficient of the highest order derivative of $\frac{\partial J_n(\theta,\varphi)}{\partial\theta}$, both equations above are fulfilled. ■

The next two results are of great importance in both the validity of Edgeworth expansions as well as in the validation of moment approximations. In fact the lemma below represents an extension of Lemma 8 in Andrews [2].

Lemma AL.6 *Suppose that $\sqrt{n}\left(\beta_n-b\left(\theta_0\right)\right)$ admits a valid Edgeworth expansion of order $s=2a+1$. Let $\{x_n\}$ denote a sequence of random vectors and there exists an $\varepsilon>0$ and a real sequence $\{a_n\}$, such that $a_n=o\left(n^{-\varepsilon}\right)$ and $P\left(\sqrt{n}\left\|x_n\right\|>a_n\right)=o\left(n^{-a}\right)$. Then $\sqrt{n}\left(\beta_n-b\left(\theta_0\right)+x_n\right)$ admits a valid Edgeworth expansion of the same order.*

Proof.
By definition $\sup_{A\in\mathcal{B}_C}\left|P\left(\sqrt{n}\left(\beta_n-b\left(\theta_0\right)\right)\in A\right)-\int_A\left(1+\sum_{i=1}^{2a}n^{-\frac{i}{2}}\pi_i\left(z\right)\right)\phi\left(z\right)dz\right|=o\left(n^{-a}\right)$ where $\mathcal{B}_C$ denotes the collection of convex Borel sets of $\mathbb{R}^q$ and

$\pi_i(z) = O(1)$. Then, $P\left(\sqrt{n}\left(\beta_n - b(\theta_0) + x_n\right) \in A\right)$
$\leq P\left(\sqrt{n}\left(\beta_n - b(\theta_0)\right) \in A - a_n\right) + o\left(n^{-a}\right)$. Also,
$\sup_{A\in\mathcal{B}_C}\left|P\left(\sqrt{n}\left(\beta_n - b(\theta_0) + x_n\right) \in A\right) - \int_{A-a_n}\left(1+\sum_{i=1}^{2a} n^{-\frac{i}{2}}\pi_i(y)\right)\phi(y)\,dy\right| \leq$

$\sup_{A\in\mathcal{B}_C}\left|P\left(\sqrt{n}\left(\beta_n - b(\theta_0)\right) \in A - a_n\right) - \int_{A-a_n}\left(1+\sum_{i=1}^{2a} n^{-\frac{i}{2}}\pi_i(y)\right)\phi(y)\,dy\right| +$
$o\left(n^{-a}\right) = o\left(n^{-a}\right)$ as $A - a_n$ is convex.
Now, $\int_{A-a_n}\left(1+\sum_{i=1}^{2a} n^{-\frac{i}{2}}\pi_i(y)\right)\phi(y)\,dy = \int_{A}\left(1+\sum_{i=1}^{2a} n^{-\frac{i}{2}}\pi_i(z-a_n)\right)\phi(z-a_n)\,dz$.
Therefore, if $H_k(z)$ denotes the k^{th} order Hermite multivariate polynomial, $L(H_k(z), a_n, i)$ and i-linear function of a_n with coefficients from $H_k(z)$, and $\phi(z-a_n) = \phi(z)\sum_{k=0}^{K}\frac{1}{k!}L(H_k(z), a_n, k) + \rho_n(z)$ where $\rho_n(z) = \frac{1}{(2K+1)!}(-1)^{K+1}L(H_k(z-a_n^*), a_n, K+1)\,\phi(z-a_n)$, and a_n^* lies between a_n and zero. If $a \leq \varepsilon$ set $K=0$, else, choose some natural $K \geq \frac{a}{\varepsilon} - 1$.
Then, $\left(1+\sum_{i=1}^{2a} n^{-\frac{i}{2}}\pi_i(z-a_n)\right)\phi(z-a_n)$
$= \phi(z)\left(1+\sum_{i=1}^{2a} n^{-\frac{i}{2}}\pi_i^*(z)\right) + q_n(z)$ where the $\pi_i^*(z)'s$ are $O(1)$ polynomials in z and $q_n(z) = \left(1+\sum_{i=1}^{2a} n^{-\frac{i}{2}}\pi_i(z-a_n)\right)\rho_n(z)$.
Hence $\int_A\left(1+\sum_{i=1}^{2a} n^{-\frac{i}{2}}\pi_i(z-a_n)\right)\phi(z-a_n)\,dz$
$= \int_A \phi(z)\left(1+\sum_{i=1}^{2a} n^{-\frac{i}{2}}\pi_i^*(z)\right)dz$
$+ \int_A \phi(z)\,q_n(z)\,dz$ and $\sup_{A\in\mathcal{B}_C}\left|\int_A q_n(z)\,dz\right|$
$\leq \sup_{A\in\mathcal{B}_C}\int_A \phi(z-a_n)\left|\left(1+\sum_{i=1}^{2a} n^{-\frac{i}{2}}\pi_i(z-a_n)\right)\rho_n(z)\right|dz$
$\leq \int_{\mathbb{R}^q}\phi(z-a_n)\left|\left(1+\sum_{i=1}^{2a} n^{-\frac{i}{2}}\pi_i(z-a_n)\right)\rho_n(z)\right|dz \leq \frac{C}{n^{a+\delta}} = o\left(n^{-a}\right)$ for some $C, \delta > 0$. Hence,
$\sup_{A\in\mathcal{B}_C}\left|R_n - \int_A q_n(z)\,dz\right| = o\left(n^{-a}\right)$, and therefore $\sup_{A\in\mathcal{B}_C}\left|R_n - \int_A q_n(z)\,dz\right|$
$\geq \sup_{A\in\mathcal{B}_C}\left|R_n - \left|\int_A \phi(z)\,q_n(z)\,dz\right|\right| \geq \left|\sup_{A\in\mathcal{B}_C}|R_n| - \sup_{A\in\mathcal{B}_C}\left|\int_A \phi(z)\,q_n(z)\,dz\right|\right| =$
$o\left(n^{-a}\right)$ and $\sup_{A\in\mathcal{B}_C}\left|P\left(\sqrt{n}\left(\beta_n - b(\theta_0) + x_n\right) \in A\right) - \int_A \phi(z)\left(1+\sum_{i=1}^{2a} n^{-\frac{i}{2}}\pi_i^*(z)\right)dz\right| =$
$o\left(n^{-a}\right)$ due to the fact that the transformation from $\pi_i(z)$ to $\pi_i^*(z)$ does not depend on A but only on a_n and $R_n = P\left(\sqrt{n}\left(\beta_n - b(\theta_0) + x_n\right) \in A\right)$
$- \int_A \phi(z)\left(1+\sum_{i=1}^{2a} n^{-\frac{i}{2}}\pi_i^*(z)\right)dz$. ∎

Corollary AC.1 *If $a \leq \varepsilon$ then $\pi_i(z) = \pi_i^*(z)$, $\forall i$, and therefore the resulting Edgeworth distribution coincides with the initial.*