STOCHASTIC EXPANSIONS AND MOMENT APPROXIMATIONS FOR THREE INDIRECT ESTIMATORS

by

Stelios Arvanitis [1]

and

Antonis Demos [2]

DISCUSSION PAPER No. 198
June 2010

Department of Economics
Athens University of Economics and Business
76 Patission Str., Athens 104 34, Greece
Tel. (++30) 210-8203911 - Fax: (++30) 210-8203301

1 Dept. of Economics, Athens University of Economics and Business.

2 Dept. of International & European Economic Studies, Athens University of Economics and Business.

Stochastic Expansions and Moment Approximations for Three Indirect Estimators

Stelios Arvanitis and Antonis Demos
Athens University of Economics and Business

January 14, 2010

Abstract

This paper is concerned with some properties of three indirect estimators that are known to be *(first order) asymptotically equivalent.* Specifically, for each one of them, we examine a) the issue of validity of the formal Edgeworth expansion of an arbitrary order. b) Given the establishment of validity, we are concerned with valid moment approximations and employ them to characterize the *bias* structure of the estimators up to this order. 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 is derived without any establishment of validity. We provide this establishment, but we also are able to massively generalize the conditions under which this second order property remains true. 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.* We also provide a further generalization of the indirect estimators by introducing recursive ones emerging from multistep optimization procedures. Upon strengthening the validity of the aforementioned moment approximations, we are able to establish higher order unbiaseness for estimators of this sort.

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

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 used for statistical estimation (e.g. moment conditions). In this respect, a collection of random elements used to define an estimation procedure in one model, can be pulled back to another and therefore used in a similar manner, thereby indirectly facilitating inference. When these collections of measures have additional structure (for example, when they are finite dimensional differentiable manifolds, as is the case with differential parametric finite dimensional statistical models), the resulting "binding" can be chosen so that (at least locally) it respects this structure, something that can facilitate the derivation of results and/or the analysis of the properties of such procedures.

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 issue of validity of the formal Edgeworth expansion of the 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 establishment of 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, is derived without any establishment of validity. We provide this establishment, but we also are able to massively generalize the conditions under which this second order property remains true. There are

[1] Although it can be extended into a semiparametric framework, see [7].

no 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.[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. 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]

The Binding Function

The central notion of indirect inference procedures is the one of the binding function. In pure terms this constitutes of a function between the measures involved in the relevant statistical models. This function can be formed as the pushover of an automorphism of the underlying probability space. Such a derivation of the binding function would be in accordance with the generic efficiency loss of indirect estimators, due to the fact that the observed sample is not subjected to the underlying automorphism, and/or that the estimating equation does not constitute a basis of the vector space spanned by the score at the true parameter values, and/or due to non linearities of the function. In any case the binding function is denoted by $b(\theta)$, where θ the parameter vector to be estimated, and what is usually discussed is not the function

[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.

itself, but a parametric representation of it (see the paragraph entitled as General Assumption Framework).

The Auxiliary and the Indirect Estimators

All three indirect estimators essentially involve two step estimation procedures. In the first step, the estimating equation that is part of the structure of the auxiliary model, is used 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 estimators are collectively denoted in the paper by β_n whereas θ_n denote the indirect ones, with n being the sample size.

We consider one type of auxiliary estimator. It 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. The last remark makes possible the computation of this matrix with respect to an initial auxiliary 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 GMR 1 and it was proposed by [9] in order for the numerical burden of the second estimator to be relaxed. The latter is termed GMR 2 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 [12] and minimizes an analogous distance between the *conditional expectation* of the auxiliary estimating vector and zero. 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 GMR 1 estimator is the one involved with the maximal numerical burden among the three.

[11] show that the GMR 2 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. [4]

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 the initial estimator (see the paragraph entitled as General Assumption Framework), on 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 in the case of GMR1 and GT (see lemma 1.2) then $b_n(\theta, W_n, \theta_n^*) = b(\theta)$, while in the case of GMR2 $b_n(\theta, W_n, \theta_n^*) = E_\theta \beta_n$ (see the preceding paragraph). Hence the stochastic weighting, essentially generalizes the structure of the functions from the inversion of which the Indirect Estimation (IE) are derived.[5]

Generalizations to Multistep Procedures

We are able to extend the definition of IE (in the particular case of the GMR2 one), through the employment of recursive multistep procedures based on the existing definitions. These are motivated by the bias structure of the GMR2 estimator as obtained later, and the fact that these kind of generalizations can lead to indirect estimators that are (globally) unbiased for any given order. We provide the analogous definitions and results in the section entitled as GMR2 recursion. It will be evident but not examined here, that analogous generalizations can be defined in ways that involve any combination of the aforementioned IE.

[4] [11] 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 [8] (chapter 4).

[5] 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.

Edgeworth and Moment Approximations of Sequences of Distributions

As previously noted we are concerned with the validity of the approximation of sequences of distributions (namely the ones emerging from the sequences of the examined estimators). 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. 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 (i). Consider now two arbitrary sequences (say M_n and M_n^*) of 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 $s = 2a + 1$. Some remarks on these definitions are the following:[6]

- Due to (i), 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 asymptotic approximation of order s sequence M_n^* is simply an element of a closed ball with center M_n and an radius that depends on a.

- 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.

- 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*

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

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

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 $\int_S fd\cdot$ on $\left(\left(\mathcal{MF}\left(S\right)\right)^{\mathbb{N}}, M_0\right)$ 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 $\int_S fd\cdot:\left(\left(\mathcal{MF}\left(S\right)\right)^{\mathbb{N}}, M_0\right) \to \left(\left(\mathbb{R}^q\right)^{\mathbb{N}}, x_0\right)$ this function 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 section entitled as "Validity of 1st moment approximation" in the case where $S = \mathbb{R}^q$ and $f = id_{\mathbb{R}^q}$. This conditions are reminiscent of 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).

Outline of the paper

We immediately provide the assumption framework needed for the definition of the examined estimators. We then provide assumptions sufficient for and derive the validity of the Edgeworth approximations. In the following sections, we provide assumptions that validate the first moment approximations given the previous results, derive the approximations for $a = \frac{1}{2}$, discuss the bias properties of the estimators, and provide multistep extensions of the GMR2 estimators that have desirable bias properties of general order. In the last section we conclude. In the appendix, we provide a series of useful to our derivations general lemmas.

1.1 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 $\mathcal{O}_{\varepsilon}\left(\theta\right)$ will denote the ε-ball around θ in a relevant metric space and let $d = \max\left(2a+2, 3\right)$.

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 parameterization, that is a (k^{th}-order) diffeomorphism (for $k \geq d$), say **par** to an open 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 θ_0 =**par**(D_0).*

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*.

Assumption A.2 *Θ and B are bounded.*

Remark R.1 *Since Θ is a bounded subset of a finite dimensional Euclidean space it is also totally bounded.*

Remark R.2 *D could be extended (restricted) so as to be homeomorphic to a compact superset (subset) of Θ, say Θ^*. In this case and in order for the differentiability properties to be retained the previous assumption could be completed with $\theta_0 \in Int(\Theta^*)$.*

It is evident that the previous remark also applies in the case of B and that the binding function is by definition **bounded**.

Assumption A.3 *$b(\theta_0) = b(\theta)$ 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.*

[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, that could be involved in properties of the statistical model, as well as in the definition and the properties of the binding function, to be later presented.

[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 [1].

Remark R.3 *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 indirect inference, hence it is termed as* ***local indirect identification****.*

We strengthen the previous assumption in our differentiable context as follows:

Assumption A.4 *For some $\varepsilon_1 \geq \varepsilon_2 > 0$, the restriction $b|_{\mathcal{O}_{\varepsilon_2}(\theta_0)} : \mathcal{O}_{\varepsilon_2}(\theta_0) \rightarrow B$ is a $k-$diffeomorphism.*

Remark R.4 *The previous assumption that $q \geq p$ and that rank $\left(\frac{\partial b}{\partial \theta'}\right) = p$, $\forall \theta \in \mathcal{O}_{\varepsilon_2}(\theta_0)$.*

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

Assumption A.5 *p, q, l are finite and $p \leq q \leq l$.*

The following set of assumptions deal with the structure of the derivatives of c as well as of the likelihood function.

Assumption A.6 *Integration with respect to the measures involved in the statistical model and derivation with respect to θ and β are commutative.*

Remark R.5 *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 [4], theorem 9.31).*

In the following we will 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.7 *$b(\theta)$ 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 $M_r \in \mathbb{R}$.*

Remark R.6 *Notice that $D^r(\cdot)$ denotes the vector containing the partial derivatives of the relevant order of the differentiated function. The above assumption is obviously true for $r = 1$ as $b(\theta)$ is Lipschitz on Θ and consequently on $\mathcal{O}_{\varepsilon_3}(\theta_0)$.*

Here after we will not *explicitly* refer to assumptions 1-8 in the statement of our results. These will be considered to formalize the most basic framework, on which the assumption bellow will operate. Also note that due to the fact that the spaces Θ and B are separable, suprema of real random elements over these spaces are typically measurable.

Assumption A.8 *$c(\cdot,\beta)$ is $\mathcal{B}_{\mathbb{R}^l}/\mathcal{B}_{\mathbb{R}^m}-$measurable for every $\beta \in B$, 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, $\sup_{\theta\in\mathcal{O}_\eta(\varphi_0)} \left\|\frac{\partial}{\partial\theta'} E_\theta\left[c(x,\beta)\right]\right\|$ and $\sup_{\theta\in\mathcal{O}_\eta(\varphi_0)} \left\|\frac{\partial}{\partial\theta_i\partial\theta_j} E_\theta\left[c(x,\beta)\right]\right\|$ are bounded $\forall i,j = 1,\ldots p$ for some $\eta > 0$, where $\varphi_0 = \left(b^{/}(\theta_0), \theta_0^{/}\right)^{/}$.*

Remark R.7 *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 and the assumptions bellow implies that $l \geq q$, and rank $\left(E_\theta \frac{\partial c(x,\beta)}{\partial \beta'}\right) = q$, rank $\left(E_\theta \frac{\partial^2 \rho(x,\beta)}{\partial\beta\partial\beta'}\right) = q$, $\forall \beta \in b\left(\mathcal{O}_{\varepsilon_2}(\theta_0)\right)$, $\forall \theta \in \mathcal{O}_{\varepsilon_2}(\theta_0)$.*

Remark R.8 *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.9 *The function c and the estimating equations $E_\theta c(x, b(\theta)) = \mathbf{0}_{l\times 1}$ can be derived as part of the structure of a second (potentially misspecified), differentiable parametric μ-dominated statistical model defined on the same measurable space, say D^*, usually termed as **auxiliary model**, with B as its parameter space. In this case the restricted binding function is a parametric representation of a relevant function (with similar properties) between the two sets of probability measures (properly restricted). For example, if $g : \mathbb{R}^m \to \mathbb{R}^m$ is $\mathcal{B}_{\mathbb{R}^m}/\mathcal{B}_{\mathbb{R}^m}-$measurable, and in the structure of the auxiliary model are the conditions $E_\beta c(x,\beta) = \mathbf{0}_{l\times 1}$ in some neighborhood of $b(\theta_0)$, then the change of variables formula implies the conditions $E_\theta c(g(x), b(\theta)) = \mathbf{0}_{l\times 1}$ of the previous assumption. It should also be noted that the binding function can be locally retrieved from the previous conditions through results of the sort of the implicit function theorem.*[9]

[9] The binding function is usually in practice anallyticaly unknown and approximated with numerical simulations. Our results do not concern this case and the rationale of this choice becomes evident later.

Remark R.10 *Conditions of the sort* $\sup_\theta E_\theta \|u_\rho\|^{q_0} < \infty$*, holding globally on* Θ *and locally on* B *are typically used in the case of the GT estimator.*

In the following we denote $\frac{1}{n}\sum_{i=1}^{n} c(x_i, \beta)$ with $c_n(\beta)$.

Remark R.11 *Similarly, conditions boundeness of quantities such as* $\sup_{\theta\in\mathcal{O}_\eta(\varphi_0)} \left\|\frac{\partial}{\partial\theta'} E_\theta [c(x,\beta)]\right\|$*, holding locally on* $\Theta \times B$ *are typically used in the case of the GT estimator and can be derived from conditions like* $\lim_{n\to\infty} \sup_{\theta\in\mathcal{O}_\eta(\varphi_0)} E_\theta \|\sqrt{n}c_n(\beta)\|^2 < \infty$ *and* $\lim_{n\to\infty} \sup_{\theta\in\mathcal{O}_\eta(\theta_0)} E_\theta \|\sqrt{n}\overline{s}_n(\theta)\|^2 < \infty$ *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.10 and R.24 for analogous conditions).*

Assumption A.9 *Let* $W(x,\beta)$*,* $W^*(x,\theta)$ *and* $W^{**}(x,\theta)$ *be* $l\times l$*,* $q \times q$ *and* $l \times l$ *(or* $q \times q$ *see the definition of the GT estimator)* μ*-almost surely positive definite random matrices such that d-differentiable* $\forall\beta \in b(\mathcal{O}_{\varepsilon_2}(\theta_0))$*,* $\forall\theta \in \mathcal{O}_{\varepsilon_2}(\theta_0)$*, 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* $E_{\theta_0}\|W(x, b(\theta_0))\|^{q_0} < \infty$*,* $E_{\theta_0}\|W^*(x,\theta_0)\|^{q_0} < \infty$ *and* $E_{\theta_0}\|W^{**}(x,\theta)\|^{q_0} < \infty$ *for* q_0 *defined above.*

Remark R.12 *This assumption essentially implies that the aforementioned 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.*

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.

1.2 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 on $\mathbb{R}^k \times \mathcal{PD}(k\times k)$ for $k \in \mathbb{N}$

$$(x, A) \to (x'Ax)^{1/2}$$

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 size of n.

Auxiliary Estimators

The auxiliary estimator (β_n) is defined next.

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

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

Remark R.13 *For large enough* n*, the estimator is defined as an* μ*-almost sure global minimum.*

Remark R.14 *When* $l = q$*, and under the assumption framework* β_n *becomes almost surely independent of the weighting matrix.*

Indirect Estimators

Given the definition of the auxiliary estimator we define the indirect ones. The remarks on the definition of the auxiliary estimator apply under the appropriate alterations to the indirect estimators due to their structure as distance minimizers. 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 [9] while the third was introduced by [12] (see also [8], chapter 4, for a summary).

GMR 1 Estimator The first GMR estimator is defined as:

Definition D.2 *The GMR 1 estimator* $\theta_n : \Omega^n \to B$ *is defined as*

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

Remark R.15 *Even the computation of the estimator relies on the analytical knowledge of the binding function. In this respect this estimator is almost always intractable. Due to this fact in applications, under the appropriate conditions, a version of this estimator is defined, in which the unknown binding function is approximated by the computation of* β_n *on a large simulated path. Notice that the corresponding estimator (sub)sequence is well defined (at least for large enough* n*) given that* b *is invertible around* θ_0*, and that* β_n *converges to* $b(\theta_0)$*.*

GMR 2 Estimator Due to the boundeness of B we have the following lemma.

Lemma 1.1 $\|E_\theta \beta_n\| < \infty$

Proof. $\|E_\theta \beta_n - b(\theta)\| \leq E_\theta \|\beta_n - b(\theta)\| \leq M_1$, where M_1 denotes the diameter of B. The result follows due to the boundeness of b. ∎

Hence the following definition becomes possible:

Definition D.3 *The GMR 2 estimator $\theta_n : \Omega^n \to B$ is defined as*

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

Remark R.16 *Again in most cases even the computation of θ_n is analytically intractable due to the fact that $E_\theta \beta_n$ is unknown. Again in applications, under the appropriate conditions, a version of this estimator is defined, in which the unknown expectation is approximated by the computation of β_n on a large set of simulated paths. Again the corresponding estimator (sub)sequence is well defined (at least for large enough n) given that the function sequence $E_\theta \beta_n$ is invertible around θ_0, and that β_n converges to $b(\theta_0)$.*

GT Estimator Due to assumption 5, the definition that comes next becomes possible. We denote by $E_\theta (c_n(\beta_n))$, the quantity $E_\theta (c_n(\beta))|_{\beta=\beta_n}$ for notational simplicity.

Remark R.17 *Due to assumption 5 the $\|E_\theta (c_n(\beta_n))\| < \infty$, $\forall\theta$, hence the following minimization procedure can be defined.*

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

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

Remark R.18 *The usual definition of the aforementioned estimator is given only when the auxiliary estimator is the M.L.E. of the auxiliary model. The currently defined one is an obvious extension.*

Remark R.19 *Again the computation of the estimator relies on the analytical knowledge of the engaged expectation, which is usually intractable. In this respect this estimator is also almost always intractable. Due to this fact in applications, under the appropriate conditions, an approximation of this estimator is defined, in which the unknown expectation is approximated by the computation of $c_n(\beta_n)$ on a large simulated path or equivalently on a large set of simulated paths. The local invertibility of the binding function implies via the implicit function theorem that $E_\theta (c_n(\beta)) = \mathbf{0}_{l\times 1}$ iff $\beta = b(\theta)$, $\forall\theta \in \mathcal{O}(\theta_0, \varepsilon_2)$, hence the corresponding estimator (sub)sequence is well defined (at least for large enough n).*

Remark R.20 *The first step estimator* θ_n^* *is again supposed to be defined as any of the indirect estimators with the restriction that the relevant weighting matrix is deterministic and independent of* θ*.*

Relationship between the three indirect estimators As the asymptotic expansions presented in the results section of the paper will show, in accordance with the relevant literature the three estimators are asymptotically first order equivalent (proviso a certain selection of the weighting matrix of GMR 1 and GMR 2 given the weighting matrix of the GT estimator). However, in the special case where $p = q$, a special relationship is revealed between the GMR 1 and the GT estimators by the following lemma.

Lemma 1.2 *Given consistency, and* $p = q = l$*, with probability* $1 - o(n^{-a})$

$$GMR1{=}GT$$

Proof. 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.8 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. ■

Remark R.21 *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.22 *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 1.3 *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.

Proof. 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. ■

Remark R.23 *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.*

2 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. Remember that every estimator considered is an extremum one, and the criterion from which it emerges is at least locally differentiable. Accommodating this fact, in order for the derivation of the aforementioned validity we employ the following steps. First, we prove that the estimators satisfy the first order conditions with probability $1 - o\left(n^{-a}\right)$. 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 $o\left(n^{-a}\right)$ real sequence with probability $1-o\left(n^{-a}\right)$. This due to corollary AC.1 implies 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 [2] an is essentially based on local differentiability, lemma AL.3 and [3] which provide a theorem of invariance of validity of Edgeworth approximations with respect to locally differentiable functions. Notice also that lemma AL.3 enables the extension of the results in non differentiable case, but this will not be pursued here.

2.1 Assumptions Specific to the Validity of the Edgeworth Approximations

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

Assumption A.10 *$\sup_{\theta \in \mathcal{O}_{\varepsilon_4}(\theta_0)} \left\| D^r E_\theta \beta_n \right\| < M_r^*$, for $0 < \varepsilon_4 \leq \varepsilon_2$, for $r = 2, \ldots, d+1$, and $M_r^* > 0$.*

Remark R.24 *Assumption A.10 along with Assumption A.7 imply that for $r = 2, \ldots, d+1$, $\sup_{\theta \in \mathcal{O}_{\min(\varepsilon_3, \varepsilon_4)}(\theta_0)} \left\| D^r \left(E_\theta \beta_n - b\left(\theta\right)\right) \right\| < M_r + M_r^*$, which in turn means that $D^{r-1}\left(E_\theta \beta_n - b\left(\theta\right)\right)$ are uniformly Lipschitz on $\mathcal{O}_{\min(\varepsilon_3, \varepsilon_4)}\left(\theta_0\right)$, 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 $B\left(\theta_0, \min\left(\varepsilon_3, \varepsilon_4\right)\right)$. Due to Assumption A.1 for $k \geq d+1$, this assumption is verified via conditions of the form $\sup_{\theta \in \mathcal{O}_{\min(\varepsilon_3, \varepsilon_4)}(\theta_0)} E_\theta \left\| \sqrt{n}\left(\beta_n - b\left(\theta\right)\right) \right\|^2 = O\left(1\right)$ and $\sup_{\theta \in \mathcal{O}_{\min(\varepsilon_3, \varepsilon_4)}(\theta_0)} E_\theta \left\| \sqrt{n} \bar{l}_n\left(\theta\right) \right\|^2 =$*

$O(1)$ where $\bar{l}_n(\theta)$ depends on derivatives of the (well defined in our setting) average likelihood function. For example for $r = 2$, we have that $\bar{l}_n(\theta) = \overline{s_n(\theta)\, s_n'(\theta)} + \overline{H}_n(\theta)$.

Assumption A.11 *$E_\theta \|f(x,\beta)\|^{q_1} < \infty$, $\forall\theta \in \mathcal{O}_{\varepsilon_2}(\theta_0)$, $\forall\beta \in b(\mathcal{O}_{\varepsilon_2}(\theta_0))$ for $q_1 = 2a+3$, $\|f(x,\beta) - f(x,\beta')\| \leq \kappa_\gamma \|\beta - \beta'\|$, $\forall\beta \in b(\mathcal{O}_{\varepsilon_2}(\theta_0))$, μ-almost surely for an almost surely positive random variable κ_γ, with $E_\theta \kappa_\gamma^{q_1} < \infty$, $\forall\theta \in \mathcal{O}_{\varepsilon_2}(\theta_0)$.*[10]

Remark R.25 *This condition that could be termed as local stochastic Lipschitz continuity condition facilitate the Edgeworth approximations of the relevant sequences of random elements.*

Assumption A.12 *The Weak Dependence assumption and the Cramer type of condition of [2] or [10] 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.26 *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 [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.13 *. The initial estimators are derived from a, relevant to assumptions 1-11, framework. The relevant sequences of distributions of the initial estimators, β_n^* and θ_n^* can be approximated by a sequence of Edgeworth distributions of order of error $o(n^{-a})$.*

Remark R.27 *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^* (see below).*

We 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.

[10] Notice the local nature of the moment existence conditions here and in assumption A.8. These are stronger that the relevant conditions of [2], and facilitate mainly the case of the GT estimator.

Auxiliary Estimator

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

Lemma 2.1 *Under assumptions A.1, A.2, A.5, A.8, A.9, and A.11-A.13 there exists an Edgeworth distribution* $\mathcal{EDG}_a(\bullet)$ *such that*

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

Proof. Notice that assumptions 1-4 in [2] correspond to assumptions A.1, A.2, A.5, A.8, A.9, and A.11-A.13. The result follows from Lemmas 5 and 9 of [2]. ■

Indirect Estimators

We next present in a more detailed manner, as described in the introduction of the present section, the analogous results for the indirect estimators. We begin first with the issue of the rate at which the probability of the event that the estimator belongs to an arbitrary neighborhood of θ_0, approaches unity. We term it $o\left(n^{-a}\right)$-consistency. Then we are occupied with the rate at which the probability of the event that the $\sqrt{n}$ transformation of the estimator lies in a σ-compact subset of $\mathbb{R}^p$, approaches unity. We term it $o\left(n^{-a}\right)$-tightness. Notice that the latter implies the former. The relevant lemmas are announced in such manner so that only $o\left(n^{-a}\right)$-tightness is explosively presented. However the $o\left(n^{-a}\right)$-consistency is established as a first step for the establishment of the tightness and therefore it lies in the proofs of the lemmas. Then, the validity of the Edgeworth approximation is established separately.

GMR 1 estimator: $o\left(n^{-a}\right)$-Consistency and $o\left(n^{-a}\right)$-Tightness The results for the GMR 1 estimator are presented here. These follow directly from the previous results and the fact that the binding function has *bounded derivatives of any of the supposed orders.*

Lemma 2.2 *Under the validity of lemmas 2.1, AL.1 and assumption A.7*

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

Proof. The initial estimator, θ_n^* is defined in assumption A.9, and by Lemma AL.1 we have that
$P_{\theta_0}\left(\|W^*(\theta_n^*) - W^*(\theta_0)\| > \varepsilon\right) = o\left(n^{-a}\right), \forall \varepsilon > 0$. Now notice that $p\lim\left(\beta_n - b(\theta)\right) = b(\theta_0) - b(\theta)$. Hence $\forall \theta \in \Theta$ and $\forall \varepsilon > 0$

$$P_{\theta_0}\left(\sup_{\theta\in\Theta}\|\beta_n - b(\theta) - [b(\theta_0) - b(\theta)]\| > \varepsilon\right) = P_{\theta_0}\left(\|\beta_n - b(\theta_0)\| > \varepsilon\right) = o\left(n^{-a}\right)$$

from Lemma 2.1 and consequently, the consistency of θ_n follows from Lemma 5 of [2].
Hence θ_n is in the interior of Θ and $\frac{\partial}{\partial\theta}J_n(\theta_n) = 0$ with probability $1 - o\left(n^{-a}\right)$, where $J_n(\theta) = \left(\beta_n - b(\theta)\right)^{/} W_n^*(\theta_n^*)\left(\beta_n - b(\theta)\right)$. Hence element by element mean value expansions of $\frac{\partial}{\partial\theta}J_n(\theta_n)$ around θ_0 and rearrangement gives: $\theta_n - \theta_0 = -\left(\frac{\partial^2}{\partial\theta\partial\theta^{/}}J_n\left(\theta_n^+\right)\right)^{-1}\frac{\partial}{\partial\theta}J_n(\theta_0)$ with probability $1 - o\left(n^{-a}\right)$, where θ_n^+ lies between θ_n and θ_0 and may be different across rows. Hence it suffices to show that $A)$ $P_{\theta_0}\left(\left\|\frac{\partial}{\partial\theta}J_n(\theta_0)\right\| > C^*\frac{\ln^{1/2} n}{n^{1/2}}\right) = o\left(n^{-a}\right)$ and $B)$ $P\left(\left\|\left(\frac{\partial^2}{\partial\theta\partial\theta^{/}}J_n\left(\theta_n^+\right)\right)^{-1}\right\| > K\right) = o\left(n^{-a}\right)$.
For $A)$ notice that $P_{\theta_0}\left(\left\|\frac{\partial}{\partial\theta}J_n(\theta_0)\right\| > C^*\frac{\ln^{1/2} n}{n^{1/2}}\right) =$
$P_{\theta_0}\left(\left\|-2\frac{\partial}{\partial\theta}b(\theta_0)^{/} W_n^*(\theta_n^*)\left(\beta_n - b(\theta_0)\right)\right\| > C^*\frac{\ln^{1/2} n}{n^{1/2}}\right) \leq$
$P_{\theta_0}\left(\|W_n^*(\theta_n^*)\|\,\|\beta_n - b(\theta_0)\| > C\frac{\ln^{1/2} n}{n^{1/2}}\right)$ where $C = \frac{C^*}{2\left\|\frac{\partial}{\partial\theta}b(\theta_0)^{/}\right\|}$ by the submultiplicative property of the norm and by assumption A.7 $\left\|\frac{\partial}{\partial\theta}b(\theta_0)^{/}\right\| > 0$.
Hence we have that for $K > 0$ we have $P_{\theta_0}\left(\|W_n^*(\theta_n^*)\| > K\right) = o\left(n^{-a}\right)$, which is true from Lemma AL.1 and $P_{\theta_0}\left(\|\beta_n - b(\theta_0)\| > C\frac{\ln^{1/2} n}{n^{1/2}}\right)$ by Lemma 2.1 and the result follows.
For $B)$ notice that $\frac{\partial^2}{\partial\theta\partial\theta^{/}}J_n\left(\theta_n^+\right) = 2\frac{\partial}{\partial\theta}b\left(\theta_n^+\right)^{/} W_n^*(\theta_n^*)\frac{\partial}{\partial\theta^{/}}b\left(\theta_n^+\right)$
$- 2\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}b\left(\theta_n^+\right)^{/} W_n^*(\theta_n^*)\left(\beta_n - b\left(\theta_n^+\right)\right)\right]_{i,j=1,\ldots,p} = B_1 + B_2$. It suffices to show that for $K > 0$ we have that $P_{\theta_0}\left(\|B_1 + B_2\| > K\right) = o\left(n^{-a}\right)$. But $P_{\theta_0}\left(\|B_1 + B_2\| > K\right) \leq P_{\theta_0}\left(\|B_1\| > \frac{K}{2}\right) + P_{\theta_0}\left(\|B_2\| > \frac{K}{2}\right)$. Now $P_{\theta_0}\left(\|B_1\| > \frac{K}{2}\right) =$
$P_{\theta_0}\left(\left\|\frac{\partial}{\partial\theta}b\left(\theta_n^+\right)^{/} W_n^*(\theta_n^*)\frac{\partial}{\partial\theta^{/}}b\left(\theta_n^+\right)\right\| > \frac{K}{4}\right) \leq$
$P_{\theta_0}\left(\left\|\frac{\partial}{\partial\theta}b\left(\theta_n^+\right)^{/}\right\|^2 \|W_n^*(\theta_n^*)\| > \frac{K}{4}\right) =$
$= P_{\theta_0}\left(\left\|\frac{\partial}{\partial\theta}b^{/}\left(\theta_n^+\right)\right\|^2 \|W_n^*(\theta_n^*)\| > \frac{K}{4} \cap \|W_n^*(\theta_n^*) - W^*(\theta_0)\| > \varepsilon\right)$
$+ P_{\theta_0}\left(\left\|\frac{\partial}{\partial\theta}b^{/}\left(\theta_n^+\right)\right\|^2 \|W_n^*(\theta_n^*)\| > \frac{K}{4} \cap \|W_n^*(\theta_n^*) - W^*(\theta_0)\| \leq \varepsilon\right)$
(for any $\varepsilon > 0$)

$\leq P_{\theta_0}\left(\left\|W_n^*\left(\theta_n^*\right)-W^*\left(\theta_0\right)\right\|>\varepsilon\right)+$
$P_{\theta_0}\left(\left\|\frac{\partial}{\partial\theta}b\left(\theta_n^+\right)^{/}\right\|^2\left\|W_n^*\left(\theta_n^*\right)\right\|>\frac{K}{4}\cap\left\|W_n^*\left(\theta_n^*\right)\right\|-\left\|W^*\left(\theta_0\right)\right\|\leq\varepsilon\right)\leq$
$\leq o\left(n^{-a}\right)+P_{\theta_0}\left(\left\|\frac{\partial}{\partial\theta}b\left(\theta_n^+\right)^{/}\right\|^2>\frac{K}{4(\varepsilon+\left\|W^*(\theta_0)\right\|)}\right)=o\left(n^{-a}\right)+P_{\theta_0}\left(\left\|\frac{\partial}{\partial\theta}b\left(\theta_n^+\right)^{/}\right\|^2>K^*\right)$
(where $K^*=\frac{K}{4(\varepsilon+\left\|W^*(\theta_0)\right\|)}$)
$=o\left(n^{-a}\right)+P_{\theta_0}\left(\left\|\frac{\partial}{\partial\theta}b\left(\theta_n^+\right)^{/}\right\|^2>K^*\cap\left\|\theta_n^+-\theta_0\right\|>\varepsilon^*\right)$
$+P_{\theta_0}\left(\left\|\frac{\partial}{\partial\theta}b\left(\theta_n^+\right)^{/}\right\|^2>K^*\cap\left\|\theta_n^+-\theta_0\right\|\leq\varepsilon^*\right)$ (for any $\varepsilon^*>0$)
$\leq o\left(n^{-a}\right)+P_{\theta_0}\left(\left\|\theta_n^+-\theta_0\right\|>\varepsilon^*\right)+$
$P_{\theta_0}\left(\left\|\frac{\partial}{\partial\theta}b^{/}\left(\theta_n^+\right)\right\|^2>K^*\cap\left\|\theta_n^+-\theta_0\right\|\leq\varepsilon^*\right)=$
$o\left(n^{-a}\right)+P_{\theta_0}\left(\left\|\frac{\partial}{\partial\theta}b\left(\theta_n^+\right)^{/}\right\|^2>K^*\cap\left\|\theta_n^+-\theta_0\right\|\leq\varepsilon^*\right)$ (from consistency of θ_n^+). Now as $\left\|\theta_n^+-\theta_0\right\|\leq\varepsilon^*$ and choosing $\varepsilon^*\leq\varepsilon_4$ we have that $\theta_n^+\in\mathcal{O}_\varepsilon\left(\theta_0\right)$ with probability $1-o\left(n^{-a}\right)$, due to assumption A.10. Hence by choosing $K^*\geq M_1^*$ we have that $P_{\theta_0}\left(\left\|\frac{\partial}{\partial\theta}b^{/}\left(\theta_n^+\right)\right\|^2>K^*\cap\left\|\theta_n^+-\theta_0\right\|\leq\varepsilon^*\right)=o\left(n^{-a}\right)$.

For B_2 we need to prove that $\exists K^*>0$ such that
$P_{\theta_0}\left(\left\|\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}b\left(\theta_n^+\right)^{/}W_n^*\left(\theta_n^*\right)\left(\beta_n-b\left(\theta_n^+\right)\right)\right]_{i,j=1,\ldots,p}\right\|>\frac{K^*}{2}\right)=o\left(n^{-a}\right)$. But
$P_{\theta_0}\left(\left\|\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}b\left(\theta_n^+\right)^{/}W_n^*\left(\theta_n^*\right)\left(\beta_n-b\left(\theta_n^+\right)\right)\right]_{i,j=1,\ldots,p}\right\|>\frac{K^*}{2}\right)\leq$
$P_{\theta_0}\left(\sum_{i=1,\ldots,p}\sum_{j=1,\ldots,p}\left|\frac{\partial^2}{\partial\theta_i\partial\theta_j}b\left(\theta_n^+\right)^{/}W_n^*\left(\theta_n^*\right)\left(\beta_n-b\left(\theta_n^+\right)\right)\right|>\frac{K^*}{2}\right)$
$\leq P_{\theta_0}\left(\max_{i,j}\left|\frac{\partial^2}{\partial\theta_i\partial\theta_j}b\left(\theta_n^+\right)^{/}W_n^*\left(\theta_n^*\right)\left(\beta_n-b\left(\theta_n^+\right)\right)\right|>\frac{K^*}{4p^2}\right)$. Hence it suffices $P_{\theta_0}\left(\left|\frac{\partial^2}{\partial\theta_i\partial\theta_j}b\left(\theta_n^+\right)^{/}W_n^*\left(\theta_n^*\right)\left(\beta_n-b\left(\theta_n^+\right)\right)\right|>\frac{K^*}{4p^2}\right)=o\left(n^{-a}\right)$ for some specific i,j. In fact we can prove that $\forall\varepsilon>0$ we have that
$P_{\theta_0}\left(\left|\frac{\partial^2}{\partial\theta_i\partial\theta_j}b\left(\theta_n^+\right)^{/}W_n^*\left(\theta_n^*\right)\left(\beta_n-b\left(\theta_n^+\right)\right)\right|>\varepsilon\right)=o\left(n^{-a}\right)$.
This, again, follows first, as for $\varepsilon^{**}>0$ $P\left(\left\|\beta_n-b\left(\theta_n^+\right)\right\|>\varepsilon^{**}\right)$
$=P\left(\left\|\beta_n-b\left(\theta_0\right)+b\left(\theta_0\right)-b\left(\theta_n^+\right)\right\|>\varepsilon^{**}\right)$
$\leq P\left(\left\|\beta_n-b\left(\theta_0\right)\right\|>\frac{\varepsilon^{**}}{2}\right)+P\left(\left\|b\left(\theta_0\right)-b\left(\theta_n^+\right)\right\|>\frac{\varepsilon^{**}}{2}\right)=o\left(n^{-a}\right)$ as the first probability is $o\left(n^{-a}\right)$ due to Lemma 2.1 and the second is also $o\left(n^{-a}\right)$ due to assumption A.3 and the consistency of θ_n^+. Second, for $\varepsilon^{***}>0$ $P\left(\left\|W_n^*\left(\theta_n^*\right)-W^*\left(\theta_0\right)\right\|>\varepsilon^{***}\right)=o\left(n^{-a}\right)$ from Lemma AL.1 and finally, for $\varepsilon^{****}>0$ $P_{\theta_0}\left(\left|\frac{\partial^2}{\partial\theta_i\partial\theta_j}b\left(\theta_n^+\right)-\frac{\partial^2}{\partial\theta_i\partial\theta_j}b\left(\theta_0\right)\right|>\varepsilon^{****}\right)=o\left(n^{-a}\right)$ from assumption A.4 and the result follows. ∎

GMR 2 Estimator: $o(n^{-a})$-Consistency and $o(n^{-a})$-Tightness The relevant results for the GMR 2 estimator are presented here.

Lemma 2.3 *Under the validity of Lemma 2.1 and assumption A.10 we have that*

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

Proof. Due to the definition of θ_n and by Lemma AL.1 we have that $P_{\theta_0}\left(\|W_n^*(x,\theta_n^*)-W^*(\theta_0)\|>\varepsilon\right)=o(n^{-a}), \forall\varepsilon>0$. Now notice that $p\lim\left(\beta_n-E_\theta\beta_n\right)=b(\theta_0)-b(\theta)$. Hence $\forall\theta\in\Theta$ and $\forall\varepsilon>0$
$P_{\theta_0}\left(\sup_{\theta\in\Theta}\|\beta_n-E_\theta\beta_n-[b(\theta_0)-b(\theta)]\|>\varepsilon\right)=$
$P_{\theta_0}\left(\sup_{\theta\in\Theta}\|\beta_n-b(\theta_0)-E_\theta\beta_n+b(\theta)\|>\varepsilon\right)\le$
$P_{\theta_0}\left(\|\beta_n-b(\theta_0)\|+\sup_{\theta\in\Theta}\|E_\theta\beta_n-b(\theta)\|>\varepsilon\right)$.
Now we know from Lemma 2.1 above that $P_{\theta_0}\left(\|\beta_n-b(\theta_0)\|>\frac{\varepsilon}{2}\right)=o(n^{-a})$.
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 $P_{\theta_0}\left(\|\beta_n-b(\theta_0)\|>\varepsilon\right)=o(n^{-a})$ and θ_0 is arbirtrary. Hence, $P_\theta\left(\|\beta_n-b(\theta)\|>\varepsilon\right)=o(n^{-a})$ for any $\theta\in\Theta$. Furthermore, as B is bounded the series $\beta_n-b(\theta)$ is uniformly integrable, and as $\|E_\theta\beta_n-b(\theta)\|\le E_\theta\|\beta_n-b(\theta)\|$ 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 $\|(E_\theta\beta_n-b(\theta))-(E_{\theta^*}\beta_n-b(\theta^*))\|\le\|E_\theta\beta_n-E_{\theta^*}\beta_n\|+\|b(\theta)-b(\theta^*)\|$.
But $\|b(\theta)-b(\theta^*)\|\le k\|\theta-\theta^*\|$ by assumption A.7. Further, $\|E_\theta\beta_n-E_{\theta^*}\beta_n\|=$
$\|(E_\theta\beta_n-b(\theta))-(E_{\theta^*}\beta_n-b(\theta))\|=$
$=\left\|\int_{\mathbb{R}^n}(\beta_n-b(\theta))\,dP_\theta-\int_{\mathbb{R}^n}(\beta_n-b(\theta))\,dP_{\theta^*}\right\|\le$
$\dim(B)\max_{i=1,\ldots,\dim(B)}\left|\int_{\mathbb{R}^n}(\beta_n-b(\theta))_i\,dP_\theta-\int_{\mathbb{R}^n}(\beta_n-b(\theta))_i\,dP_{\theta^*}\right|$
$\le\dim(B)\max_{i=1,\ldots,\dim(B)}\int_{\mathbb{R}^n}|(\beta_n-b(\theta))_i|\,|dP_\theta-dP_{\theta^*}|\le$
$\dim(B)\,M_1\int_{\mathbb{R}^n}|dP_\theta-dP_{\theta^*}|=\dim(B)\,M_1TVD(P_\theta,P_{\theta^*})\le\dim(B)\,M_1C\|\theta-\theta^*\|$
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 $\|(E_\theta\beta_n-b(\theta))-(E_{\theta^*}\beta_n-b(\theta^*))\|\le[k+\dim(B)\,M_1C]\,\|\theta-\theta^*\|$

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

and consequently, $P_{\theta_0}\left(\sup_{\theta\in\Theta}\left\|\beta_n - E_\theta\beta_n - \left[b\left(\theta_0\right) - b\left(\theta\right)\right]\right\| > \varepsilon\right) = o\left(n^{-a}\right)$, and the $o\left(n^{-a}\right)$ consistency follows.
Now, let us call $J_n\left(\theta\right) = \left(\beta_n - E_\theta\beta_n\right)^{/} W^*\left(x,\theta_n^*\right)\left(\beta_n - E_\theta\beta_n\right)$ then $\theta_n - \theta_0 = -\left(\frac{\partial^2}{\partial\theta\partial\theta^{/}} J_n\left(\theta_n^+\right)\right)^{-1} \frac{\partial}{\partial\theta} J_n\left(\theta_0\right)$ with probability $1 - o\left(n^{-a}\right)$ where $\frac{\partial}{\partial\theta} J_n\left(\theta_0\right) = \frac{\partial}{\partial\theta} J_n\left(\theta\right)\Big|_{\theta=\theta_0}$ and θ_n^+ lies between θ_n and θ_0 and may be different across rows.
Hence it suffices to first show that $P\left(\left\|\frac{\partial}{\partial\theta} J_n\left(\theta_0\right)\right\| > C\frac{\ln^{1/2} n}{n^{1/2}}\right) = o\left(n^{-a}\right)$ and second that for some $K > 0$ $P\left(\left\|\left(\frac{\partial^2}{\partial\theta\partial\theta^{/}} J_n\left(\theta_n^+\right)\right)^{-1}\right\| > K\right) = o\left(n^{-a}\right)$ and apply Lemma 5 of [2].
Now $\frac{\partial}{\partial\theta} J_n\left(\theta\right)\Big|_{\theta=\theta_0} = -2\frac{\partial E_\theta \beta_n^{/}}{\partial\theta} W_n^*\left(x,\theta_n^*\right)\left(\beta_n - E_\theta\beta_n\right)\Big|_{\theta=\theta_0}$ and we know from Lemma 2.1 above that $P_{\theta_0}\left(\left\|\beta_n - b\left(\theta_0\right)\right\| > C_1\frac{\ln^{1/2} n}{n^{1/2}}\right) = o\left(n^{-a}\right)$ for some $C_1 > 0$. Hence $P_{\theta_0}\left(\left\|\beta_n - E_{\theta_0}\beta_n\right\| > C_1\frac{\ln^{1/2} n}{n^{1/2}}\right) =$
$P_{\theta_0}\left(\left\|\beta_n - b\left(\theta_0\right) - \left(E_{\theta_0}\beta_n - b\left(\theta_0\right)\right)\right\| > C_1\frac{\ln^{1/2} n}{n^{1/2}}\right) \le$
$P_{\theta_0}\left(\left\|\beta_n - b\left(\theta_0\right)\right\| + \left\|E_{\theta_0}\beta_n - b\left(\theta_0\right)\right\| > C_1\frac{\ln^{1/2} n}{n^{1/2}}\right)$. Now

$$P_{\theta_0}\left(\left\|\beta_n - E_{\theta_0}\beta_n\right\| > C_1\frac{\ln^{1/2} n}{n^{1/2}}\right) = o(n^{-a}),$$

as for $a > 0$, we have that $\left\|E_{\theta_0}\beta_n - b\left(\theta_0\right)\right\| \le E_{\theta_0}\left\|\beta_n - b\left(\theta_0\right)\right\| =$
$E_{\theta_0}\left[\left\|\beta_n - b\left(\theta_0\right)\right\| I\left(\left\|\beta_n - b\left(\theta_0\right)\right\| > C_3\frac{\ln^{1/2} n}{n^{1/2}}\right)\right]$
$+ E_{\theta_0}\left[\left\|\beta_n - b\left(\theta_0\right)\right\| I\left(\left\|\beta_n - b\left(\theta_0\right)\right\| \le C_3\frac{\ln^{1/2} n}{n^{1/2}}\right)\right] \le$
$\le BE_{\theta_0}\left[I\left(\left\|\beta_n - b\left(\theta_0\right)\right\| > C_3\frac{\ln^{1/2} n}{n^{1/2}}\right)\right] + C_3\frac{\ln^{1/2} n}{n^{1/2}} E_{\theta_0}\left[I\left(\left\|\beta_n - b\left(\theta_0\right)\right\| \le C_3\frac{\ln^{1/2} n}{n^{1/2}}\right)\right]$
(where B is the bound of $\left\|\beta_n - b\left(\theta_0\right)\right\|$, see assumption A.2)
$= BP_{\theta_0}\left(\left\|\beta_n - b\left(\theta_0\right)\right\| > C_3\frac{\ln^{1/2} n}{n^{1/2}}\right) + C_3\frac{\ln^{1/2} n}{n^{1/2}} P_{\theta_0}\left(\left\|\beta_n - b\left(\theta_0\right)\right\| \le C_3\frac{\ln^{1/2} n}{n^{1/2}}\right) =$
$Bo\left(n^{-a}\right) + C_3\frac{\ln^{1/2} n}{n^{1/2}}\left(1 - o\left(n^{-a}\right)\right) = o\left(n^{-a}\right) + C_3\frac{\ln^{1/2} n}{n^{1/2}} = O\left(\frac{\ln^{1/2} n}{n^{1/2}}\right)$. In this case we have that $P_{\theta_0}\left(\left\|\beta_n - b\left(\theta_0\right)\right\| + \left\|E_{\theta_0}\beta_n - b\left(\theta_0\right)\right\| > C_1\frac{\ln^{1/2} n}{n^{1/2}}\right) \le$
$P_{\theta_0}\left(\left\|\beta_n - b\left(\theta_0\right)\right\| + o\left(\frac{\ln^{1/2} n}{n^{1/2}}\right) > C_1\frac{\ln^{1/2} n}{n^{1/2}}\right) \le P_{\theta_0}\left(\left\|\beta_n - b\left(\theta_0\right)\right\| > C_4\frac{\ln^{1/2} n}{n^{1/2}}\right)$
for some $C_4 > 0$ and we know that this probability is $o\left(n^{-a}\right)$ and the result follows. For $a = 0$ we have that the GMR2 is asymptotically equivalent to GMR1 (Gourieroux et al. 1993).
Further, due to assumption A.10 it follows that

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

For the second, $\exists K>0$ such that $P\left(\left\|\left(\frac{\partial^2}{\partial\theta\partial\theta^{/}}J_n\left(\theta_n^+\right)\right)^{-1}\right\|>K\right)=o\left(n^{-a}\right)$,

notice that $\frac{\partial^2}{\partial\theta\partial\theta^{/}}J_n\left(\theta_n^+\right)=2\frac{\partial E_{\theta_n^+}\beta_n^{/}}{\partial\theta}W_n^*\left(x,\theta_n^*\right)\frac{\partial E_{\theta_n^+}\beta_n}{\partial\theta^{/}}$
$-2\left[\frac{\partial E_{\theta_n^+}\beta_n^{/}}{\partial\theta_i\partial\theta_j}W_n^*\left(x,\theta_n^*\right)\left(\beta_n-E_{\theta_n^+}\beta_n\right)\right]_{i,j=1,\ldots,p}=A+B$. It suffices to show that for $K^*>0$ we have that $P_{\theta_0}\left(\|A+B\|>K^*\right)=o\left(n^{-a}\right)$. But $P_{\theta_0}\left(\|A+B\|>K^*\right)\leq P_{\theta_0}\left(\|A\|>\frac{K^*}{2}\right)+P_{\theta_0}\left(\|B\|>\frac{K^*}{2}\right)$.

Now $P_{\theta_0}\left(\|A\|>\frac{K^*}{2}\right)=P_{\theta_0}\left(\left\|\frac{\partial E_{\theta_n^+}\beta_n^{/}}{\partial\theta}W_n\left(x,\theta_n^*\right)\frac{\partial E_{\theta_n^+}\beta_n}{\partial\theta^{/}}\right\|>\frac{K^*}{4}\right)\leq$

$P_{\theta_0}\left(\left\|\frac{\partial E_{\theta_n^+}\beta_n^{/}}{\partial\theta}\right\|^2\left\|W_n\left(x,\theta_n^*\right)\right\|>\frac{K^*}{4}\right)=$

$=P_{\theta_0}\left(\left\|\frac{\partial E_{\theta_n^+}\beta_n^{/}}{\partial\theta}\right\|^2\left\|W_n\left(x,\theta_n^*\right)\right\|>\frac{K^*}{4}\cap\left\|W_n^*\left(x,\theta_n^*\right)-W^*\left(\theta_0\right)\right\|>\varepsilon\right)$

$+P_{\theta_0}\left(\left\|\frac{\partial E_{\theta_n^+}\beta_n^{/}}{\partial\theta}\right\|^2\left\|W_n\left(x,\theta_n^*\right)\right\|>\frac{K^*}{4}\cap\left\|W_n^*\left(x,\theta_n^*\right)-W^*\left(\theta_0\right)\right\|\leq\varepsilon\right)\leq$ (for any $\varepsilon>0$)

$\leq P_{\theta_0}\left(\left\|W_n^*\left(x,\theta_n^*\right)-W^*\left(\theta_0\right)\right\|>\varepsilon\right)+$

$P_{\theta_0}\left(\left\|\frac{\partial E_{\theta_n^+}\beta_n^{/}}{\partial\theta}\right\|^2\left\|W_n\left(x,\theta_n^*\right)\right\|>\frac{K^*}{4}\cap\left\|W_n^*\left(x,\theta_n^*\right)\right\|-\left\|W^*\left(\theta_0\right)\right\|\leq\varepsilon\right)\leq$

$\leq o\left(n^{-a}\right)+P_{\theta_0}\left(\left\|\frac{\partial E_{\theta_n^+}\beta_n^{/}}{\partial\theta}\right\|^2>\frac{K^*}{4\left(\varepsilon+\left\|W^*\left(\theta_0\right)\right\|\right)}\right)=o\left(n^{-a}\right)+P_{\theta_0}\left(\left\|\frac{\partial E_{\theta_n^+}\beta_n^{/}}{\partial\theta}\right\|^2>K^{**}\right)=$

(where $K^{**}=\frac{K^*}{4\left(\varepsilon+\left\|W^*\left(\theta_0\right)\right\|\right)}$)

$=o\left(n^{-a}\right)+P_{\theta_0}\left(\left\|\frac{\partial E_{\theta_n^+}\beta_n^{/}}{\partial\theta}\right\|^2>K^{**}\cap\left\|\theta_n^+-\theta_0\right\|>\varepsilon^*\right)$

$+P_{\theta_0}\left(\left\|\frac{\partial E_{\theta_n^+}\beta_n^{/}}{\partial\theta}\right\|^2>K^{**}\cap\left\|\theta_n^+-\theta_0\right\|\leq\varepsilon^*\right)\leq$ (for any $\varepsilon^*>0$)

$\leq o\left(n^{-a}\right)+P_{\theta_0}\left(\left\|\theta_n^+-\theta_0\right\|>\varepsilon^*\right)+$

$P_{\theta_0}\left(\left\|\frac{\partial E_{\theta_n^+}\beta_n^{/}}{\partial\theta}\right\|^2>K^{**}\cap\left\|\theta_n^+-\theta_0\right\|\leq\varepsilon^*\right)=$

$o\left(n^{-a}\right)+P_{\theta_0}\left(\left\|\frac{\partial E_{\theta_n^+}\beta_n^{/}}{\partial\theta}\right\|^2>K^{**}\cap\left\|\theta_n^+-\theta_0\right\|\leq\varepsilon^*\right)$ (from consistency of θ_n^+)

. Now as $\left\|\theta_n^+-\theta_0\right\|\leq\varepsilon^*$ and choosing $\varepsilon^*\leq\varepsilon_4$ we have that $\theta_n^+\in\mathcal{O}_\varepsilon\left(\theta_0\right)$ with probability $1-o\left(n^{-a}\right)$, due to assumption A.10. Hence by choosing $K^{**}\geq$

M_1^* we have that $P_{\theta_0}\left(\left\|\frac{\partial E_{\theta_n^+}\beta_n^{/}}{\partial\theta}\right\|^2 > K^{**}\cap\left\|\theta_n^+-\theta_0\right\|\leq\varepsilon^*\right) = o\left(n^{-a}\right)$.
For B we need to prove that $\exists K^*>0$ such that
$P_{\theta_0}\left(\left\|\left[\frac{\partial^2 E_{\theta_n^+}\beta_n^{/}}{\partial\theta_i\partial\theta_j}W_n^*\left(x,\theta_n^*\right)\left(\beta_n-E_{\theta_n^+}\beta_n\right)\right]_{i,j=1,\ldots,p}\right\| > \frac{K^*}{2}\right) = o\left(n^{-a}\right)$. In fact we can prove that $\forall\varepsilon>0$ we have that
$P_{\theta_0}\left(\left\|\left[\frac{\partial^2 E_{\theta_n^+}\beta_n^{/}}{\partial\theta_i\partial\theta_j}W_n^*\left(x,\theta_n^*\right)\left(\beta_n-E_{\theta_n^+}\beta_n\right)\right]_{i,j=1,\ldots,p}\right\| > \varepsilon\right) = o\left(n^{-a}\right)$. This, again, follows from assumption A.10 and the $o\left(n^{-a}\right)$ consistency of θ_n^+ and $P\left(\left\|\beta_n-E_{\theta_n^+}\beta_n\right\|>\varepsilon^{**}\right) = o\left(n^{-a}\right)$, $\forall\varepsilon^{**}>0$. ■

GT Estimator: $o\left(n^{-a}\right)$-Consistency and $o\left(n^{-a}\right)$-Tightness The relevant results for the GT estimator are presented here.

Lemma 2.4 *Under the validity of Lemma 2.1 we have that*

$$P_{\theta_0}\left(\left\|\theta_n-\theta_0\right\| > C_3\frac{\ln^{1/2}n}{n^{1/2}}\right) = o\left(n^{-a}\right)\ \textit{for some}\ C_3>0$$

where θ_n is the GT estimator.

Proof. For notational convenience we set $c_n\left(\beta\right)=\frac{1}{n}\sum c\left(x_i,\beta\right)$ and denote by $E_\theta\left(c_n\left(\beta_n\right)\right)$, the quantity $E_\theta\left(c_n\left(\beta\right)\right)|_{\beta=\beta_n}$. The definition of θ_n is: $\theta_n = \arg\min_{\theta\in\Theta}J_n\left(\theta\right) = \arg\min_{\theta\in\Theta}\left(E_\theta\left(c_n\left(\beta_n\right)\right)\right)^{/}W_n^{**}\left(\theta_n^*\right)E_\theta\left(c_n\left(\beta_n\right)\right)$ and we have that $W_n^{**}\left(\theta_n^*\right)=\frac{1}{n}\sum_i W^{**}\left(x_i,\theta_n^*\right)$, where θ_n^* as in assumption A.13, and by Lemma AL.1 we have that $P_{\theta_0}\left(\left\|W_n^{**}\left(x,\theta_n^*\right)-W^{**}\left(\theta_0\right)\right\|>\varepsilon\right)=o\left(n^{-a}\right)$, $\forall\varepsilon>0$. Further,

$$P_{\theta_0}\left(\sup_\theta\left\|E_\theta c_n\left(\beta_n\right)-E_\theta c_n\left(b\left(\theta_0\right)\right)\right\|>\varepsilon\right) = o\left(n^{-a}\right)\ \forall\varepsilon>0 \qquad (1)$$

as $P_{\theta_0}\left(\sup_\theta\left\|E_\theta c_n\left(\beta_n\right)-E_\theta c_n\left(b\left(\theta_0\right)\right)\right\|>\varepsilon\right) \leq P_{\theta_0}\left(\sup_\theta\left(E_\theta u_c\right)\left\|\beta_n-b\left(\theta_0\right)\right\|>\varepsilon\right) = o\left(n^{-a}\right)$ due to Lemma 2.1 above and by assumption A.8. (Notice that $E_\theta c_n\left(\beta_n\right) = E_\theta c_n\left(\beta\right)|_{\beta=\beta_n}$ and consequently $\left\|E_\theta c_n\left(\beta_n\right)-E_\theta c_n\left(b\left(\theta_0\right)\right)\right\| \leq E_\theta\left\|c_n\left(\beta_n\right)-c_n\left(b\left(\theta_0\right)\right)\right\| \leq \left(E_\theta u_c\right)\left\|\beta_1-\beta_2\right\||_{\substack{\beta_1=\beta_n\\ \beta_2=b(\theta_0)}}$). Consequently, the consistency of θ_n follows from Lemma 5 of [2].
Hence θ_n is in the interior of Θ and $\frac{\partial}{\partial\theta}J_n\left(\theta_n\right)=0$ with probability $1-o\left(n^{-a}\right)$. Hence element by element mean value expansions of $\frac{\partial}{\partial\theta}J_n\left(\theta_n\right)$ around θ_0 and rearrangement gives: $\theta_n-\theta_0 = -\left(\frac{\partial^2}{\partial\theta\partial\theta^{/}}J_n\left(\theta_n^+\right)\right)^{-1}\frac{\partial}{\partial\theta}J_n\left(\theta_0\right)$ 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 1^{st}) $P_{\theta_0}\left(\left\|\frac{\partial}{\partial\theta}J_n(\theta_0)\right\| > C\frac{\ln^{1/2}n}{n^{1/2}}\right) = o(n^{-a})$ and 2^{nd}) $P\left(\left\|\left(\frac{\partial^2}{\partial\theta\partial\theta^{/}}J_n\left(\theta_n^+\right)\right)^{-1}\right\| > K\right) = o(n^{-a})$.

For 1^{st}) notice that $P_{\theta_0}\left(\left\|\frac{\partial}{\partial\theta}J_n(\theta_0)\right\| > C^*\frac{\ln^{1/2}n}{n^{1/2}}\right) =$
$P_{\theta_0}\left(\left\|2\frac{\partial}{\partial\theta}E_{\theta_0}\left[c^{/}(\beta_n)\right]W_n^{**}(\theta_n^*)E_{\theta_0}\left[c(\beta_n)\right]\right\| > C^*\frac{\ln^{1/2}n}{n^{1/2}}\right) \leq$
$P_{\theta_0}\left(\left\|\frac{\partial}{\partial\theta}E_{\theta_0}c^{/}(\beta_n)\right\|\left\|W_n^{**}(\theta_n^*)\right\|\left\|E_{\theta_0}\left[c(\beta_n)\right]\right\| > C\frac{\ln^{1/2}n}{n^{1/2}}\right)$ where $C = \frac{C^*}{2}$ by the submultiplicative property of the norm. Hence it suffices to show that 1_i) $P_{\theta_0}\left(\left\|E_{\theta_0}\left[c(\beta_n)\right]\right\| > C\frac{\ln^{1/2}n}{n^{1/2}}\right) = o(n^{-a})$ 1_{ii}) for some $K > 0$ we have that $P_{\theta_0}\left(\left\|\frac{\partial}{\partial\theta}E_{\theta_0}\left[c^{/}(\beta_n)\right]\right\| > K\right) = o(n^{-a})$, and 1_{iii}) for $K^* > 0$ we have $P_{\theta_0}\left(\left\|W_n^{**}(\theta_n^*)\right\| > K^*\right) = o(n^{-a})$, which is true from Lemma AL.1. For 1_i) notice that $P_{\theta_0}\left(\left\|E_{\theta_0}\left[c(\beta_n)\right]\right\| > C\frac{\ln^{1/2}n}{n^{1/2}}\right) = P_{\theta_0}\left(\left\|E_{\theta_0}\left[c(\beta_n)\right] - E_{\theta_0}\left[c(b(\theta_0))\right]\right\| > C\frac{\ln^{1/2}n}{n^{1/2}}\right)$ (as $E_{\theta_0}\left[c(b(\theta_0))\right] = 0$) $\leq P_{\theta_0}\left(E_{\theta_0}\left\|c(\beta_n) - c(b(\theta_0))\right\| > C\frac{\ln^{1/2}n}{n^{1/2}}\right)$
$\leq P_{\theta_0}\left(\left[E_{\theta_0}\left\|u_c(x_i)\right\|\right]\left\|\beta_n - b(\theta_0)\right\| > C\frac{\ln^{1/2}n}{n^{1/2}}\right) = P_{\theta_0}\left(\left\|\beta_n - b(\theta_0)\right\| > \frac{C}{E_{\theta_0}\|u_c(x_i)\|}\frac{\ln^{1/2}n}{n^{1/2}}\right) =$
$o(n^{-a})$ by Lemma 2.1 above. Finally, for 1_{ii}) it suffices to show that $\forall\varepsilon > 0$ we have that $P_{\theta_0}\left(\left\|\frac{\partial}{\partial\theta}E_{\theta_0}\left[c^{/}(\beta_n)\right] - \frac{\partial}{\partial\theta}E_{\theta_0}\left[c^{/}(b(\theta_0))\right]\right\| > \varepsilon\right) = o(n^{-a})$. Now by assumption A.8 we have that $P_{\theta_0}\left(\left\|\frac{\partial}{\partial\theta}E_{\theta_0}\left[c^{/}(\beta_n)\right] - \frac{\partial}{\partial\theta}E_{\theta_0}\left[c^{/}(b(\theta_0))\right]\right\| > \varepsilon\right) \leq$
$P_{\theta_0}\left(M^*\left\|E_{\theta_0}\left[c(\beta_n)\right] - E_{\theta_0}\left[c(b(\theta_0))\right]\right\| > \varepsilon\right) = o(n^{-a})$ from 1_i) above.

For 2^{nd}), i.e. $P\left(\left\|\left(\frac{\partial^2}{\partial\theta\partial\theta^{/}}J_n\left(\theta_n^+\right)\right)^{-1}\right\| > K\right) = o(n^{-a})$ notice that $\frac{\partial^2}{\partial\theta\partial\theta^{/}}J_n\left(\theta_n^+\right) =$
$2\left\{\frac{\partial^2}{\partial\theta_i\partial\theta_j}E_{\theta_n^+}\left[c^{/}(\beta_n)\right]W_n^{**}(\theta_n^*)E_{\theta_n^+}\left[c(\beta_n)\right]\right\}_{i,j=1,\ldots,p} + 2\frac{\partial}{\partial\theta}E_{\theta_n^+}\left[c^{/}(\beta_n)\right]W_n^{**}(\theta_n^*)\frac{\partial}{\partial\theta^{/}}E_{\theta_n^+}\left[c(\beta_n)\right]$.

We show that $\exists K^* > 0$ such that $P\left(\left\|\frac{\partial^2}{\partial\theta\partial\theta^{/}}J_n\left(\theta_n^+\right)\right\| > K^*\right) = o(n^{-a})$. But
$P\left(\left\|\frac{\partial^2}{\partial\theta\partial\theta^{/}}J_n\left(\theta_n^+\right)\right\| > K^*\right) \leq P\left(\left\|2\frac{\partial}{\partial\theta}E_{\theta_n^+}\left[c^{/}(\beta_n)\right]W_n^{**}(\theta_n^*)\frac{\partial}{\partial\theta^{/}}E_{\theta_n^+}\left[c(\beta_n)\right]\right\| > \frac{K^*}{2}\right) +$
$P\left(\left\|2\left\{\frac{\partial^2}{\partial\theta_i\partial\theta_j}E_{\theta_n^+}\left[c^{/}(\beta_n)\right]W_n^{**}(\theta_n^*)E_{\theta_n^+}\left[c(\beta_n)\right]\right\}_{i,j=1,\ldots,p}\right\| > \frac{K^*}{2}\right)$

Now $P_{\theta_0}\left(\left\|2\frac{\partial}{\partial\theta}E_{\theta_n^+}\left[c^{/}(\beta_n)\right]W_n^{**}(\theta_n^*)\frac{\partial}{\partial\theta^{/}}E_{\theta_n^+}\left[c(\beta_n)\right]\right\| > \frac{K^*}{2}\right) \leq$
$P_{\theta_0}\left(\left\|\frac{\partial}{\partial\theta}E_{\theta_n^+}\left[c^{/}(\beta_n)\right]\right\|^2\left\|W_n^{**}(\theta_n^*)\right\| > \frac{K^*}{4}\right) =$
$= P_{\theta_0}\left(\left\|\frac{\partial}{\partial\theta}E_{\theta_n^+}\left[c^{/}(\beta_n)\right]\right\|^2\left\|W_n^{**}(\theta_n^*)\right\| > \frac{K^*}{4} \cap \left\|W_n^{**}(\theta_n^*) - W^{**}(\theta_0)\right\| > \varepsilon\right) +$
$P_{\theta_0}\left(\left\|\frac{\partial}{\partial\theta}E_{\theta_n^+}\left[c^{/}(\beta_n)\right]\right\|^2\left\|W_n^{**}(\theta_n^*)\right\| > \frac{K^*}{4} \cap \left\|W_n^{**}(\theta_n^*) - W^{**}(\theta_0)\right\| \leq \varepsilon\right)$ (for any $\varepsilon > 0$)
$\leq P_{\theta_0}\left(\left\|W_n^{**}(\theta_n^*) - W^{**}(\theta_0)\right\| > \varepsilon\right) +$
$P_{\theta_0}\left(\left\|\frac{\partial}{\partial\theta}E_{\theta_n^+}\left[c^{/}(\beta_n)\right]\right\|^2\left\|W_n^{**}(\theta_n^*)\right\| > \frac{K^*}{4} \cap \left\|W_n^{**}(\theta_n^*) - W^{**}(\theta_0)\right\| \leq \varepsilon\right)$

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

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

Now for $P_{\theta_0}\left(\left\|\frac{\partial}{\partial\theta}E_{\theta_n^+}\left[c^/\left(\beta_n\right)\right]\right\|>K^{**}\right)=o\left(n^{-a}\right)$ and $\varphi_0=\left(b^/\left(\theta_0\right),\theta_0^/\right)^/$ we have that

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

$P_{\theta_0}\left(\left\|\frac{\partial}{\partial\theta}E_{\theta_n^+}\left[c^/\left(\beta_n\right)\right]\right\|>K^{**}\cap\left\{\begin{pmatrix}\beta_n\\ \theta_n^+\end{pmatrix}\in\mathcal{O}_\eta\left(\varphi_0\right)\right\}\right)+$

$P_{\theta_0}\left(\left\|\frac{\partial}{\partial\theta}E_{\theta_n^+}\left[c^/\left(\beta_n\right)\right]\right\|>K^{**}\cap\left\{\begin{pmatrix}\beta_n\\ \theta_n^+\end{pmatrix}\notin\mathcal{O}_\eta\left(\varphi_0\right)\right\}\right) \leq$

$P_{\theta_0}\left(\sup_{\theta\in\mathcal{O}_\eta\left(\varphi_0\right)}\left\|\frac{\partial}{\partial\theta}E_{\theta_n^+}\left[c^/\left(\beta_n\right)\right]\right\|>K^{**}\right)+$

$P_{\theta_0}\left(\begin{pmatrix}\beta_n\\ \theta_n^+\end{pmatrix}\notin\mathcal{O}_\eta\left(\varphi_0\right)\right) \leq$

$P_{\theta_0}\left(\left\{\beta_n\notin\mathcal{O}_\eta\left(b\left(\theta_0\right)\right)\right\}\cup\left\{\theta_n^+\notin\mathcal{O}_\eta\left(\theta_0\right)\right\}\right) \leq$

$P_{\theta_0}\left(\beta_n\notin\mathcal{O}_\eta\left(b\left(\theta_0\right)\right)\right)+P_{\theta_0}\left(\theta_n\notin\mathcal{O}_\eta\left(\theta_0\right)\right)\leq o\left(n^{-a}\right)$ where K^{**} can be chosen as greater than the maximum between the previous choice and an upper bound of $\sup_{\theta\in\mathcal{O}_\eta\left(\varphi_0\right)}\left\|\frac{\partial}{\partial\theta}E_\theta\left[c^/\left(\beta\right)\right]\right\|$ which exists due to assumption A.8.

Hence $\frac{\partial}{\partial\theta}E_{\theta_n^+}\left[c^/\left(\beta_n\right)\right]W_n^{**}\left(\theta_n^*\right)\frac{\partial}{\partial\theta^/}E_{\theta_n^+}\left[c\left(\beta_n\right)\right]\rightarrow\frac{\partial}{\partial\theta}E_{\theta_0}\left[c^/\left(b\left(\theta_0\right)\right)\right]W^{**}\left(\theta_0\right)\frac{\partial}{\partial\theta^/}E_{\theta_0}\left[c\left(b\left(\theta_0\right)\right)\right]$ with probability $o\left(n^{-a}\right)$ and $\frac{\partial}{\partial\theta}E_{\theta_0}\left[c^/\left(b\left(\theta_0\right)\right)\right]W^{**}\left(\theta_0\right)\frac{\partial}{\partial\theta^/}E_{\theta_0}\left[c\left(b\left(\theta_0\right)\right)\right]$ nonsingular. This follows from the fact that $\forall\theta\in\Theta$ we have that $E_\theta\left[c\left(b\left(\theta\right)\right)\right]=0$ and by the Implicit Function Theorem we have that $\frac{\partial}{\partial\theta}E_\theta\left[c^/\left(b\left(\theta\right)\right)\right]+\frac{\partial}{\partial\theta}b^/\left(\theta\right)\frac{\partial}{\partial\beta}E_\theta\left[c^/\left(b\left(\theta\right)\right)\right]=0$ and it follows that $\frac{\partial}{\partial\theta}E_\theta\left[c^/\left(b\left(\theta\right)\right)\right]=-\frac{\partial}{\partial\theta}b^/\left(\theta\right)\frac{\partial}{\partial\beta}E_\theta\left[c^/\left(b\left(\theta\right)\right)\right]$. Now $\frac{\partial}{\partial\theta}b^/\left(\theta\right)$ is a $p\times q$ matrix and $rank\left(\frac{\partial}{\partial\theta}b^/\left(\theta\right)\right)=p$, by assumption A.7 above, whereas $\frac{\partial}{\partial\beta}E_\theta\left[c^/\left(b\left(\theta\right)\right)\right]$ is an $q\times l$ matrix with $rank\left(\frac{\partial}{\partial\beta}E_\theta\left[c^/\left(b\left(\theta\right)\right)\right]\right)=q$, by assumption A.8 above, hence $rank\left(\frac{\partial}{\partial\theta}b^/\left(\theta\right)\frac{\partial}{\partial\beta}E_\theta\left[c^/\left(b\left(\theta\right)\right)\right]\right)=p$ and it follows that $rank\left(\frac{\partial}{\partial\theta}E_\theta\left[c^/\left(b\left(\theta\right)\right)\right]\right)=p$. It follows that as $W^{**}\left(\theta_0\right)$ is nonsingular, by assumption A.9 above, $rank\left(\frac{\partial}{\partial\theta}E_{\theta_0}\left[c^/\left(b\left(\theta_0\right)\right)\right]W^{**}\left(\theta_0\right)\frac{\partial}{\partial\theta^/}E_{\theta_0}\left[c\left(b\left(\theta_0\right)\right)\right]\right)=p$.

Further we have to prove that

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

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

$P_{\theta_0}\left(\sum_{i=1,\ldots,p}\sum_{j=1,\ldots,p}\left|\frac{\partial^2}{\partial\theta_i\partial\theta_j}E_{\theta_n^+}\left[c^/\left(\beta_n\right)\right]W_n^{**}\left(\theta_n^*\right)E_{\theta_n^+}\left[c\left(\beta_n\right)\right]\right|>\frac{K^*}{4}\right)$

$\leq P_{\theta_0}\left(\max_{i,j}\left|\frac{\partial^2}{\partial\theta_i\partial\theta_j}E_{\theta_n^+}\left[c^/\left(\beta_n\right)\right]W_n^{**}\left(\theta_n^*\right)E_{\theta_n^+}\left[c\left(\beta_n\right)\right]\right|>\frac{K^*}{4p^2}\right)$. In fact we

can prove that $\forall\varepsilon>0$ we have that
$P_{\theta_0}\left(\left|\frac{\partial^2}{\partial\theta_i\partial\theta_j}E_{\theta_n^+}\left[c^{/}(\beta_n)\right]W_n^{**}(\theta_n^*)E_{\theta_n^+}\left[c(\beta_n)\right]\right|>\varepsilon\right)=o\left(n^{-a}\right)$. But
$P_{\theta_0}\left(\left|\frac{\partial^2}{\partial\theta_i\partial\theta_j}E_{\theta_n^+}\left[c^{/}(\beta_n)\right]W_n^{**}(\theta_n^*)E_{\theta_n^+}\left[c(\beta_n)\right]\right|>\varepsilon\right)=$
$P_{\theta_0}\left(\left\|\frac{\partial^2}{\partial\theta_i\partial\theta_j}E_{\theta_n^+}\left[c^{/}(\beta_n)\right]W_n^{**}(\theta_n^*)E_{\theta_n^+}\left[c(\beta_n)\right]\right\|>\varepsilon\right)\leq$
$P_{\theta_0}\left(\left\|\frac{\partial^2}{\partial\theta_i\partial\theta_j}E_{\theta_n^+}\left[c^{/}(\beta_n)\right]\right\|\left\|W_n^{**}(\theta_n^*)\right\|\left\|E_{\theta_n^+}\left[c(\beta_n)\right]\right\|>\varepsilon\right)$
$E_{\theta_n^+}\left[c(\beta_n)\right]\rightarrow E_{\theta_0}\left[c(b(\theta_0))\right]=0$ as $E_\theta\left[c(b(\theta))\right]=0\ \forall\theta\in\mathcal{O}_{\varepsilon_2}(\theta_0)$ due to continuous mapping.
$P_{\theta_0}\left(\left\|\frac{\partial^2}{\partial\theta_i\partial\theta_j}E_{\theta_n^+}\left[c^{/}(\beta_n)\right]\right\|\left\|W_n^{**}(\theta_n^*)\right\|\left\|E_{\theta_n^+}\left[c(\beta_n)\right]\right\|>\varepsilon\right)=$
$P_{\theta_0}\left(\left\|\frac{\partial^2}{\partial\theta_i\partial\theta_j}E_{\theta_n^+}\left[c^{/}(\beta_n)\right]\right\|\left\|W_n^{**}(\theta_n^*)\right\|\left\|E_{\theta_n^+}\left[c(\beta_n)\right]\right\|>\varepsilon\cap\left\|W_n^{**}(\theta_n^*)-W^{**}(\theta_0)\right\|>\varepsilon_1\right)$
$+P_{\theta_0}\left(\left\|\frac{\partial^2}{\partial\theta_i\partial\theta_j}E_{\theta_n^+}\left[c^{/}(\beta_n)\right]\right\|\left\|W_n^{*}(\theta_n^*)\right\|\left\|E_{\theta_n^+}\left[c(\beta_n)\right]\right\|>\varepsilon\cap\left\|W_n^{**}(\theta_n^*)-W^{**}(\theta_0)\right\|\leq\varepsilon_1\right)$
$\leq P_{\theta_0}\left(\left\|W_n^{**}(\theta_n^*)-W^{**}(\theta_0)\right\|>\varepsilon_1\right)$
$+P_{\theta_0}\left(\left\|\frac{\partial^2}{\partial\theta_i\partial\theta_j}E_{\theta_n^+}\left[c^{/}(\beta_n)\right]\right\|\left\|W_n^{**}(\theta_n^*)\right\|\left\|E_{\theta_n^+}\left[c(\beta_n)\right]\right\|>\varepsilon\cap\left\|W_n^{**}(\theta_n^*)\right\|-\left\|W^{**}(\theta_0)\right\|\leq\varepsilon_1\right)=$

$o\left(n^{-a}\right)+$
$P_{\theta_0}\left(\left\|\frac{\partial^2}{\partial\theta_i\partial\theta_j}E_{\theta_n^+}\left[c^{/}(\beta_n)\right]\right\|\left\|W_n^{**}(\theta_n^*)\right\|\left\|E_{\theta_n^+}\left[c(\beta_n)\right]\right\|>\varepsilon\cap\left\|W_n^{**}(\theta_n^*)\right\|\leq\varepsilon_1+\left\|W^{**}(\theta_0)\right\|\right)$
$\leq o\left(n^{-a}\right)+P_{\theta_0}\left(\left\|\frac{\partial^2}{\partial\theta_i\partial\theta_j}E_{\theta_n^+}\left[c^{/}(\beta_n)\right]\right\|\left\|E_{\theta_n^+}\left[c(\beta_n)\right]\right\|>\varepsilon^*\right)$ where $\varepsilon^*=\frac{\varepsilon}{\varepsilon_1+\|W^*(\theta_0)\|}$
To prove that $P_{\theta_0}\left(\left\|\frac{\partial^2}{\partial\theta_i\partial\theta_j}E_{\theta_n^+}\left[c^{/}(\beta_n)\right]\right\|\left\|E_{\theta_n^+}\left[c(\beta_n)\right]\right\|>\varepsilon^*\right)=o\left(n^{-a}\right)$ it suffices to prove that for $K>0$ $P_{\theta_0}\left(\left\|\frac{\partial^2}{\partial\theta_i\partial\theta_j}E_{\theta_n^+}\left[c^{/}(\beta_n)\right]\right\|>K\right)=o\left(n^{-a}\right)$ and for $\varepsilon^{**}>0$ $P_{\theta_0}\left(\left\|E_{\theta_n^+}\left[c(\beta_n)\right]\right\|>\varepsilon^{**}\right)=o\left(n^{-a}\right)$. For the second order derivatives we have that $\forall i,j=1,\ldots,p$
$P_{\theta_0}\left(\left\|\frac{\partial^2}{\partial\theta_i\partial\theta_j}E_{\theta_n^+}\left[c(\beta_n)\right]\right\|>K\right)$ $=$
$P_{\theta_0}\left(\left\|\frac{\partial^2}{\partial\theta_i\partial\theta_j}E_{\theta_n^+}\left[c(\beta_n)\right]\right\|>K\cap\left\{\begin{pmatrix}\beta_n\\ \theta_n^+\end{pmatrix}\in\mathcal{O}_\eta(\varphi_0)\right\}\right)+$
$P_{\theta_0}\left(\left\|\frac{\partial^2}{\partial\theta_i\partial\theta_j}E_{\theta_n^+}\left[c(\beta_n)\right]\right\|>K\cap\left\{\begin{pmatrix}\beta_n\\ \theta_n^+\end{pmatrix}\notin\mathcal{O}_\eta(\varphi_0)\right\}\right)$ $\leq$
$P_{\theta_0}\left(\sup_{\theta\in\mathcal{O}_\eta(\varphi_0)}\left\|\frac{\partial^2}{\partial\theta_i\partial\theta_j}E_{\theta_n^+}\left[c^{/}(\beta_n)\right]\right\|>K\right)+$
$P_{\theta_0}\left(\begin{pmatrix}\beta_n\\ \theta_n^+\end{pmatrix}\notin\mathcal{O}_\eta(\varphi_0)\right)$ $\leq$
$P_{\theta_0}\left(\left\{\beta_n\notin\mathcal{O}_\eta(b(\theta_0))\right\}\cup\left\{\theta_n^+\notin\mathcal{O}_\eta(\theta_0)\right\}\right)$ $\leq$
$P_{\theta_0}\left(\beta_n\notin\mathcal{O}_\eta(b(\theta_0))\right)+P_{\theta_0}\left(\theta_n\notin\mathcal{O}_\eta(\theta_0)\right)\leq o\left(n^{-a}\right)$ where again K can be chosen as greater than the maximum between the previous choice and an upper bound of $\sup_{\theta\in\mathcal{O}_\eta(\varphi_0)}\left\|\frac{\partial^2}{\partial\theta_i\partial\theta_j}E_\theta\left[c(\beta)\right]\right\|$ which exists due to assumption A.8. Further, for $\varepsilon^{**}>0$ (small) $P_{\theta_0}\left(\left\|E_{\theta_n^+}\left[c(\beta_n)\right]\right\|>\varepsilon^{**}\right)=$

$P_{\theta_0}\left(\left\|E_{\theta_n^+}\left[c\left(\beta_n\right)\right]-E_{\theta_0}\left[c\left(b\left(\theta_0\right)\right)\right]\right\|>\varepsilon^{**}\right)=o\left(n^{-a}\right)$ from above.
It follows that $\frac{\partial^2}{\partial\theta_i\partial\theta_j}E_{\theta_n^+}\left[c^{/}\left(\beta_n\right)\right]W_n^{**}\left(\theta_n^*\right)E_{\theta_n^+}\left[c\left(\beta_n\right)\right]\to 0$ with probability $o\left(n^{-a}\right)$.
Hence $\frac{\partial^2}{\partial\theta\partial\theta^{/}}J_n\left(\theta_n^+\right)\to 2\frac{\partial}{\partial\theta}E_{\theta_0}\left[c^{/}\left(b\left(\theta_0\right)\right)\right]W^{**}\left(\theta_0\right)\frac{\partial}{\partial\theta^{/}}E_{\theta_0}\left[c\left(b\left(\theta_0\right)\right)\right]$, a non-singular matrix, with probability $o\left(n^{-a}\right)$. It follows that $\left(\frac{\partial^2}{\partial\theta\partial\theta^{/}}J_n\left(\theta_n^+\right)\right)^{-1}\to$ $\left(2\frac{\partial}{\partial\theta}E_{\theta_0}\left[c^{/}\left(b\left(\theta_0\right)\right)\right]W^{**}\left(\theta_0\right)\frac{\partial}{\partial\theta^{/}}E_{\theta_0}\left[c\left(b\left(\theta_0\right)\right)\right]\right)^{-1}$ with probability $o\left(n^{-a}\right)$ and, for $K>0$, $P\left(\left(\frac{\partial^2}{\partial\theta\partial\theta^{/}}J_n\left(\theta_n^+\right)\right)^{-1}>K\right)=o\left(n^{-a}\right)$.
Consequently, as $P_{\theta_0}\left(\left\|\frac{\partial}{\partial\theta}J_n\left(\theta_0\right)\right\|>C\frac{\ln^{1/2}n}{n^{1/2}}\right)=o\left(n^{-a}\right)$ the result follows by Lemma 5 of [2]. ■

Existence of Edgeworth Expansions of Indirect Estimators

Lemma 2.5 *Under the validity of Lemmas 2.2, 2.3, 2.4 and assumptions A.12-A.13 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 $s=2a+1$.*

Proof. i) For GMR1 we apply lemma AL.2 where $\theta_n=$GMR1, $\varphi_n=\begin{pmatrix}\beta_n\\ \theta_n^*\end{pmatrix}$ and the application is justified by the fact that that provision 1 holds due to 2.1, 2.2, and A.13, 2 follows from A.7, A.9, A.11 and A.13 and 3 follows from lemma 5 of [2] and A.13. Let $S_n=\begin{pmatrix}\frac{1}{n}\sum_{i=1}^n f\left(x_i,b\left(\theta_0\right),\theta_0\right)\\ \beta_n-b\left(\theta_0\right)\end{pmatrix}$ where f is defined in A.11. Denote by S, $\begin{pmatrix}\frac{1}{n}\sum_{i=1}^n Ef\left(x_i,b\left(\theta_0\right),\theta_0\right)\\ \mathbf{0}_{q\times 1}\end{pmatrix}$. By remark R.26 and lemma 2.1 $\sqrt{n}\left(S_n-S\right)$ 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 [3] $\sqrt{n}G\left(S_n\right)$ has an Edgeworth expansion of the same order.
ii) For GT the proof is analogous to (i) apart from the fact that 2.4 has to be evoked instead of 2.2. The only thing different is J_n which obeys the provisions of AL.2 additionally due to assumption A.8.
iii) For GMR2 we apply again lemma AL.2 where $\theta_n=$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 2.1, 2.3, and A.13, 2 follows from A.10, A.9, A.11 and A.13 and 3 follows from lemma 5 of [2] and A.13. Notice that in this case R_n^* is expanded

by $\begin{pmatrix} D^1 E_{\theta_0}\beta_n \\ . \\ . \\ . \\ D^{d-1} E_{\theta_0}\beta_n \end{pmatrix}$. Now. define $S_n^* = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n f\left(x_i, b\left(\theta_0\right), \theta_0\right) \\ \beta_n - E_{\theta_0}\beta_n \\ D^1 E_{\theta_0}\beta_n \\ . \\ . \\ . \\ D^{d-1} E_{\theta_0}\beta_n \end{pmatrix}$ then

$$\sqrt{n}\left(S_n^* - E_{\theta_0} S_n^*\right) = \sqrt{n}\begin{pmatrix} \frac{1}{n}\sum_{i=1}^n \left[f\left(x_i, b\left(\theta_0\right), \theta_0\right) - E_{\theta_0} f\left(x_i, b\left(\theta_0\right), \theta_0\right)\right] \\ \beta_n - E_{\theta_0}\beta_n \\ 0 \\ . \\ . \\ . \\ 0 \end{pmatrix}$$

has an Edgeworth expansion of order $s = 2a+1$. This is justified by assumptions A.11 and A.12 for $\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \left[f\left(x_i, b\left(\theta_0\right), \theta_0\right) - E_{\theta_0} f\left(x_i, b\left(\theta_0\right), \theta_0\right)\right] \right)$ and by Lemma AL.3 of Appendix 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 3.1 below and remarks R.28 and R.29). $S^* = \begin{pmatrix} S \\ 0 \\ . \\ . \\ . \\ 0 \end{pmatrix}$ 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 [3] $\sqrt{n}G\left(S_n^*\right)$ has an Edgeworth expansion of the same order. ■

3 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 the paragraph "Edgeworth and Moment Approximations of Sequences of Distributions" that the validity of the former do not imply the validity of the latter. We provide a general lemma which, utilizes the Edgeworth expansions along with further assumptions the required approximations are validated. These assumptions are uniform integrability ones, 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.14

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

Remark R.28 *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 (1993), Lemma 2).*

Assumption A.15

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

Remark R.29 *In fact if Q_n is the Edgeworth distribution we have that $A=n^a\int_{\mathbb{R}^q\backslash\mathcal{O}_{K(\ln n)^{\epsilon}}(0)}\|x\|\left|dQ_n\right|=n^a\int_{\mathbb{R}^q\backslash\mathcal{O}_{K(\ln n)^{\epsilon}}(0)}\|f_n\left(z\right)\|d\Phi+o\left(1\right)$ where Φ is the multivariate standard normal cumulative distribution function, and as $f_n\left(z\right)$ is a polynomial in z we get: $A-o\left(1\right)\leq n^a\int_{\mathbb{R}^q\backslash\mathcal{O}_{K(\ln n)^{\epsilon}}(0)}\left\|\sum_{i=0}^{2a}n^{-\frac{i}{2}}f_i\left(z\right)\right\|d\Phi\leq\sum_{i=0}^{2a}n^{\frac{2a-i}{2}}\int_{\mathbb{R}^q\backslash\mathcal{O}_{K(\ln n)^{\epsilon}}(0)}\|f_i\left(z\right)\|d\Phi$ where $f_i\left(z\right)$ appropriate polynomials in z. Now $n^{\frac{2a-i}{2}}\int_{\mathbb{R}^q\backslash\mathcal{O}_{K(\ln n)^{\epsilon}}(0)}\|f_i\left(z\right)\|d\Phi\leq Cn^{\frac{2a-i}{2}}\int_{\mathbb{R}^q\backslash\mathcal{O}_{K(\ln n)^{\epsilon}}(0)}\|z\|^{2\lambda_i}d\Phi=Cn^{\frac{2a-i}{2}}\int_{\mathbb{R}^q\backslash\mathcal{O}_{K(\ln n)^{\epsilon}}(0)}\left(\sum_{j=1}^{q}z_j^2\right)^{\lambda_i}d\Phi$. 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, $A-o\left(1\right)\leq Cn^{\frac{2a-i}{2}}\sum_{l=1}^{q^{\lambda_i}}\int_{\mathbb{R}^q\backslash\mathcal{O}_{K(\ln n)^{\epsilon}}(0)}\prod_{j=1}^{q}z_j^{k_{j,l}}d\Phi=$

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

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

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

$Cn^{\frac{2a-i}{2}}\left(\frac{\pi}{2}\right)^{-\frac{q}{2}}\sum_{l=1}^{q^{\lambda_i}}\prod_{j=1}^{q}2^{\frac{k_{j,l}-1}{2}}\Gamma\left(\frac{k_{j,l}-1}{2},\frac{K^2(\ln n)^{2\epsilon}}{2}\right)$ where $\Gamma\left(\bullet,\bullet\right)$ is the incomplete Gamma function (see e.g. [13] 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. [13] formula 8.357). Hence* $A\leq$ $C(\pi)^{-\frac{q}{2}}2^{\frac{3q}{2}}\sum_{l=1}^{q^{\lambda_i}}\prod_{j=1}^{q}(\ln n)^{\epsilon\left(k_{j,l}-3\right)}K^{k_{j,l}-3}\exp\left(\frac{(2a-i)\ln n-K^2(\ln n)^{2\epsilon}}{2}\right)+o(1)$. *Now for* $\epsilon>\frac{1}{2}$, *and* $K>0$ *we have that* $C(\pi)^{-\frac{q}{2}}2^{\frac{3q}{2}}\sum_{l=1}^{q^{\lambda_i}}\prod_{j=1}^{q}(\ln n)^{\epsilon\left(k_{j,l}-3\right)}K^{k_{j,l}-3}\exp\left(\frac{(2a-i)\ln n-(\ln n)^{2\epsilon}}{2}\right)\to 0$ *as* $n\to\infty$.

Lemma 3.1 *Given the assumptions A.14 and A.15 above then*

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

Proof. Assume now that $\sup_{A\in\mathcal{B}_C}|P_n(A)-Q_n(A)|=O(n^{-a-\eta})$, where $\mathcal{B}_C$ denote the collection of convex Borel sets of $\mathbb{R}^q$ and $\eta>0$. Now
$n^a\left\|\int_{\mathbb{R}^q}x\,(dP_n-dQ_n)\right\|=n^a\left\|\int_{B(0,K(\ln n)^\epsilon)}x\,(dP_n-dQ_n)\right\|+n^a\left\|\int_{\mathbb{R}^q\backslash B(0,K(\ln n)^\epsilon)}x\,(dP_n-dQ_n)\right\|$
$\leq n^a\left\|\int_{B(0,K(\ln n)^\epsilon)}x\,(dP_n-dQ_n)\right\|+n^a\left\|\int_{\mathbb{R}^q\backslash B(0,K(\ln n)^\epsilon)}x\,dP_n\right\|+n^a\left\|\int_{\mathbb{R}^q\backslash B(0,K(\ln n)^\epsilon)}x\,dQ_n\right\|$
$\leq n^a\int_{B(0,K(\ln n)^\epsilon)}\|x\|\,|dP_n-dQ_n|+n^a\int_{\mathbb{R}^q\backslash B(0,K(\ln n)^\epsilon)}\|x\|\,dP_n+n^a\int_{\mathbb{R}^q\backslash B(0,K(\ln n)^\epsilon)}\|x\|\,|dQ_n|$
$\leq n^aK(\ln n)^\epsilon\int_{B(0,K(\ln n)^\epsilon)}|dP_n-dQ_n|+n^a\int_{\mathbb{R}^q\backslash B(0,K(\ln n)^\epsilon)}\|x\|\,dP_n+n^a\int_{\mathbb{R}^q\backslash B(0,K(\ln n)^\epsilon)}\|x\|\,|dQ_n|$
$\leq K(\ln n)^\epsilon\sup_{A\in\mathcal{B}_C}n^a|P_n(A)-Q_n(A)|+n^a\int_{\mathbb{R}^q\backslash B(0,K(\ln n)^\epsilon)}\|x\|\,dP_n$
$+\,n^a\int_{\mathbb{R}^q\backslash B(0,K(\ln n)^\epsilon)}\|x\|\,|dQ_n|$
Let P_n be the distribution of $\sqrt{n}(\theta_n-\theta_0)$. 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 $\sqrt{n}(\Theta-\theta_0)$.
$n^a\int_{[\mathbb{R}^q\backslash B(0,K(\ln n)^\epsilon)]\cap\sqrt{n}(\Theta-\theta_0)}\|x\|\,dP_n=n^a\int_{\sqrt{n}(\Theta-\theta_0)\backslash B(0,K(\ln n)^\epsilon)}\|x\|\,dP_n$ 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 $B(0,\rho)\supseteq\Theta-\theta_0$ and ρ exists as Θ is bounded by assumption.
Hence $n^a\left\|\int_{\mathbb{R}^q}x\,(dP_n-dQ_n)\right\|\leq K(\ln n)^\epsilon\sup_{A\in\mathcal{B}_C}n^a|P_n(A)-Q_n(A)|$
$+n^{a+\frac{1}{2}}\rho P\left(\sqrt{n}(\theta_n-\theta_0)\in\sqrt{n}(\Theta-\theta_0)\backslash B(0,K(\ln n)^\epsilon)\right)+n^a\int_{\mathbb{R}^q\backslash B(0,K(\ln n)^\epsilon)}\|x\|\,|dQ_n|$.
As $\sup_{A\in\mathcal{B}_C}n^a|P_n(A)-Q_n(A)|=O(n^{-\eta})$ for $\eta>0$, we have that $K(\ln n)^\epsilon\sup_{A\in\mathcal{B}_C}n^a|P_n(A)-Q_n(A)|=o(1)$ and the result follows due to assumptions A.14 and A.15 above. ■

Remark R.30 *Due to the two previous remarks and Lemma 2 in [14], it suffices that* $\sqrt{n}\,(\theta - \theta_0)$ *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.

3.1 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* $\|E\,(x_n - y_n)\| = o\,(n^{-a})$.

Remark R.31 *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 $\sup_{A \in \mathcal{B}_C} |\mathcal{EDG}\,(A) - \Phi\,(A_n)| = o\,(n^{-a})$ for suitable choice of the sequence $\{A_n\}$ emerging from a bijective correspondence $A \to A_n$. Hence $\sup_{A \in \mathcal{B}_C} |P\,(x_n \in A) - P\,(z \in A_n)| = o\,(n^{-a})$ 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 $P\,(z \in A_n) = P\,((g_n\,(z) + o\,(n^{-a})) \in A) = P\,(g_n\,(z) \in A) + o\,(n^{-a})$ for a suitable choice of a polynomial in z function sequence and the smoothness of Φ (see [14] for the definition of smoothness of a distribution, which is implied by analytical smoothness in the case where a density exists), we have that that $\sup_{A \in \mathcal{B}_C} |P\,(x_n \in A) - P\,(g_n\,(z) \in A)|$. We then employ lemma 3.1 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.31, something that will be needed in the case of GMR2. We present the following lemma that concerns approximations of inverse matrices that will be useful in what follows.

Lemma 3.2 *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

$$P(\|R_n(z)\| > \gamma_n) = o\left(n^{-a}\right)$$

where $\gamma_n = o(n^{-a})$.

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 $n^a P(\|R_n(z)\| > cn^{-a-\varepsilon}) = o(1)$ But $n^a P(\|R_n(z)\| > cn^{-a-\varepsilon}) \leq n^a P\left(\frac{1}{n^{a+\frac{1}{2}}} \|R(z)\| > cn^{-a-\varepsilon}\right) \leq n^a \frac{E\|R(z)\|^k}{\left(cn^{\frac{1}{2}-\varepsilon}\right)^k} = n^{a-\frac{k}{2}+k\varepsilon} m$ 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 $k > 2a$. ∎

Assumption A.16 *Any initial estimator has an analogous first moment approximation with the one that it defines.*

Remark R.32 *This assumption is in the spirit of assumption A.13 and can be justified in our set up.*

Auxiliary Estimators

We begin with the auxiliary estimator β_n. Next lemma summarizes the results.

Lemma 3.3 *If $\sqrt{n}(\beta_n - b(\theta_0))$ has a valid Edgeworth expansion of third order*

$$\sqrt{n}(\beta_n - b(\theta_0)) \underset{1/2}{\sim} k_1 + \frac{k_2}{\sqrt{n}}$$

where

$$k_1 = -Q^{-1}(\theta_0, b(\theta_0)) c_\beta^{/}(b(\theta_0)) W(b(\theta_0)) c(z, b(\theta_0))$$

and

$$\begin{aligned} k_2 &= -Q^{-1}(\theta_0, b(\theta_0))\, c^{/}_{\beta}(b(\theta_0))\, W(b(\theta_0))\, c^{*}(z, b(\theta_0)) - Q^{-1}(\theta_0, b(\theta_0))\, A(z, \theta_0, b(\theta_0))\, k_1 \\ &\quad -Q^{-1}(\theta_0, b(\theta_0)) \left[c^{/}_{\beta}(b(\theta_0))\, w(z, b(\theta_0)) + c_{\beta}(z, b(\theta_0))\, W(b(\theta_0))\right] c(z, b(\theta_0)) \\ &\quad -\frac{1}{2} Q^{-1}(\theta_0, b(\theta_0))\, c^{/}_{\beta}(b(\theta_0))\, W(b(\theta_0)) \left[k_1^{/} c_{\beta,\beta'}(b(\theta_0))_j\, k_1\right]_{j=1,\ldots,l} \\ &\quad -Q^{-1}(\theta_0, b(\theta_0))\, c^{/}_{\beta}(b(\theta_0)) \left[W_{\beta^{/}}(b(\theta_0))_{rj}\, k_1^{*}\right]_{r,i=1,\ldots,l} c(z, b(\theta_0)) \\ &\quad -Q^{-1}(\theta_0, b(\theta_0)) \left[k_1^{/} c_{\beta,\beta'}(b(\theta_0))_j\right]_{j=1,\ldots,l} W(b(\theta_0))\, c(z, b(\theta_0)) \\ &\quad -Q^{-1}(\theta_0, b(\theta_0))\, c^{/}_{\beta}(b(\theta_0)) \left[W_{\beta^{/}}(b(\theta_0))_{rj}\, k_1^{*}\right]_{r,i=1,\ldots,l} c_{\beta}(b(\theta_0))\, k_1 \\ &\quad -Q^{-1}(\theta_0, b(\theta_0)) \left[k_1^{/} c_{\beta,\beta'}(b(\theta_0))_j\right]_{j=1,\ldots,l} W(b(\theta_0))\, c_{\beta}(b(\theta_0))\, k_1 \end{aligned}$$

where

$$Q(\theta_0, b(\theta_0)) = E_{\theta_0} \frac{\partial}{\partial \beta} c^{/}(x_1, b(\theta_0))\, W(b(\theta_0))\, E_{\theta_0} \frac{\partial}{\partial \beta^{/}} c(x_1, b(\theta_0))$$

and

$$A(z_1, z_2, \theta_0, b(\theta_0)) = 2Sym \left[\begin{array}{c} E_{\theta_0} \frac{\partial}{\partial \beta} c^{/}(x_1, b(\theta_0))\, W(b(\theta_0))\, c^{/}_{\beta}(z_1, b(\theta_0)) \\ +\frac{1}{2} E_{\theta_0} \frac{\partial}{\partial \beta} c^{/}(x_1, b(\theta_0))\, w(z_2, b(\theta_0))\, E_{\theta_0} \frac{\partial}{\partial \beta^{/}} c(x_1, b(\theta_0)) \end{array} \right]$$

where $Sym[A] = \frac{1}{2}(A + A')$, $\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$ *and* k_1^{*} *is the relevant term of the analogous expansion of the first step auxiliary estimator due to assumption A.16.*

Proof. $\beta_n = \arg\min_{\beta \in B} \frac{1}{n} \sum_i c^{/}(x_i, \beta) \frac{1}{n} \sum_i W(x_i, \beta_n^{*}) \frac{1}{n} \sum_i c^{/}(x_i, \beta) \Rightarrow$
$\left[\frac{\partial}{\partial \beta} \frac{1}{n} \sum c'(x_i, \beta_n)\right] \frac{1}{n} \sum W(x_i, \beta_n^{*}) \frac{1}{\sqrt{n}} \sum c(x_i, \beta_n) = \mathbf{0}_q \Rightarrow$
$\left[\frac{\partial}{\partial \beta} \frac{1}{n} \sum c'(x_i, b(\theta_0)) + \frac{1}{\sqrt{n}} \left[\sqrt{n}\left(\beta_{1,n} - b(\theta_0)\right)' \frac{\partial^2}{\partial \beta \partial \beta'} \frac{1}{n} \sum c_j(x_i, b(\theta_0))\right]_{j=1,\ldots,l}\right] \times$
$\left[\frac{1}{n} \sum W(x_i, b(\theta_0)) + \frac{1}{\sqrt{n}} \left[\frac{\partial}{\partial \beta^{/}} \frac{1}{n} \sum W_{rj}(x_i, b(\theta_0)) \sqrt{n}(\beta_n^{*} - b(\theta_0))\right]_{r,i=1,\ldots,l}\right] \times$
$\left[\begin{array}{c} \frac{1}{\sqrt{n}} \sum c(x_i, b(\theta_0)) + \frac{\partial}{\partial \beta^{/}} \frac{1}{n} \sum c(x_i, b(\theta_0)) \sqrt{n}\left(\beta_{1,n} - b(\theta_0)\right) \\ +\frac{1}{2\sqrt{n}} \left[\sqrt{n}\left(\beta_{1,n} - b(\theta_0)\right)' \frac{\partial^2}{\partial \beta \partial \beta'} \frac{1}{n} \sum c_j(x_i, b(\theta_0)) \sqrt{n}\left(\beta_{1,n} - b(\theta_0)\right)\right]_{j=1,\ldots,l} \end{array} \right] =$
$\mathbf{0}_q \Rightarrow$
$\frac{\partial}{\partial \beta} \frac{1}{n} \sum c'(x_i, b(\theta_0)) \frac{1}{n} \sum W(x_i, b(\theta_0)) \frac{1}{\sqrt{n}} \sum c(x_i, b(\theta_0)) +$

$\frac{\partial}{\partial\beta}\frac{1}{n}\sum c'\left(x_i,b\left(\theta_0\right)\right)\frac{1}{n}\sum W\left(x_i,b\left(\theta_0\right)\right)\frac{\partial}{\partial\beta'}\frac{1}{n}\sum c\left(x_i,b\left(\theta_0\right)\right)\sqrt{n}\left(\beta_{1,n}-b\left(\theta_0\right)\right)$
$+\frac{1}{2\sqrt{n}}\frac{\partial}{\partial\beta}\frac{1}{n}\sum c'\left(x_i,b\left(\theta_0\right)\right)\frac{1}{n}\sum W\left(x_i,b\left(\theta_0\right)\right)\times$
$\left[\sqrt{n}\left(\beta_{1,n}-b\left(\theta_0\right)\right)'\frac{\partial^2}{\partial\beta\partial\beta'}\frac{1}{n}\sum c_j\left(x_i,b\left(\theta_0\right)\right)\sqrt{n}\left(\beta_{1,n}-b\left(\theta_0\right)\right)\right]_{j=1,\ldots,l}$
$+\begin{pmatrix}\frac{\partial}{\partial\beta}\frac{1}{n}\sum c'\left(x_i,b\left(\theta_0\right)\right)\frac{1}{\sqrt{n}}\left[\frac{\partial}{\partial\beta'}\frac{1}{n}\sum W_{rj}\left(x_i,b\left(\theta_0\right)\right)\sqrt{n}\left(\beta_n^*-b\left(\theta_0\right)\right)\right]_{r,i=1,\ldots,l}\\+\frac{1}{\sqrt{n}}\left[\sqrt{n}\left(\beta_{1,n}-b\left(\theta_0\right)\right)'\frac{\partial^2}{\partial\beta\partial\beta'}\frac{1}{n}\sum c_j\left(x_i,b\left(\theta_0\right)\right)\right]_{j=1,\ldots,l}\frac{1}{n}\sum W\left(x_i,b\left(\theta_0\right)\right)\end{pmatrix}\times$
$\left[\frac{1}{\sqrt{n}}\sum c\left(x_i,b\left(\theta_0\right)\right)+\frac{\partial}{\partial\beta'}\frac{1}{n}\sum c\left(x_i,b\left(\theta_0\right)\right)\sqrt{n}\left(\beta_{1,n}-b\left(\theta_0\right)\right)\right] \quad = \quad \mathbf{0}_q$
As the term $\frac{1}{\sqrt{n}}\left[\sqrt{n}\left(\beta_{1,n}-b\left(\theta_0\right)\right)'\frac{\partial^2}{\partial\beta\partial\beta'}\sqrt{n}\frac{1}{n}\sum c_j\left(x_i,b\left(\theta_0\right)\right)\right]_{j=1,\ldots,l}\times$
$\frac{1}{\sqrt{n}}\left[\frac{\partial}{\partial\beta'}\frac{1}{n}\sum W_{rj}\left(x_i,b\left(\theta_0\right)\right)\sqrt{n}\left(\beta_n^*-b\left(\theta_0\right)\right)\right]_{r,i=1,\ldots,l}\frac{1}{n}\sum c\left(x_i,b\left(\theta_0\right)\right)=o\left(n^{-\frac{1}{2}}\right)$

$\sqrt{n}\left(\beta_{1,n}-b\left(\theta_0\right)\right)=-\left[\frac{\partial}{\partial\beta}\frac{1}{n}\sum c'\left(x_i,b\left(\theta_0\right)\right)\frac{1}{n}\sum W\left(x_i,b\left(\theta_0\right)\right)\frac{\partial}{\partial\beta'}\frac{1}{n}\sum c\left(x_i,b\left(\theta_0\right)\right)\right]^{-1}\times$

$\frac{\partial}{\partial\beta}\frac{1}{n}\sum c'\left(x_i,b\left(\theta_0\right)\right)\frac{1}{n}\sum W\left(x_i,b\left(\theta_0\right)\right)\frac{1}{\sqrt{n}}\sum c\left(x_i,b\left(\theta_0\right)\right)$
$-\frac{1}{2\sqrt{n}}\left[\frac{\partial}{\partial\beta}\frac{1}{n}\sum c'\left(x_i,b\left(\theta_0\right)\right)\frac{1}{n}\sum W\left(x_i,b\left(\theta_0\right)\right)\frac{\partial}{\partial\beta'}\frac{1}{n}\sum c\left(x_i,b\left(\theta_0\right)\right)\right]^{-1}\frac{\partial}{\partial\beta}\frac{1}{n}\sum c'\left(x_i,b\left(\theta_0\right)\right)\times$
$\frac{1}{n}\sum W\left(x_i,b\left(\theta_0\right)\right)\left[\sqrt{n}\left(\beta_{1,n}-b\left(\theta_0\right)\right)'\frac{\partial^2}{\partial\beta\partial\beta'}\frac{1}{n}\sum c_j\left(x_i,b\left(\theta_0\right)\right)\sqrt{n}\left(\beta_{1,n}-b\left(\theta_0\right)\right)\right]_{j=1,\ldots,l}$
$-\left[\frac{\partial}{\partial\beta}\frac{1}{n}\sum c'\left(x_i,b\left(\theta_0\right)\right)\frac{1}{n}\sum W\left(x_i,b\left(\theta_0\right)\right)\frac{\partial}{\partial\beta'}\frac{1}{n}\sum c\left(x_i,b\left(\theta_0\right)\right)\right]^{-1}\times$
$\begin{pmatrix}\frac{\partial}{\partial\beta}\frac{1}{n}\sum c'\left(x_i,b\left(\theta_0\right)\right)\frac{1}{\sqrt{n}}\left[\frac{\partial}{\partial\beta'}\frac{1}{n}\sum W_{rj}\left(x_i,b\left(\theta_0\right)\right)\sqrt{n}\left(\beta_n^*-b\left(\theta_0\right)\right)\right]_{r,i=1,\ldots,l}\\+\frac{1}{\sqrt{n}}\left[\sqrt{n}\left(\beta_{1,n}-b\left(\theta_0\right)\right)'\frac{\partial^2}{\partial\beta\partial\beta'}\frac{1}{n}\sum c_j\left(x_i,b\left(\theta_0\right)\right)\right]_{j=1,\ldots,l}\frac{1}{n}\sum W\left(x_i,b\left(\theta_0\right)\right)\end{pmatrix}\frac{1}{\sqrt{n}}\sum c\left(x_i,b\left(\theta_0\right)\right)$
$-\left[\frac{\partial}{\partial\beta}\frac{1}{n}\sum c'\left(x_i,b\left(\theta_0\right)\right)\frac{1}{n}\sum W\left(x_i,b\left(\theta_0\right)\right)\frac{\partial}{\partial\beta'}\frac{1}{n}\sum c\left(x_i,b\left(\theta_0\right)\right)\right]^{-1}\times$
$\left(\frac{\partial}{\partial\beta}\frac{1}{n}\sum c'\left(x_i,b\left(\theta_0\right)\right)\frac{1}{\sqrt{n}}\left[\frac{\partial}{\partial\beta'}\frac{1}{n}\sum W_{rj}\left(x_i,b\left(\theta_0\right)\right)\sqrt{n}\left(\beta_n^*-b\left(\theta_0\right)\right)\right]_{r,i=1,\ldots,l}\right)\times$
$\frac{\partial}{\partial\beta'}\frac{1}{n}\sum c\left(x_i,b\left(\theta_0\right)\right)\sqrt{n}\left(\beta_{1,n}-b\left(\theta_0\right)\right)$
$-\left[\frac{\partial}{\partial\beta}\frac{1}{n}\sum c'\left(x_i,b\left(\theta_0\right)\right)\frac{1}{n}\sum W\left(x_i,b\left(\theta_0\right)\right)\frac{\partial}{\partial\beta'}\frac{1}{n}\sum c\left(x_i,b\left(\theta_0\right)\right)\right]^{-1}\times$
$\frac{1}{\sqrt{n}}\left[\sqrt{n}\left(\beta_{1,n}-b\left(\theta_0\right)\right)'\frac{\partial^2}{\partial\beta\partial\beta'}\frac{1}{n}\sum c_j\left(x_i,b\left(\theta_0\right)\right)\right]_{j=1,\ldots,l}\frac{1}{n}\sum W\left(x_i,b\left(\theta_0\right)\right)\times$
$\frac{\partial}{\partial\beta'}\frac{1}{n}\sum c\left(x_i,b\left(\theta_0\right)\right)\sqrt{n}\left(\beta_{1,n}-b\left(\theta_0\right)\right).$

Employing now the moment approximations for the analogous terms of $\frac{1}{n}\sum_{i=1}^{n}\left(f\left(x_i,b\left(\theta_0\right),\theta_0\right)-E\left(f\left(x_i,b\left(\theta_0\right),\theta_0\right)\right)\right)$, due to remark R.26 and lemma 3.1 and holding terms up to the relevant order, $\frac{\partial}{\partial\beta}\frac{1}{n}\sum c^{/}\left(x_i,b\left(\theta_0\right)\right)=c_{\beta}^{/}\left(b\left(\theta_0\right)\right)+\frac{1}{\sqrt{n}}c_{\beta}^{/}\left(z,b\left(\theta_0\right)\right)$, $\frac{1}{n}\sum W\left(x_i,b\left(\theta_0\right)\right)=W\left(b\left(\theta_0\right)\right)+\frac{1}{\sqrt{n}}w\left(z,b\left(\theta_0\right)\right)$, and $\frac{1}{\sqrt{n}}\sum c\left(x_i,b\left(\theta_0\right)\right)=c\left(z,b\left(\theta_0\right)\right)+\frac{1}{\sqrt{n}}c^*\left(z,b\left(\theta_0\right)\right)$ where

$z \sim N(0,\Sigma)$ and the elements of $c_\beta(z,b(\theta_0))$, $w(z,b(\theta_0))$ and $c(z,b(\theta_0))$ are finite polynomials in z with $O(1)$ coefficients and $E_{\theta_0}c(z,b(\theta_0)) = E(c^*(z,b(\theta_0))) = 0$, we get that

$$\left[\frac{\partial}{\partial\beta}\frac{1}{n}\sum c'(x_i,b(\theta_0))\frac{1}{n}\sum W(x_i,b(\theta_0))\frac{\partial}{\partial\beta'}\frac{1}{n}\sum c(x_i,b(\theta_0))\right]^{-1} =$$

$$= \left[\begin{array}{c} c_\beta'(b(\theta_0))W(b(\theta_0))c_\beta(b(\theta_0)) \\ +\frac{1}{\sqrt{n}}\left[\begin{array}{c} c_\beta'(b(\theta_0))W(b(\theta_0))c_\beta(z,b(\theta_0)) \\ +c_\beta'(z,b(\theta_0))W(b(\theta_0))c_\beta(b(\theta_0)) \\ +c_\beta'(b(\theta_0))w(z,b(\theta_0))c_\beta(b(\theta_0)) \end{array}\right] \end{array}\right]^{-1} =$$

$$\left[\begin{array}{c} Q(\theta_0,b(\theta_0)) \\ +\frac{1}{\sqrt{n}}\left[\begin{array}{c} c_\beta'(b(\theta_0))W(b(\theta_0))c_\beta(z,b(\theta_0)) \\ +c_\beta'(z,b(\theta_0))W(b(\theta_0))c_\beta(b(\theta_0)) \\ +c_\beta'(b(\theta_0))w(z,b(\theta_0))c_\beta(b(\theta_0)) \end{array}\right] \end{array}\right]^{-1} =$$

$$Q^{-1}(\theta_0,b(\theta_0)) - \frac{1}{\sqrt{n}}Q^{-1}(\theta_0,b(\theta_0))A(z,\theta_0,b(\theta_0))Q^{-1}(\theta_0,b(\theta_0)) + o\left(n^{-\frac{1}{2}}\right)$$

due to lemma 3.2 where

$$Q(\theta_0,b(\theta_0)) = c_\beta'(b(\theta_0))W(b(\theta_0))c_\beta(b(\theta_0))$$

and

$$A(z,\theta_0,b(\theta_0)) = 2\mathrm{Sym}\left[\begin{array}{c} c_\beta'(b(\theta_0))W(b(\theta_0))c_\beta(z,b(\theta_0)) \\ +\frac{1}{2}c_\beta'(b(\theta_0))w(z,b(\theta_0))c_\beta(b(\theta_0)) \end{array}\right]$$

where $\mathrm{Sym}[A] = \frac{1}{2}(A+A')$ see Corollary 1 [14],

$\frac{\partial}{\partial\beta}\frac{1}{n}\sum c'(x_i,b(\theta_0))\frac{1}{n}\sum W(x_i,b(\theta_0)) =$
$c_\beta'(b(\theta_0))W(b(\theta_0)) + \frac{1}{\sqrt{n}}\left[c_\beta'(b(\theta_0))w(z,b(\theta_0)) + c_\beta'(z,b(\theta_0))W(b(\theta_0))\right] +$
$o\left(n^{-\frac{1}{2}}\right)$

and $\frac{\partial^2}{\partial\beta\partial\beta'}\frac{1}{n}\sum c_j(x_i,b(\theta_0)) = c_{\beta,\beta'}(b(\theta_0)) + \frac{1}{\sqrt{n}}c_{\beta,\beta'}(z,b(\theta_0))$ and $\frac{\partial}{\partial\beta'}\frac{1}{n}\sum W_{rj}(x_i,b(\theta_0)) = W_{\beta'}(b(\theta_0))_{rj} + \frac{1}{\sqrt{n}}w_{\beta'}(z,b(\theta_0))_{rj}$ due to remark R.26 and lemma 3.1

$\frac{\partial}{\partial\beta}\frac{1}{n}\sum c'(x_i,b(\theta_0)) = c_\beta'(b(\theta_0)) + \frac{1}{\sqrt{n}}c_\beta'(z,b(\theta_0))$,
$\frac{1}{n}\sum W(x_i,b(\theta_0)) = W(b(\theta_0)) + \frac{1}{\sqrt{n}}w(z,b(\theta_0))$,
$\frac{1}{\sqrt{n}}\sum c(x_i,b(\theta_0)) = c(z,b(\theta_0)) + \frac{1}{\sqrt{n}}c^*(z,b(\theta_0))$
$\frac{\partial^2}{\partial\beta\partial\beta'}\frac{1}{n}\sum c_j(x_i,b(\theta_0)) = c_{\beta,\beta'}(b(\theta_0))_j + \frac{1}{\sqrt{n}}c_{\beta,\beta'}(z,b(\theta_0))_j$
$\frac{\partial}{\partial\beta'}\frac{1}{n}\sum W_{rj}(x_i,b(\theta_0)) = W_{\beta'}(b(\theta_0))_{rj} + \frac{1}{\sqrt{n}}w_{\beta'}(z,b(\theta_0))_{rj}$

Hence $\sqrt{n}\left(\beta_{1,n} - b(\theta_0)\right) = -Q^{-1}(\theta_0,b(\theta_0))c_\beta'(b(\theta_0))W(b(\theta_0))c(z,b(\theta_0))$
$-\frac{1}{\sqrt{n}}Q^{-1}(\theta_0,b(\theta_0))c_\beta'(b(\theta_0))W(b(\theta_0))c^*(z,b(\theta_0))$

$+\frac{1}{\sqrt{n}}Q^{-1}\left(\theta_0,b\left(\theta_0\right)\right)A\left(z,\theta_0,b\left(\theta_0\right)\right)Q^{-1}\left(\theta_0,b\left(\theta_0\right)\right)c_{\beta}^{/}\left(b\left(\theta_0\right)\right)W\left(b\left(\theta_0\right)\right)c\left(z,b\left(\theta_0\right)\right)$
$-\frac{1}{\sqrt{n}}Q^{-1}\left(\theta_0,b\left(\theta_0\right)\right)\left[c_{\beta}^{/}\left(b\left(\theta_0\right)\right)w\left(z,b\left(\theta_0\right)\right)+c_{\beta}\left(z,b\left(\theta_0\right)\right)W\left(b\left(\theta_0\right)\right)\right]c\left(z,b\left(\theta_0\right)\right)$
$-\frac{1}{2\sqrt{n}}Q^{-1}\left(\theta_0,b\left(\theta_0\right)\right)c_{\beta}^{/}\left(b\left(\theta_0\right)\right)W\left(b\left(\theta_0\right)\right)$
$\times\qquad\left[\sqrt{n}\left(\beta_{1,n}-b\left(\theta_0\right)\right)'c_{\beta,\beta'}\left(b\left(\theta_0\right)\right)_j\sqrt{n}\left(\beta_{1,n}-b\left(\theta_0\right)\right)\right]_{j=1,\ldots,l}$
$-\frac{1}{\sqrt{n}}Q^{-1}\left(\theta_0,b\left(\theta_0\right)\right)c_{\beta}^{/}\left(b\left(\theta_0\right)\right)\left[W_{\beta^{/}}\left(b\left(\theta_0\right)\right)_{rj}k_{1,n}^{*}\right]_{r,i=1,\ldots,l}c\left(z,b\left(\theta_0\right)\right)$
$-\frac{1}{\sqrt{n}}Q^{-1}\left(\theta_0,b\left(\theta_0\right)\right)\left[\sqrt{n}\left(\beta_{1,n}-b\left(\theta_0\right)\right)'c_{\beta,\beta'}\left(b\left(\theta_0\right)\right)_j\right]_{j=1,\ldots,l}W\left(b\left(\theta_0\right)\right)c\left(z,b\left(\theta_0\right)\right)$
$-\frac{1}{\sqrt{n}}Q^{-1}\left(\theta_0,b\left(\theta_0\right)\right)c_{\beta}^{/}\left(b\left(\theta_0\right)\right)\left[W_{\beta^{/}}\left(b\left(\theta_0\right)\right)_{rj}k_{1}^{*}\right]_{r,i=1,\ldots,l}c_{\beta}\left(b\left(\theta_0\right)\right)\sqrt{n}\left(\beta_{1,n}-b\left(\theta_0\right)\right)$
$-\frac{1}{\sqrt{n}}Q^{-1}\left(\theta_0,b\left(\theta_0\right)\right)\left[\sqrt{n}\left(\beta_{1,n}-b\left(\theta_0\right)\right)'c_{\beta,\beta'}\left(b\left(\theta_0\right)\right)_j\right]_{j=1,\ldots,l}W\left(b\left(\theta_0\right)\right)c_{\beta}\left(b\left(\theta_0\right)\right)$
$\times\sqrt{n}\left(\beta_{1,n}-b\left(\theta_0\right)\right)$
where $k_{1,n}^{*}$ is given in $\sqrt{n}\left(\beta_{1,n}^{*}-b\left(\theta_0\right)\right)\underset{1/2}{\sim}k_1^{*}+\frac{k_2^{*}}{\sqrt{n}}$. Now setting
$k_1=-Q^{-1}\left(\theta_0,b\left(\theta_0\right)\right)c_{\beta}^{/}\left(b\left(\theta_0\right)\right)W\left(b\left(\theta_0\right)\right)c\left(z,b\left(\theta_0\right)\right)$ we get
$\sqrt{n}\left(\beta_{1,n}-b\left(\theta_0\right)\right)=k_1-\frac{1}{\sqrt{n}}Q^{-1}\left(\theta_0,b\left(\theta_0\right)\right)c_{\beta}^{/}\left(b\left(\theta_0\right)\right)W\left(b\left(\theta_0\right)\right)c^{*}\left(z,b\left(\theta_0\right)\right)$
$-\frac{1}{\sqrt{n}}Q^{-1}\left(\theta_0,b\left(\theta_0\right)\right)A\left(z,\theta_0,b\left(\theta_0\right)\right)k_1$
$-\frac{1}{\sqrt{n}}Q^{-1}\left(\theta_0,b\left(\theta_0\right)\right)\left[c_{\beta}^{/}\left(b\left(\theta_0\right)\right)w\left(z,b\left(\theta_0\right)\right)+c_{\beta}\left(z,b\left(\theta_0\right)\right)W\left(b\left(\theta_0\right)\right)\right]c\left(z,b\left(\theta_0\right)\right)$
$-\frac{1}{2\sqrt{n}}Q^{-1}\left(\theta_0,b\left(\theta_0\right)\right)c_{\beta}^{/}\left(b\left(\theta_0\right)\right)W\left(b\left(\theta_0\right)\right)\left[k_1^{/}c_{\beta,\beta'}\left(b\left(\theta_0\right)\right)_j k_1\right]_{j=1,\ldots,l}$
$-\frac{1}{\sqrt{n}}Q^{-1}\left(\theta_0,b\left(\theta_0\right)\right)c_{\beta}^{/}\left(b\left(\theta_0\right)\right)\left[W_{\beta^{/}}\left(b\left(\theta_0\right)\right)_{rj}k_{1}^{*}\right]_{r,i=1,\ldots,l}c\left(z,b\left(\theta_0\right)\right)$
$-\frac{1}{\sqrt{n}}Q^{-1}\left(\theta_0,b\left(\theta_0\right)\right)\left[k_1^{/}c_{\beta,\beta'}\left(b\left(\theta_0\right)\right)_j\right]_{j=1,\ldots,l}W\left(b\left(\theta_0\right)\right)c\left(z,b\left(\theta_0\right)\right)$
$-\frac{1}{\sqrt{n}}Q^{-1}\left(\theta_0,b\left(\theta_0\right)\right)c_{\beta}^{/}\left(b\left(\theta_0\right)\right)\left[W_{\beta^{/}}\left(b\left(\theta_0\right)\right)_{rj}k_{1}^{*}\right]_{r,i=1,\ldots,l}c_{\beta}\left(b\left(\theta_0\right)\right)k_1$
$-\frac{1}{\sqrt{n}}Q^{-1}\left(\theta_0,b\left(\theta_0\right)\right)\left[k_1^{/}c_{\beta,\beta'}\left(b\left(\theta_0\right)\right)_j\right]_{j=1,\ldots,l}W\left(b\left(\theta_0\right)\right)c_{\beta}\left(b\left(\theta_0\right)\right)k_1\quad\blacksquare$

Remark R.33 *It is easy to see that when $l=q$ the results do not depend on the weighting matrix as expected.*

Remark R.34 *$E_{\theta_0}k_{1,n}$ 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 3.1, $E_{\theta_0}k_{2,n}$ 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 [15]).*

Indirect Estimators

We proceed to state the main results concerning the expansions of the three indirect estimators. These reveal a quite different behavior of GMR 2 from the other two, due to the fact that the computation of the particular estimator is based upon the term $E_\theta \beta_n$.

GMR 1 Estimator We begin with the GMR 1 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 and the behavior of the weighting matrix and through this of the initial estimator θ_n^*.

Lemma 3.4 *If $\sqrt{n}\left(\theta_n - \theta_0\right)$ has a valid Edgeworth expansion of third order*

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

where

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

$$\begin{aligned}
q_2 \;=\;& \left(\frac{\partial b'\left(\theta_0\right)}{\partial \theta} W^*\left(b\left(\theta_0\right)\right) \frac{\partial b\left(\theta_0\right)}{\partial \theta'}\right)^{-1} \frac{\partial b'\left(\theta_0\right)}{\partial \theta} w^*\left(z, \theta_0\right) k_1 \\
& -\left(\frac{\partial b'\left(\theta_0\right)}{\partial \theta} W^*\left(b\left(\theta_0\right)\right) \frac{\partial b\left(\theta_0\right)}{\partial \theta'}\right)^{-1} \frac{\partial b'\left(\theta_0\right)}{\partial \theta} w^*\left(z, \theta_0\right) \\
& \times \frac{\partial b\left(\theta_0\right)}{\partial \theta'} \left(\frac{\partial b'\left(\theta_0\right)}{\partial \theta} W^*\left(b\left(\theta_0\right)\right) \frac{\partial b\left(\theta_0\right)}{\partial \theta'}\right)^{-1} \frac{\partial b'\left(\theta_0\right)}{\partial \theta} W^*\left(b\left(\theta_0\right)\right) k_1 \\
& +\left(\frac{\partial b'\left(\theta_0\right)}{\partial \theta} W^*\left(b\left(\theta_0\right)\right) \frac{\partial b\left(\theta_0\right)}{\partial \theta'}\right)^{-1} \frac{\partial b'\left(\theta_0\right)}{\partial \theta} W^*\left(b\left(\theta_0\right)\right) k_2 \\
& -\frac{1}{2}\left(\frac{\partial b'\left(\theta_0\right)}{\partial \theta} W^*\left(b\left(\theta_0\right)\right) \frac{\partial b\left(\theta_0\right)}{\partial \theta'}\right)^{-1} \frac{\partial b'\left(\theta_0\right)}{\partial \theta} W^*\left(b\left(\theta_0\right)\right) \left[tr q_{1,n} q_{1,n}' \frac{\partial^2 b_j}{\partial \theta \partial \theta'}\right]_{j=1,\ldots,l} \\
& +\left(\frac{\partial b'\left(\theta_0\right)}{\partial \theta} W^*\left(b\left(\theta_0\right)\right) \frac{\partial b\left(\theta_0\right)}{\partial \theta'}\right)^{-1} \frac{\partial b'\left(\theta_0\right)}{\partial \theta} \left[\frac{\partial}{\partial \theta} \frac{1}{n} W_{rj}^*\left(\theta_0\right) q_1^*\right]_{r,i=1,\ldots,l} A k_1 \\
& +\left(\frac{\partial b'\left(\theta_0\right)}{\partial \theta} W^*\left(b\left(\theta_0\right)\right) \frac{\partial b\left(\theta_0\right)}{\partial \theta'}\right)^{-1} \left[q_{1,n}' \frac{\partial^2 b_j\left(\theta_0\right)}{\partial \theta \partial \theta'}\right]_{j=1,\ldots,q} W^*\left(b\left(\theta_0\right)\right) A k_1,
\end{aligned}$$

$A = Id_{q\times q} - \frac{\partial b'(\theta_0)}{\partial \theta}\left(\frac{\partial b'(\theta_0)}{\partial \theta} W^*\left(b\left(\theta_0\right)\right) \frac{\partial b(\theta_0)}{\partial \theta'}\right)^{-1} \frac{\partial b'(\theta_0)}{\partial \theta} W^*\left(b\left(\theta_0\right)\right)$, *and q_1^* is the relevant term of the analogous expansion of the first step auxiliary estimator due to assumption A.16.*

Proof. Utilizing assumption A.16 we have that

$$\frac{\partial b'(\theta_n)}{\partial\theta}\frac{1}{n}\sum W^*(x_i,\theta_n^*)\sqrt{n}(\beta_n-b(\theta_n))=\mathbf{0}_p\Rightarrow$$

$$\left(\frac{\partial b'(\theta_0)}{\partial\theta}+\frac{1}{\sqrt{n}}\left[\sqrt{n}(\theta_n-\theta_0)'\frac{\partial^2 b_j(\theta_0)}{\partial\theta\partial\theta'}\right]_{j=1,\ldots,p}\right)\times$$

$$\left(\frac{1}{n}\sum W^*(x_i,\theta_0)+\frac{1}{\sqrt{n}}\left[\frac{\partial}{\partial\theta'}\frac{1}{n}\sum W^*_{rj}(x_i,\theta_0)\sqrt{n}(\theta_n^*-\theta_0)\right]_{r,i=1,\ldots,q}\right)\times$$

$$\left(\sqrt{n}(\beta_n-b(\theta_0))-\frac{\partial b(\theta_0)}{\partial\theta'}\sqrt{n}(\theta_n-\theta_0)-\frac{1}{2\sqrt{n}}\left[\sqrt{n}(\theta_n-\theta_0)'\frac{\partial^2 b_j(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}(\theta_n-\theta_0)\right]_{j=1,\ldots,q}\right)=$$

$$\mathbf{0}_p\Rightarrow$$

$$\begin{bmatrix}\frac{\partial b'(\theta_0)}{\partial\theta}\frac{1}{n}\sum W^*(x_i,\theta_0)\\ +\frac{1}{\sqrt{n}}\frac{\partial b'(\theta_0)}{\partial\theta}\left[\frac{\partial}{\partial\theta'}\frac{1}{n}\sum W^*_{rj}(x_i,\theta_0)\sqrt{n}(\theta_n^*-\theta_0)\right]_{r,i=1,\ldots,q}\\ +\frac{1}{\sqrt{n}}\left[\sqrt{n}(\theta_n-\theta_0)'\frac{\partial^2 b_j(\theta_0)}{\partial\theta\partial\theta'}\right]_{j=1,\ldots,q}\frac{1}{n}\sum W^*(x_i,\theta_0)\end{bmatrix}$$

$$\left(k_1+\frac{k_2}{\sqrt{n}}-\frac{\partial b(\theta_0)}{\partial\theta'}\sqrt{n}(\theta_n-\theta_0)-\frac{1}{2\sqrt{n}}\left[\sqrt{n}(\theta_n-\theta_0)'\frac{\partial^2 b_j(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}(\theta_n-\theta_0)\right]_{j=1,\ldots,q}\right)=$$

$$\mathbf{0}_p\Rightarrow$$

$$\frac{\partial b'(\theta_0)}{\partial\theta}\frac{1}{n}\sum W^*(x_i,\theta_0)k_1+\frac{\partial b'(\theta_0)}{\partial\theta}\frac{1}{n}\sum W^*(x_i,\theta_0)\frac{k_2}{\sqrt{n}}$$

$$-\frac{\partial b'(\theta_0)}{\partial\theta}\frac{1}{n}\sum W^*(x_i,\theta_0)\frac{\partial b(\theta_0)}{\partial\theta'}\sqrt{n}(\theta_n-\theta_0)-\frac{1}{\sqrt{n}}\frac{\partial b'(\theta_0)}{\partial\theta}\left[\frac{\partial}{\partial\theta'}\frac{1}{n}\sum W^*_{rj}(x_i,\theta_0)q_1^*\right]_{r,i=1,\ldots,q}\times$$

$$\frac{\partial b(\theta_0)}{\partial\theta'}\sqrt{n}(\theta_n-\theta_0)$$

$$-\frac{1}{2\sqrt{n}}\frac{\partial b'(\theta_0)}{\partial\theta}\frac{1}{n}\sum W^*(x_i,\theta_0)\left[\sqrt{n}(\theta_n-\theta_0)'\frac{\partial^2 b_j(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}(\theta_n-\theta_0)\right]_{j=1,\ldots,q}$$

$$+\frac{1}{\sqrt{n}}\frac{\partial b'(\theta_0)}{\partial\theta}\left[\frac{\partial}{\partial\theta'}\frac{1}{n}\sum W^*_{rj}(x_i,\theta_0)q_1^*\right]_{r,i=1,\ldots,q}k_1$$

$$+\frac{1}{\sqrt{n}}\left[\sqrt{n}(\theta_n-\theta_0)'\frac{\partial^2 b_j(\theta_0)}{\partial\theta\partial\theta'}\right]_{j=1,\ldots,q}\frac{1}{n}\sum W^*(x_i,\theta_0)k_1$$

$$-\frac{1}{\sqrt{n}}\left[\sqrt{n}(\theta_n-\theta_0)'\frac{\partial^2 b_j(\theta_0)}{\partial\theta\partial\theta'}\right]_{j=1,\ldots,q}\frac{1}{n}\sum W^*(x_i,\theta_0)\frac{\partial b(\theta_0)}{\partial\theta'}\sqrt{n}(\theta_n-\theta_0)=\mathbf{0}_p\Rightarrow$$

$$\sqrt{n}(\theta_n-\theta_0)=\left(\frac{\partial b'(\theta_0)}{\partial\theta}\frac{1}{n}\sum W^*(x_i,\theta_0)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\frac{\partial b'(\theta_0)}{\partial\theta}\frac{1}{n}\sum W^*(x_i,\theta_0)k_1+$$

$$\left(\frac{\partial b'(\theta_0)}{\partial\theta}\frac{1}{n}\sum W^*(x_i,\theta_0)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\frac{\partial b'(\theta_0)}{\partial\theta}\frac{1}{n}\sum W^*(x_i,\theta_0)\frac{k_2}{\sqrt{n}}$$

$$-\frac{1}{2\sqrt{n}}\left(\frac{\partial b'(\theta_0)}{\partial\theta}\frac{1}{n}\sum W^*(x_i,\theta_0)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\frac{\partial b'(\theta_0)}{\partial\theta}\frac{1}{n}\sum W^*(x_i,\theta_0)\left[tr\left(q_1q_1'\frac{\partial^2 b_j(\theta_0)}{\partial\theta\partial\theta'}\right)\right]_{j=1,\ldots,q}$$

$$+\frac{1}{\sqrt{n}}\left(\frac{\partial b'(\theta_0)}{\partial\theta}\frac{1}{n}\sum W^*(x_i,\theta_0)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\frac{\partial b'(\theta_0)}{\partial\theta}\left[\frac{\partial}{\partial\theta'}\frac{1}{n}\sum W^*_{rj}(x_i,\theta_0)q_1^*\right]_{r,i=1,\ldots,q}\times$$

$$\left[Id_{q\times q}-\frac{\partial b(\theta_0)}{\partial\theta'}\left(\frac{\partial b'(\theta_0)}{\partial\theta}\frac{1}{n}\sum W^*(x_i,\theta_0)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\frac{\partial b'(\theta_0)}{\partial\theta}\frac{1}{n}\sum W^*(x_i,\theta_0)\right]k_1$$

$$+\frac{1}{\sqrt{n}}\left(\frac{\partial b'(\theta_0)}{\partial\theta}\frac{1}{n}\sum W^*(x_i,\theta_0)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\left[q_1'\frac{\partial^2 b_j(\theta_0)}{\partial\theta\partial\theta'}\right]_{j=1,\ldots,q}\frac{1}{n}\sum W^*(x_i,\theta_0)\times$$

$$\left[Id_{q\times q}-\frac{\partial b(\theta_0)}{\partial\theta'}\left(\frac{\partial b'(\theta_0)}{\partial\theta}\frac{1}{n}\sum W^*(x_i,\theta_0)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\frac{\partial b'(\theta_0)}{\partial\theta}\frac{1}{n}\sum W^*(x_i,\theta_0)\right]k_1=>$$

Employing now the moment approximations for the analogous terms of $\frac{1}{n}\sum_{i=1}^{n}\left(f\left(x_i,b\left(\theta_0\right),\theta_0\right)-E\left(f\left(x_i,b\left(\theta_0\right),\theta_0\right)\right)\right)$, due to remark R.26 and lemma 3.1 and holding terms up to the relevant order,

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

$$\left(\frac{\partial b'(\theta_0)}{\partial\theta}W^*\left(b\left(\theta_0\right)\right)\frac{\partial b(\theta_0)}{\partial\theta'}+\frac{1}{\sqrt{n}}\frac{\partial b'(\theta_0)}{\partial\theta}w^*\left(z,\theta_0\right)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}=$$

$$\left(\frac{\partial b'(\theta_0)}{\partial\theta}W^*\left(b\left(\theta_0\right)\right)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}$$

$$-\frac{1}{\sqrt{n}}\left(\frac{\partial b'(\theta_0)}{\partial\theta}W^*\left(b\left(\theta_0\right)\right)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\frac{\partial b'(\theta_0)}{\partial\theta}w^*\left(z,\theta_0\right)\frac{\partial b(\theta_0)}{\partial\theta'}\left(\frac{\partial b'(\theta_0)}{\partial\theta}W^*\left(b\left(\theta_0\right)\right)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1},$$

due to lemma 3.2. Further

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

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

$$+\frac{1}{\sqrt{n}}\left(\frac{\partial b'(\theta_0)}{\partial\theta}W^*\left(b\left(\theta_0\right)\right)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\frac{\partial b'(\theta_0)}{\partial\theta}w^*\left(z_1,\theta_0\right)$$

$$-\frac{1}{\sqrt{n}}\left(\frac{\partial b'(\theta_0)}{\partial\theta}W^*\left(b\left(\theta_0\right)\right)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\frac{\partial b'(\theta_0)}{\partial\theta}w^*\left(z_1,\theta_0\right)\frac{\partial b(\theta_0)}{\partial\theta'}$$

$$\times\left(\frac{\partial b'(\theta_0)}{\partial\theta}W^*\left(b\left(\theta_0\right)\right)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\frac{\partial b'(\theta_0)}{\partial\theta}W^*\left(b\left(\theta_0\right)\right)$$

It follows that

$$\sqrt{n}\left(\theta_n-\theta_0\right)=\left(\frac{\partial b'(\theta_0)}{\partial\theta}W^*\left(b\left(\theta_0\right)\right)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\frac{\partial b'(\theta_0)}{\partial\theta}W^*\left(b\left(\theta_0\right)\right)k_1+$$

$$\frac{1}{\sqrt{n}}\left(\frac{\partial b'(\theta_0)}{\partial\theta}W^*\left(b\left(\theta_0\right)\right)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\frac{\partial b'(\theta_0)}{\partial\theta}w^*\left(z,\theta_0\right)k_1$$

$$-\frac{1}{\sqrt{n}}\left(\frac{\partial b'(\theta_0)}{\partial\theta}W^*\left(b\left(\theta_0\right)\right)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\frac{\partial b'(\theta_0)}{\partial\theta}w^*\left(z,\theta_0\right)\frac{\partial b(\theta_0)}{\partial\theta'}\left(\frac{\partial b'(\theta_0)}{\partial\theta}W^*\left(b\left(\theta_0\right)\right)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\times$$

$$\frac{\partial b'(\theta_0)}{\partial\theta}W^*\left(b\left(\theta_0\right)\right)k_1$$

$$+\frac{1}{\sqrt{n}}\left(\frac{\partial b'(\theta_0)}{\partial\theta}W^*\left(b\left(\theta_0\right)\right)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\frac{\partial b'(\theta_0)}{\partial\theta}W^*\left(b\left(\theta_0\right)\right)k_2$$

$$-\frac{1}{2\sqrt{n}}\left(\frac{\partial b'(\theta_0)}{\partial\theta}W^*\left(b\left(\theta_0\right)\right)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\frac{\partial b'(\theta_0)}{\partial\theta}W^*\left(b\left(\theta_0\right)\right)\left[tr\left(q_1q_1'\frac{\partial^2 b_j(\theta_0)}{\partial\theta\partial\theta'}\right)\right]_{j=1,\ldots,q}$$

$$+\frac{1}{\sqrt{n}}\left(\frac{\partial b'(\theta_0)}{\partial\theta}W^*\left(b\left(\theta_0\right)\right)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\frac{\partial b'(\theta_0)}{\partial\theta}\left[\frac{\partial}{\partial\theta'}\frac{1}{n}\sum W_{rj}^*\left(x_i,\theta_0\right)q_1^*\right]_{r,i=1,\ldots,q}\times$$

$$\left[Id_{q\times q}-\frac{\partial b(\theta_0)}{\partial\theta'}\left(\frac{\partial b'(\theta_0)}{\partial\theta}W^*\left(b\left(\theta_0\right)\right)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\frac{\partial b'(\theta_0)}{\partial\theta}W^*\left(b\left(\theta_0\right)\right)\right]k_1$$

$$+\frac{1}{\sqrt{n}}\left(\frac{\partial b'(\theta_0)}{\partial\theta}W^*\left(b\left(\theta_0\right)\right)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\left[q_1'\frac{\partial^2 b_j(\theta_0)}{\partial\theta\partial\theta'}\right]_{j=1,\ldots,q}W^*\left(b\left(\theta_0\right)\right)\times$$

$$\left[Id_{q\times q}-\frac{\partial b(\theta_0)}{\partial\theta'}\left(\frac{\partial b'(\theta_0)}{\partial\theta}W^*\left(b\left(\theta_0\right)\right)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\frac{\partial b'(\theta_0)}{\partial\theta}W^*\left(b\left(\theta_0\right)\right)\right]k_1\ \blacksquare$$

The following corollary is trivial.

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

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

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

Proof. Simply notice that $\frac{\partial b'(\theta_0)}{\partial \theta}\left(\frac{\partial b'(\theta_0)}{\partial \theta}\frac{1}{n}\sum W^*\left(\, x_i, \theta_0\right)\frac{\partial b(\theta_0)}{\partial \theta'}\right)^{-1}\frac{\partial b'(\theta_0)}{\partial \theta} W^*(\theta_0) = Id_{q\times q}$ when $p = q$. ■

Remark R.35 *From the corollary it is evident that in the case where $b(\theta) = \theta$, 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.*

GMR 2 Estimator We continue with the case of the GMR 2 estimator. Although the caveat met before, that there are non trivial terms in the expansion due to non linearities, due to the relation of the relevant dimensions and due to the presence of stochastic weighting, the expansion contains the term $-E_{\theta_0} k_{2,n}$ something that is not present in the other two, and a fact that is attributed to the computation of $E_\theta \beta_n$. This result that it is known from the work of [11] and [8] 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 to the expansion result, concerns the approximation of derivatives of $E_\theta \beta_n$. It follows easily under the framework established by assumption A.10 (see remark R.24) and the results on the auxiliary estimators.

Lemma 3.5

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

$$\left\|\frac{\partial^2}{\partial \theta \partial \theta'}\left(E_\theta \beta_n\right)_j|_{\theta=\theta_0} - \frac{\partial^2 b_j(\theta_0)}{\partial \theta \partial \theta'}\right\| = o(1)\,,\ j = 1, \dots, q$$

Proof. From assumption A.10 and remark R.24 we have that $\left\|D^r\left(E_\theta\left(\beta_n - b(\theta) - \frac{k_{1,n}(\theta)}{\sqrt{n}}\right)\right)|_{\theta=\theta_0}\right\| = \left\|D^r\left(E_\theta\left(\beta_n - b(\theta)\right)\right)|_{\theta=\theta_0}\right\| \le M\left\|E_{\theta_0}\left(\beta_n - b(\theta_0)\right)\right\| = o(1)$, $r = 1, 2$. ■

We are now ready to state the expansion.

Lemma 3.6 *If* $\sqrt{n}\left(\theta_n-\theta_0\right)$ *has a valid Edgeworth expansion of third order, then*

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

where

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

$$\begin{aligned}
q_2 \;=\; & B^{-1}\left(\frac{\partial b\left(\theta_0\right)}{\partial\theta'}\right)'w^*\left(z,\theta_0\right)k_1-B^{-1}w^*\left(z,\theta_0\right)B^{-1}\left(\frac{\partial b\left(\theta_0\right)}{\partial\theta'}\right)'k_1\\
&+B^{-1}\left(\frac{\partial b\left(\theta_0\right)}{\partial\theta'}\right)'W^*\left(b\left(\theta_0\right)\right)\left(k_2-E_{\theta_0}k_2\right)\\
&-\frac{1}{2}B^{-1}\left(\frac{\partial b\left(\theta_0\right)}{\partial\theta'}\right)'W^*\left(b\left(\theta_0\right)\right)\left[tr\left(q_1q_1^{/}\frac{\partial^2}{\partial\theta\partial\theta^{/}}b\left(\theta_0\right)_j\right)\right]_{j=1,\dots,q}\\
&+B^{-1}\frac{\partial b^{/}\left(\theta_0\right)}{\partial\theta}\left[\frac{\partial}{\partial\theta^{/}}W^*_{rj}\left(b\left(\theta_0\right)\right)q_1^*\right]_{r,i=1,\dots,q}\\
&\times\left[Id_{q\times q}-\frac{\partial}{\partial\theta^{/}}E_{\theta_0}\beta_nB^{-1}\frac{\partial}{\partial\theta}b^{/}\left(\theta_0\right)W^*\left(b\left(\theta_0\right)\right)\right]k_1\\
&+B^{-1}\left[\frac{\partial^2b\left(\theta_0\right)_j}{\partial\theta\partial\theta'}q_1\right]_{j=1,\dots,q}W^*\left(b\left(\theta_0\right)\right)\\
&\times\left[Id_{q\times q}-\frac{\partial b\left(\theta_0\right)}{\partial\theta'}B^{-1}\frac{\partial b^{/}\left(\theta_0\right)}{\partial\theta}W^*\left(b\left(\theta_0\right)\right)\right]k_1,
\end{aligned}$$

$B=\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)'W^*\left(b\left(\theta_0\right)\right)\frac{\partial b(\theta_0)}{\partial\theta'}$, *and* q_1^* *is the relevant term of the analogous expansion of the first step auxiliary estimator due to assumption A.16.*

Proof. Employing again the procedure as in the relevant proofs before and utilizing assumption A.16, remark R.26, lemmas 3.1, 3.2 and 3.5 we have that $\frac{\partial}{\partial\theta}E_{\theta_n}\left(\beta_n^{/}\right)\frac{1}{n}\sum W^*\left(x_i,\theta_n^*\right)\sqrt{n}\left(\beta_n-E_{\theta_n}\beta_n\right)=\mathbf{0}_p\Rightarrow$

$$\left(\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}+\frac{1}{\sqrt{n}}\left[\frac{\partial^2}{\partial\theta\partial\theta'}\left(E_{\theta_0}\beta_n\right)_j\sqrt{n}\left(\theta_n-\theta_0\right)\right]_{j=1,\dots,q}\right)\times$$

$$\left(\frac{1}{n}\sum W^*\left(x_i,\theta_0\right)+\frac{1}{\sqrt{n}}\left[\frac{\partial}{\partial\theta^{/}}\frac{1}{n}\sum W^*_{rj}\left(x_i,\theta_0\right)\sqrt{n}\left(\theta_n^*-\theta_0\right)\right]_{r,i=1,\dots,q}\right)$$

$$\begin{pmatrix}\sqrt{n}\left(\beta_n-b\left(\theta_0\right)\right)-\sqrt{n}\left(E_{\theta_0}\beta_n-b\left(\theta_0\right)\right)-\frac{\partial}{\partial\theta^{/}}E_{\theta_0}\beta_n\sqrt{n}\left(\theta_n-\theta_0\right)\\-\frac{1}{2\sqrt{n}}\left[\sqrt{n}\left(\theta_n-\theta_0\right)^{/}\frac{\partial^2}{\partial\theta\partial\theta^{/}}E_{\theta_0}\left(\beta_n\right)_j\sqrt{n}\left(\theta_n-\theta_0\right)\right]_{j=1,\dots,q}\end{pmatrix}=\mathbf{0}_p\Rightarrow$$

$$\begin{pmatrix} \frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\frac{1}{n}\sum W^*(x_i,\theta_0) + \frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\left[\frac{\partial}{\partial\theta^{/}}\frac{1}{n}\sum W_{rj}^*(x_i,\theta_0)\sqrt{n}(\theta_n^*-\theta_0)\right]_{r,i=1,\ldots,q} \\ +\frac{1}{\sqrt{n}}\left[\frac{\partial^2}{\partial\theta\partial\theta'}(E_{\theta_0}\beta_n)_j\sqrt{n}(\theta_n-\theta_0)\right]_{j=1,\ldots,q}\frac{1}{n}\sum W^*(x_i,\theta_0) \end{pmatrix}\times$$
$$\begin{pmatrix} k_1+\frac{k_2-E_{\theta_0}k_2}{\sqrt{n}}-\frac{\partial}{\partial\theta^{/}}E_{\theta_0}\beta_n\sqrt{n}(\theta_n-\theta_0) \\ -\frac{1}{2\sqrt{n}}\left[\sqrt{n}(\theta_n-\theta_0)^{/}\frac{\partial^2}{\partial\theta\partial\theta^{/}}E_{\theta_0}(\beta_n)_j\sqrt{n}(\theta_n-\theta_0)\right]_{j=1,\ldots,q} \end{pmatrix}=\mathbf{0}_p =>$$
$$\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\frac{1}{n}\sum W^*(x_i,\theta_0)k_1+\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\frac{1}{n}\sum W^*(x_i,\theta_0)\frac{k_2-E_{\theta_0}k_2}{\sqrt{n}}-$$
$$\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\frac{1}{n}\sum W^*(x_i,\theta_0)\frac{\partial}{\partial\theta^{/}}E_{\theta_0}\beta_n\sqrt{n}(\theta_n-\theta_0)$$
$$-\frac{1}{2\sqrt{n}}\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\frac{1}{n}\sum W^*(x_i,\theta_0)\left[tr\left(\sqrt{n}(\theta_n-\theta_0)\sqrt{n}(\theta_n-\theta_0)^{/}\frac{\partial^2}{\partial\theta\partial\theta^{/}}E_{\theta_0}(\beta_n)_j\right)\right]_{j=1,\ldots,q}$$
$$+\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\left[\frac{\partial}{\partial\theta^{/}}\frac{1}{n}\sum W_{rj}^*(x_i,\theta_0)q_1^*\right]_{r,i=1,\ldots,q}k_1+$$
$$\frac{1}{\sqrt{n}}\left[\frac{\partial^2}{\partial\theta\partial\theta'}(E_{\theta_0}\beta_n)_j\sqrt{n}(\theta_n-\theta_0)\right]_{j=1,\ldots,q}\frac{1}{n}\sum W^*(x_i,\theta_0)k_1$$
$$-\begin{pmatrix} \frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\left[\frac{\partial}{\partial\theta^{/}}\frac{1}{n}\sum W_{rj}^*(x_i,\theta_0)q_1^*\right]_{r,i=1,\ldots,q} \\ +\frac{1}{\sqrt{n}}\left[\frac{\partial^2}{\partial\theta\partial\theta'}(E_{\theta_0}\beta_n)_j\sqrt{n}(\theta_n-\theta_0)\right]_{j=1,\ldots,q}\frac{1}{n}\sum W^*(x_i,\theta_0) \end{pmatrix}\frac{\partial}{\partial\theta^{/}}E_{\theta_0}\beta_n\sqrt{n}(\theta_n-\theta_0)=$$
$$\mathbf{0}_p =>$$
$$\begin{pmatrix} \frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\frac{1}{n}\sum W^*(x_i,\theta_0)\frac{\partial}{\partial\theta^{/}}E_{\theta_0}\beta_n \\ +\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\left[\frac{\partial}{\partial\theta^{/}}\frac{1}{n}\sum W_{rj}^*(x_i,\theta_0)q_1^*\right]_{r,i=1,\ldots,q}\frac{\partial}{\partial\theta^{/}}E_{\theta_0}\beta_n \end{pmatrix}\sqrt{n}(\theta_n-\theta_0)=$$
$$\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\frac{1}{n}\sum W^*(x_i,\theta_0)k_1+\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\frac{1}{n}\sum W^*(x_i,\theta_0)\frac{k_2-E_{\theta_0}k_2}{\sqrt{n}}-$$
$$\frac{1}{2\sqrt{n}}\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\frac{1}{n}\sum W^*(x_i,\theta_0)\left[tr\left(\sqrt{n}(\theta_n-\theta_0)\sqrt{n}(\theta_n-\theta_0)^{/}\frac{\partial^2}{\partial\theta\partial\theta^{/}}E_{\theta_0}(\beta_n)_j\right)\right]_{j=1,\ldots,q}$$
$$+\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\left[\frac{\partial}{\partial\theta^{/}}\frac{1}{n}\sum W_{rj}^*(x_i,\theta_0)q_1^*\right]_{r,i=1,\ldots,q}k_1$$
$$+\frac{1}{\sqrt{n}}\left[\frac{\partial^2}{\partial\theta\partial\theta'}(E_{\theta_0}\beta_n)_j\sqrt{n}(\theta_n-\theta_0)\right]_{j=1,\ldots,q}\frac{1}{n}\sum W^*(x_i,\theta_0)k_1$$
$$-\frac{1}{\sqrt{n}}\left[\frac{\partial^2}{\partial\theta\partial\theta'}(E_{\theta_0}\beta_n)_j\sqrt{n}(\theta_n-\theta_0)\right]_{j=1,\ldots,q}\frac{1}{n}\sum W^*(x_i,\theta_0)\frac{\partial}{\partial\theta^{/}}E_{\theta_0}\beta_n\sqrt{n}(\theta_n-\theta_0)=>$$
$$\sqrt{n}(\theta_n-\theta_0)=-\frac{1}{\sqrt{n}}\left(\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\frac{1}{n}\sum W^*(x_i,\theta_0)\frac{\partial}{\partial\theta^{/}}E_{\theta_0}\beta_n\right)^{-1}$$
$$\times\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\left[\frac{\partial}{\partial\theta^{/}}\frac{1}{n}\sum W_{rj}^*(x_i,\theta_0)q_1^*\right]_{r,i=1,\ldots,q}\frac{\partial}{\partial\theta^{/}}E_{\theta_0}\beta_n\sqrt{n}(\theta_n-\theta_0)$$
$$+\left(\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\frac{1}{n}\sum W^*(x_i,\theta_0)\frac{\partial}{\partial\theta^{/}}E_{\theta_0}\beta_n\right)^{-1}\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\frac{1}{n}\sum W^*(x_i,\theta_0)k_1$$
$$+\left(\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\frac{1}{n}\sum W^*(x_i,\theta_0)\frac{\partial}{\partial\theta^{/}}E_{\theta_0}\beta_n\right)^{-1}\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\frac{1}{n}\sum W^*(x_i,\theta_0)\frac{k_2-E_{\theta_0}k_2}{\sqrt{n}}$$
$$-\frac{1}{2\sqrt{n}}\left(\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\frac{1}{n}\sum W^*(x_i,\theta_0)\frac{\partial}{\partial\theta^{/}}E_{\theta_0}\beta_n\right)^{-1}\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\frac{1}{n}\sum W^*(x_i,\theta_0)$$
$$\times\left[tr\left(\sqrt{n}(\theta_n-\theta_0)\sqrt{n}(\theta_n-\theta_0)^{/}\frac{\partial^2}{\partial\theta\partial\theta^{/}}E_{\theta_0}(\beta_n)_j\right)\right]_{j=1,\ldots,q}$$

$$+\frac{1}{\sqrt{n}}\left(\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\frac{1}{n}\sum W^*(x_i,\theta_0)\frac{\partial}{\partial\theta^{/}}E_{\theta_0}\beta_n\right)^{-1}\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\left[\frac{\partial}{\partial\theta^{/}}\frac{1}{n}\sum W_{rj}^*(x_i,\theta_0)q_1^*\right]_{r,i=1,\dots,q}k_1$$
$$+\frac{1}{\sqrt{n}}\left(\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\frac{1}{n}\sum W^*(x_i,\theta_0)\frac{\partial}{\partial\theta^{/}}E_{\theta_0}\beta_n\right)^{-1}$$
$$\times\left[\frac{\partial^2}{\partial\theta\partial\theta'}(E_{\theta_0}\beta_n)_j\sqrt{n}(\theta_n-\theta_0)\right]_{j=1,\dots,q}\frac{1}{n}\sum W^*(x_i,\theta_0)k_1$$
$$-\frac{1}{\sqrt{n}}\left(\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\frac{1}{n}\sum W^*(x_i,\theta_0)\frac{\partial}{\partial\theta^{/}}E_{\theta_0}\beta_n\right)^{-1}\left[\frac{\partial^2}{\partial\theta\partial\theta'}(E_{\theta_0}\beta_n)_j\sqrt{n}(\theta_n-\theta_0)\right]_{j=1,\dots,q}$$
$$\times\frac{1}{n}\sum W^*(x_i,\theta_0)\frac{\partial}{\partial\theta^{/}}E_{\theta_0}\beta_n\sqrt{n}(\theta_n-\theta_0) =>$$
$$\sqrt{n}(\theta_n-\theta_0)=\left(\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\frac{1}{n}\sum W^*(x_i,\theta_0)\frac{\partial}{\partial\theta^{/}}E_{\theta_0}\beta_n\right)^{-1}\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\frac{1}{n}\sum W^*(x_i,\theta_0)k_1$$
$$+\frac{1}{\sqrt{n}}\left(\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\frac{1}{n}\sum W^*(x_i,\theta_0)\frac{\partial}{\partial\theta^{/}}E_{\theta_0}\beta_n\right)^{-1}\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\frac{1}{n}\sum W^*(x_i,\theta_0)(k_2-E_{\theta_0}k_2)$$
$$-\frac{1}{2\sqrt{n}}\left(\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\frac{1}{n}\sum W^*(x_i,\theta_0)\frac{\partial}{\partial\theta^{/}}E_{\theta_0}\beta_n\right)^{-1}\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\times$$
$$\frac{1}{n}\sum W^*(x_i,\theta_0)\left[tr\left(q_1q_1^{/}\frac{\partial^2}{\partial\theta\partial\theta'}E_{\theta_0}(\beta_n)_j\right)\right]_{j=1,\dots,q}$$
$$+\frac{1}{\sqrt{n}}\left(\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\frac{1}{n}\sum W^*(x_i,\theta_0)\frac{\partial}{\partial\theta^{/}}E_{\theta_0}\beta_n\right)^{-1}\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\left[\frac{\partial}{\partial\theta^{/}}\frac{1}{n}\sum W_{rj}^*(x_i,\theta_0)q_1^*\right]_{r,i=1,\dots,q}\times$$
$$\left[Id_{q\times q}-\frac{\partial}{\partial\theta^{/}}E_{\theta_0}\beta_n\left(\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\frac{1}{n}\sum W^*(x_i,\theta_0)\frac{\partial}{\partial\theta^{/}}E_{\theta_0}\beta_n\right)^{-1}\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\frac{1}{n}\sum W^*(x_i,\theta_0)\right]k_1$$
$$+\frac{1}{\sqrt{n}}\left(\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\frac{1}{n}\sum W^*(x_i,\theta_0)\frac{\partial}{\partial\theta^{/}}E_{\theta_0}\beta_n\right)^{-1}\left[\frac{\partial^2}{\partial\theta\partial\theta'}(E_{\theta_0}\beta_n)_j q_1\right]_{j=1,\dots,q}\frac{1}{n}\sum W^*(x_i,\theta_0)\times$$
$$\left[Id_{q\times q}-\frac{\partial}{\partial\theta^{/}}E_{\theta_0}\beta_n\left(\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\frac{1}{n}\sum W^*(x_i,\theta_0)\frac{\partial}{\partial\theta^{/}}E_{\theta_0}\beta_n\right)^{-1}\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}\frac{1}{n}\sum W^*(x_i,\theta_0)\right]k_1$$

Now $\frac{\partial}{\partial\theta}E_{\theta_0}\beta_n^{/}=\left(\frac{\partial b(\theta_0)}{\partial\theta'}+o(1)\right)'=\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)'+B_n$ with $\|B_n\|=o(1)$ and $\frac{1}{n}\sum W^*(x_i,\theta_0)=W^*(b(\theta_0))+\frac{1}{\sqrt{n}}w^*(z,\theta_0)$, where $z\sim N(0,\Sigma)$ and the elements of $w^*(z,\theta_0)$ are finite polynomials in z, it follows that

$$\sqrt{n}(\theta_n-\theta_0)=\left(\left(\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)'+B_n\right)\left(W^*(b(\theta_0))+\frac{1}{\sqrt{n}}w^*(z,\theta_0)\right)\left(\frac{\partial b(\theta_0)}{\partial\theta'}+B_n\right)\right)^{-1}$$
$$\times\left(\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)'+B_n\right)\left(W^*(b(\theta_0))+\frac{1}{\sqrt{n}}w^*(z,\theta_0)\right)k_1$$
$$+\frac{1}{\sqrt{n}}\left(\left(\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)'+B_n\right)\left(W^*(b(\theta_0))+\frac{1}{\sqrt{n}}w^*(z,\theta_0)\right)\left(\frac{\partial b(\theta_0)}{\partial\theta'}+B_n\right)\right)^{-1}$$
$$\times\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)'W^*(b(\theta_0))(k_2-E_{\theta_0}k_2)$$
$$-\frac{1}{2\sqrt{n}}\left(\left(\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)'+B_n\right)\left(W^*(b(\theta_0))+\frac{1}{\sqrt{n}}w^*(z,\theta_0)\right)\left(\frac{\partial b(\theta_0)}{\partial\theta'}+B_n\right)\right)^{-1}\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)'$$
$$\times W^*(b(\theta_0))\left[tr\left(q_1q_1^{/}\frac{\partial^2}{\partial\theta\partial\theta'}E_{\theta_0}(\beta_n)_j\right)\right]_{j=1,\dots,q}$$

$$+\frac{1}{\sqrt{n}}\left(\left(\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)'+B_n\right)\left(W^*\left(b\left(\theta_0\right)\right)+\frac{1}{\sqrt{n}}w^*\left(z,\theta_0\right)\right)\left(\frac{\partial b(\theta_0)}{\partial\theta'}+B_n\right)\right)^{-1}$$

$$\times\frac{\partial b^{/}(\theta_0)}{\partial\theta}\left[\frac{\partial}{\partial\theta^{/}}\frac{1}{n}\sum W_{rj}^*\left(x_i,\theta_0\right)q_1^*\right]_{r,i=1,\ldots,q}$$

$$\times\left[\begin{array}{c} Id_{q\times q}-\frac{\partial}{\partial\theta^{/}}E_{\theta_0}\beta_n\left(\left(\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)'+B_n\right)\left(W^*\left(b\left(\theta_0\right)\right)+\frac{1}{\sqrt{n}}w^*\left(z,\theta_0\right)\right)\left(\frac{\partial b(\theta_0)}{\partial\theta'}+B_n\right)\right)^{-1} \\ \frac{\partial b^{/}(\theta_0)}{\partial\theta}W^*\left(b\left(\theta_0\right)\right)\end{array}\right]k_1$$

$$+\frac{1}{\sqrt{n}}\left(\left(\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)'+B_n\right)\left(W^*\left(b\left(\theta_0\right)\right)+\frac{1}{\sqrt{n}}w^*\left(z,\theta_0\right)\right)\left(\frac{\partial b(\theta_0)}{\partial\theta'}+B_n\right)\right)^{-1}\left[\frac{\partial^2}{\partial\theta\partial\theta'}\left(E_{\theta_0}\beta_n\right)_j q_1\right]_{j=1,\ldots,q}$$

$$\times\left(W^*\left(b\left(\theta_0\right)\right)+\frac{1}{\sqrt{n}}w^*\left(z,\theta_0\right)\right)$$

$$\times\left[\begin{array}{c} Id_{q\times q}-\frac{\partial b(\theta_0)}{\partial\theta'}\left(\left(\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)'+B_n\right)\left(W^*\left(b\left(\theta_0\right)\right)+\frac{1}{\sqrt{n}}w^*\left(z,\theta_0\right)\right)\left(\frac{\partial b(\theta_0)}{\partial\theta'}+B_n\right)\right)^{-1} \\ \frac{\partial b^{/}(\theta_0)}{\partial\theta}W^*\left(b\left(\theta_0\right)\right)\end{array}\right]k_1$$

As now $\left(\left(\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)'+B_n\right)\left(W^*\left(b\left(\theta_0\right)\right)+\frac{1}{\sqrt{n}}w^*\left(z,\theta_0\right)\right)\left(\frac{\partial b(\theta_0)}{\partial\theta'}+B_n\right)\right)^{-1}=$

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

$$+K_n-\frac{1}{\sqrt{n}}\left(\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)'W^*\left(b\left(\theta_0\right)\right)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}w^*\left(z,\theta_0\right)\left(\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)'W^*\left(b\left(\theta_0\right)\right)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}$$

where $\|K_n\|=o\left(1\right)$, we get:

$$\sqrt{n}\left(\theta_n-\theta_0\right)=\left(\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)'W^*\left(b\left(\theta_0\right)\right)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)'k_1+A_nk_1$$

$$\frac{1}{\sqrt{n}}\left(\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)'W^*\left(b\left(\theta_0\right)\right)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)'w^*\left(z,\theta_0\right)k_1$$

$$-\frac{1}{\sqrt{n}}\left(\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)'W^*\left(b\left(\theta_0\right)\right)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}w^*\left(z,\theta_0\right)\left(\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)'W^*\left(b\left(\theta_0\right)\right)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)'k_1$$

$$+\frac{1}{\sqrt{n}}\left(\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)'W^*\left(b\left(\theta_0\right)\right)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)'W^*\left(b\left(\theta_0\right)\right)\left(k_2-E_{\theta_0}k_2\right)$$

$$-\frac{1}{2\sqrt{n}}\left(\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)'W^*\left(b\left(\theta_0\right)\right)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)'W^*\left(b\left(\theta_0\right)\right)\left[tr\left(q_1q_1^{/}\frac{\partial^2}{\partial\theta\partial\theta^{/}}b\left(\theta_0\right)_j\right)\right]_{j=1,\ldots,q}$$

$$+\frac{1}{\sqrt{n}}\left(\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)'W^*\left(b\left(\theta_0\right)\right)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\frac{\partial b^{/}(\theta_0)}{\partial\theta}\left[\frac{\partial}{\partial\theta^{/}}W_{rj}^*\left(b\left(\theta_0\right)\right)q_1^*\right]_{r,i=1,\ldots,q}\times$$

$$\left[Id_{q\times q}-\frac{\partial}{\partial\theta^{/}}E_{\theta_0}\beta_n\left(\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)'W^*\left(b\left(\theta_0\right)\right)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\frac{\partial}{\partial\theta}b^{/}\left(\theta_0\right)W^*\left(b\left(\theta_0\right)\right)\right]k_1$$

$$+\frac{1}{\sqrt{n}}\left(\left(\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)'W^*\left(b\left(\theta_0\right)\right)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\right)\left[\frac{\partial^2 b(\theta_0)_j}{\partial\theta\partial\theta'}q_1\right]_{j=1,\ldots,q}W^*\left(b\left(\theta_0\right)\right)\times$$

$$\left[Id_{q\times q}-\frac{\partial b(\theta_0)}{\partial\theta'}\left(\left(\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)'W^*\left(b\left(\theta_0\right)\right)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\right)\frac{\partial b'(\theta_0)}{\partial\theta}W^*\left(b\left(\theta_0\right)\right)\right]k_1$$

where $A_n=\left(\left(\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)'W^*\left(b\left(\theta_0\right)\right)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}B_n+K_n\left(\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)'+B_n\right)\right)W^*\left(b\left(\theta_0\right)\right)$ with $\|A_n\|=o\left(1\right)$. ■

Remark R.36 *As expected the two estimators are first order equivalent as their q_1 terms coincide.*

Remark R.37 *The term* $\left(\frac{\partial b(\theta_0)'}{\partial\theta}W^*\left(b\left(\theta_0\right)\right)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\frac{\partial b(\theta_0)'}{\partial\theta}W^*\left(b\left(\theta_0\right)\right)\left(k_2-E_{\theta_0}k_2\right)$ *is obtained due to the presence of $E_\theta\beta_n$ in the definition of the estimator and not of $b\left(\theta\right)$ or something similar as in the cases of GMR 1 and GT estimators.*

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

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

$$q_2=\left(\frac{\partial b\left(\theta_0\right)}{\partial\theta'}\right)^{-1}\left(k_2-E_{\theta_0}k_2\right)-\frac{1}{2}\left(\frac{\partial b\left(\theta_0\right)}{\partial\theta'}\right)^{-1}\left[trq_1q_1'\frac{\partial^2 b_j}{\partial\theta\partial\theta'}\right]_{j=1,\ldots,l}$$

Proof. Trivial. ■

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

Proof. Trivial. ■

Remark R.38 *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 GMR 2 obviously performs a second order bias correction. If in addition $E_\theta\beta_n$ is linear, then the estimator is totally unbiased (see [11]).*

Remark R.39 *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 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 $(\theta_0, b(\theta_0))$. Let $s_n(\theta)$ and $H_n(\theta)$ denote the gradient (score) and the Hessian of the loglikelihood function of $\mathcal{D}$ respectively. In order for the identification of terms, the (local) identity $E_\theta c_n(b(\theta)) = \mathbf{0}_{w_i}$ is differentiated thus providing the following useful lemma.

Lemma 3.7 *If $\sqrt{n}(\theta_n - \theta_0)$ has a valid Edgeworth expansion of third order*

$$E_\theta \frac{\partial c_n(b(\theta))}{\partial \beta'} \frac{\partial b(\theta)}{\partial \theta'} = -E_\theta c_n(b(\theta)) s_n'(\theta)$$

$$E_\theta \frac{\partial^2}{\partial \theta \partial \beta'} c_n(b(\theta))_j = \frac{\partial b'(\theta)}{\partial \theta} E_\theta \frac{\partial^2}{\partial \beta \partial \beta'} c_n(b(\theta))_j = -E_\theta s_n(\theta) \frac{\partial}{\partial \beta'} c_n(b(\theta))_j$$

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

Proof. Notice that for all θ in $\mathcal{O}_{\varepsilon_2}(\theta_0)$ the following identity is well defined $\forall i = 1, 2$

$$E_\theta c_n(b(\theta)) = \mathbf{0}_l$$

using assumptions A.7, A.8. Taking derivatives $\frac{\partial}{\partial \theta'} E_\theta c_n(b(\theta)) = \mathbf{0}_{w \times p} \Rightarrow E_\theta \frac{\partial c_n(b(\theta))}{\partial \beta'} \frac{\partial b(\theta)}{\partial \theta'} + E_\theta c_n(b(\theta)) s_n'(\theta) = \mathbf{0}_{w \times p} \Rightarrow E_\theta \frac{\partial c_n(b(\theta))}{\partial \beta'} \frac{\partial b(\theta)}{\partial \theta'} = -E_\theta c_n(b(\theta)) s_n'(\theta)$. Also, since $\frac{\partial}{\partial \beta} E_\theta c_n(\beta) = E_\theta \frac{\partial}{\partial \beta} c_n(\beta)$, we have $\frac{\partial^2}{\partial \beta \partial \theta'} E_\theta c_n(b(\theta))_j = \mathbf{0}_{q \times p} \Rightarrow \frac{\partial}{\partial \theta'} E_\theta \frac{\partial}{\partial \beta} c_n(b(\theta))_j = \mathbf{0}_{q \times p} \Rightarrow E_\theta \frac{\partial^2}{\partial \beta \partial \beta'} c_n(b(\theta))_j \frac{\partial b(\theta)}{\partial \theta'} + E_\theta \frac{\partial}{\partial \beta} c_n(b(\theta))_j s_n'(\theta) = \mathbf{0}_{q \times p}$
$\Rightarrow E_\theta \frac{\partial^2}{\partial \beta \partial \theta'} c_n(b(\theta))_j \frac{\partial b(\theta)}{\partial \theta'} = E_\theta \frac{\partial^2}{\partial \beta \partial \beta'} c_n(b(\theta))_j \frac{\partial b(\theta)}{\partial \theta'} = -E_\theta \frac{\partial}{\partial \beta} c_n(b(\theta))_j s_n'(\theta)$
and therefore the second result of the lemma follows. Then $\frac{\partial^2}{\partial \theta \partial \theta'} E_\theta c_n(b(\theta))_j = \frac{\partial}{\partial \theta'} \left(\frac{\partial}{\partial \theta} E_\theta c_n(b(\theta))_j \right) = \frac{\partial}{\partial \theta'} \left(E_\theta \frac{\partial c_n(b(\theta))_j}{\partial \theta} + E_\theta c_n(b(\theta))_j s_n(\theta) \right)$
$= E_\theta \frac{\partial^2 c_n(b(\theta))_j}{\partial \theta \partial \theta'} + E_\theta \frac{\partial c_n(b(\theta))_j}{\partial \theta} s_n'(\theta) + E_\theta s_n(\theta) \frac{\partial c_n(b(\theta))_j}{\partial \theta'} + E_\theta c_n(b(\theta))_j H_n(\theta) + E_\theta c_n(b(\theta))_j s_n(\theta) s_n'(\theta)$
$= \frac{\partial \beta'(\theta)}{\partial \theta} \left(E_\theta \frac{\partial^2 c_n(b(\theta))_j}{\partial \beta \partial \beta'} \right) \frac{\partial b(\theta)}{\partial \theta'} + \left[E_\theta \frac{\partial c_n(b(\theta))_j}{\partial \beta'} \frac{\partial^2 \beta(\theta)}{\partial \theta \partial \theta_r} \right]_{r=1,\ldots,p} - 2 \frac{\partial b'(\theta)}{\partial \theta} E_\theta \frac{\partial^2}{\partial \beta \partial \beta'} c_n(b(\theta))_j \frac{\partial b(\theta)}{\partial \theta'}$
$+ E_\theta c_n(b(\theta))_j H_n(\theta) + E_\theta c_n(b(\theta))_j s_n(\theta) s_n'(\theta)$
$= -\frac{\partial \beta'(\theta)}{\partial \theta} \left(E_\theta \frac{\partial^2 c_n(b(\theta))_j}{\partial \beta \partial \beta'} \right) \frac{\partial b(\theta)}{\partial \theta'} + \left[E_\theta \frac{\partial c_n(b(\theta))_j}{\partial \beta'} \frac{\partial^2 \beta(\theta)}{\partial \theta \partial \theta_r} \right]_{r=1,\ldots,p} + E_\theta c_n(b(\theta))_j H_n(\theta) +$

$E_\theta c_n (b(\theta))_j s_n(\theta) s_n'(\theta)$ by the second part of the lemma and therefore since $\frac{\partial^2}{\partial\theta\partial\theta'} E_\theta c_n (b(\theta))_j = \mathbf{0}_{p\times p}$ the third result follows. ■

Lemma 3.8 *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

$$q_1 = J^{-1} \frac{\partial b'(\theta_0)}{\partial \theta} E_{\theta_0} \frac{\partial c_n (b(\theta_0))'}{\partial \beta} W^{**}(b(\theta_0)) E_{\theta_0} \frac{\partial c_n (b(\theta_0))}{\partial \beta'} k_1,$$

$$\begin{aligned}
q_2 \quad = \quad & J^{-1} \frac{\partial b^{/}(\theta_0)}{\partial \theta} E_{\theta_0} \frac{\partial (c_n(b(\theta_0)))^{/}}{\partial \beta} W^{**}(b(\theta_0)) E_{\theta_0} \frac{\partial c_n(b(\theta_0))}{\partial \beta'} k_2 \\
& + J^{-1} A E_{\theta_0} \frac{\partial c_n(b(\theta_0))}{\partial \beta'} B k_1 \\
& + J^{-1} \frac{\partial b^{/}(\theta_0)}{\partial \theta} E_{\theta_0} \frac{\partial (c_n(b(\theta_0)))^{/}}{\partial \beta} \left[\frac{\partial}{\partial \theta^{/}} W_{rj}^{**}(b(\theta_0)) q_1^* \right]_{r,j=1,\dots,l} \\
& \times E_{\theta_0} \frac{\partial c_n(b(\theta_0))}{\partial \beta'} \left(k_1 - \frac{\partial b(\theta_0)}{\partial \theta^{/}} q_1 \right) \\
& - J^{-1} \left(\left[q_1^{/} \left(\begin{array}{c} \frac{\partial b^{/}(\theta_0)}{\partial \theta} E_{\theta_0} \frac{\partial^2 c_n(b(\theta_0))_j}{\partial \beta \partial \beta'} \frac{\partial b(\theta_0)}{\partial \theta'} \\ - \left[E_{\theta_0} \frac{\partial c_n(b(\theta_0))_j}{\partial \beta'} \frac{\partial b(\theta_0)}{\partial \theta_r \partial \theta'} \right]_{r=1,\dots p} \end{array} \right) \right]_{j=1,\dots,l} \right)^{/} W^{**}(b(\theta_0)) \\
& \times E_{\theta_0} \frac{\partial c_n(b(\theta_0))}{\partial \beta'} \left(k_1 - \frac{\partial b(\theta_0)}{\partial \theta^{/}} q_1 \right) \\
& + \frac{1}{2} J^{-1} \frac{\partial b^{/}(\theta_0)}{\partial \theta} E_{\theta_0} \frac{\partial (c_n(b(\theta_0)))^{/}}{\partial \beta} W^{**}(b(\theta_0)) \\
& \times \left(\begin{array}{c} \left[tr k_1 k_1^{/} E_{\theta_0} \frac{\partial^2 c_n(b(\theta_0))_j}{\partial \beta \partial \beta'} \right]_{j=1,\dots,l} \\ - \left[tr q_1 q_1^{/} \left[E_{\theta_0} \frac{\partial c_{i,n}^*(b(\theta_0))_j}{\partial \beta'} \frac{\partial^2 b(\theta_0)}{\partial \theta_r \partial \theta'} \right]_{r=1,\dots p} \right]_{j=1,\dots,l} \end{array} \right) \\
& + \frac{1}{2} J^{-1} \frac{\partial b^{/}(\theta_0)}{\partial \theta} E_{\theta_0} \frac{\partial (c_n(b(\theta_0)))^{/}}{\partial \beta} W^{**}(b(\theta_0)) \\
& \times \left(\begin{array}{c} \left[tr q_1 q_1^{/} \frac{\partial b^{/}(\theta_0)}{\partial \theta} E_{\theta_0} \frac{\partial^2}{\partial \beta \partial \beta'} c_n(b(\theta_0))_j \frac{\partial b(\theta_0)}{\partial \theta'} \right]_{j=1,\dots,l} \\ -2 \left[tr k_1 q_1^{/} \frac{\partial b'(\theta_0)}{\partial \theta} E_{\theta_0} \frac{\partial^2 c_n(b(\theta_0))_j}{\partial \beta \partial \beta'} \right]_{j=1,\dots,l} \end{array} \right),
\end{aligned}$$

$$
\begin{aligned}
J &= \frac{\partial b'(\theta_0)}{\partial \theta} E_{\theta_0} \frac{\partial c_{i,n}(b(\theta_0))'}{\partial \beta} W^{**}(b(\theta_0)) E_{\theta_0} \frac{\partial c_{i,n}(b(\theta_0))}{\partial \beta'} \frac{\partial b(\theta_0)}{\partial \theta'}, \\
A &= \frac{\partial b^{/}(\theta_0)}{\partial \theta} E_{\theta_0} \frac{\partial (c_n(b(\theta_0)))^{/}}{\partial \beta} w^{**}(z,\theta_0) \\
&+ \left(\left[k_1' E_{\theta_0} \frac{\partial^2 c_n(b(\theta_0))_j}{\partial \beta \partial \beta'} \frac{\partial b(\theta_0)}{\partial \theta'} \right]_{j=1,\dots,l} \right)^{/} W^{**}(b(\theta_0)), \\
B &= Id_{q\times q} - \frac{\partial b(\theta_0)}{\partial \theta'} J^{-1} \frac{\partial b^{/}(\theta_0)}{\partial \theta} E_{\theta_0} \frac{\partial (c_n(b(\theta_0)))^{/}}{\partial \beta} W^{**}(b(\theta_0)) E_{\theta_0} \frac{\partial c_n(b(\theta_0))}{\partial \beta'},
\end{aligned}
$$

and q_1^* is the relevant term of the analogous expansion of the first step auxiliary estimator due to assumption A.16.

Proof. Utilizing assumption A.16, remark R.26, lemmas 3.1, 3.2 and 3.7 we have that $\frac{\partial E_\theta(c_n(\beta)/\sigma(\beta))}{\partial \theta'} = E_\theta(c_n(\beta) s_n'(\theta)) =$
$E_{\theta_0} c_n(b(\theta_0)) s_n'(\theta_0) + \frac{1}{\sqrt{n}} \left[\sqrt{n}(\beta - b(\theta_0))' E_{\theta_0} \frac{\partial c_n(b(\theta_0))_j}{\partial \beta} s_n'(\theta_0) \right]_{j=1,\dots,l}$
$+ \frac{1}{\sqrt{n}} \left[\sqrt{n}(\theta - \theta_0)' \left(E_{\theta_0} c_n(b(\theta_0))_j s_n(\theta_0) s_n'(\theta_0) + E_{\theta_0} c_n(b(\theta_0))_j H_n(\theta_0) \right) \right]_{j=1,\dots,l}$.
Hence $E_{\theta_n}(c_n(\beta_n) s_n'(\theta_n)/\sigma(\beta_n)) =$
$E_{\theta_0} c_n(b(\theta_0)) s_n'(\theta_0) + \frac{1}{\sqrt{n}} \left[\sqrt{n}(\beta_n - b(\theta_0))' E_{\theta_0} \frac{\partial c_n(b(\theta_0))_j}{\partial \beta} s_n'(\theta_0) \right]_{j=1,\dots,l}$
$+ \frac{1}{\sqrt{n}} \left[\sqrt{n}(\theta_n - \theta_0)' \left(E_{\theta_0} c_n(b(\theta_0))_j s_n(\theta_0) s_n'(\theta_0) + E_{\theta_0} c_n(b(\theta_0))_j H_n(\theta_0) \right) \right]_{j=1,\dots,l} =$
$= -E_{\theta_0} \frac{\partial c_n(b(\theta_0))}{\partial \beta'} \frac{\partial b(\theta_0)}{\partial \theta'} - \frac{1}{\sqrt{n}} \left[k_1' E_{\theta_0} \frac{\partial^2}{\partial \beta \partial \beta'} c_n(b(\theta_0))_j \frac{\partial b(\theta_0)}{\partial \theta'} \right]_{j=1,\dots,w_i}$
$+ \frac{1}{\sqrt{n}} \left[\sqrt{n}(\theta_n - \theta_0)' \left(\frac{\partial b^{/}(\theta_0)}{\partial \theta} E_{\theta_0} \frac{\partial^2}{\partial \beta \partial \beta'} c_n(b(\theta_0))_j \frac{\partial b(\theta_0)}{\partial \theta'} - \left[E_{\theta_0} \frac{\partial c_n(b(\theta_0))_j}{\partial \beta'} \frac{\partial b(\theta_0)}{\partial \theta_r \partial \theta'} \right]_{r=1,\dots p} \right) \right]_{j=1,\dots,l}$
Also $\sqrt{n} E_\theta(c_n(\beta)/\sigma(\beta)) = E_{\theta_0} \frac{\partial c_n(b(\theta_0))}{\partial \beta'} \sqrt{n}(\beta - b(\theta_0)) + E_{\theta_0} c_n(b(\theta_0)) s_n'(\theta_0) \sqrt{n}(\theta - \theta_0)$
$+ \frac{1}{2\sqrt{n}} \left[tr \sqrt{n}(\beta - b(\theta_0)) \sqrt{n}(\beta - b(\theta_0))' E_{\theta_0} \frac{\partial^2 c_n(b(\theta_0))_j}{\partial \beta \partial \beta'} \right]_{j=1,\dots,l}$
$+ \frac{1}{2\sqrt{n}} \left[tr \sqrt{n}(\beta - b(\theta_0)) \sqrt{n}(\theta - \theta_0)' E_{\theta_0} s_n(\theta_0) \frac{\partial c_n(b(\theta_0))}{\partial \beta'} \right]_{j=1,\dots,l}$
$+ \frac{1}{2\sqrt{n}} \left[tr \sqrt{n}(\theta - \theta_0) \sqrt{n}(\beta - b(\theta_0))' E_{\theta_0} \frac{\partial c_n(b(\theta_0))_j}{\partial \beta} s_n'(\theta_0) \right]_{j=1,\dots,l}$
$+ \frac{1}{2\sqrt{n}} \left[tr \sqrt{n}(\theta - \theta_0) \sqrt{n}(\theta - \theta_0)' \times \begin{pmatrix} E_{\theta_0} c_n(b(\theta_0))_j s_n(\theta_0) s_n'(\theta_0) \\ + E_{\theta_0} c_n(b(\theta_0))_j H_n(\theta_0) \end{pmatrix} \right]_{j=1,\dots,l}$
and it follows that
$\sqrt{n} E_{\theta_n}(c_n(\beta_n)/\sigma(\beta_n)) = E_{\theta_0} \frac{\partial c_n(b(\theta_0))}{\partial \beta'} \sqrt{n}(\beta_n - b(\theta_0)) + E_{\theta_0} c_n(b(\theta_0)) s_n'(\theta_0) \sqrt{n}(\theta_n - \theta_0)$
$+ \frac{1}{2\sqrt{n}} \left[tr \sqrt{n}(\beta_n - b(\theta_0)) \sqrt{n}(\beta_n - b(\theta_0))' E_{\theta_0} \frac{\partial^2 c_n(b(\theta_0))_j}{\partial \beta \partial \beta'} \right]_{j=1,\dots,l}$

$$+\frac{1}{2\sqrt{n}}\left[tr\sqrt{n}\left(\beta_n-b\left(\theta_0\right)\right)\sqrt{n}\left(\theta-\theta_0\right)'E_{\theta_0}s_n\left(\theta_0\right)\frac{\partial c_n(b(\theta_0))}{\partial\beta'}\right]_{j=1,\dots,l}$$
$$+\frac{1}{2\sqrt{n}}\left[tr\sqrt{n}\left(\theta_n-\theta_0\right)\sqrt{n}\left(\beta_n-b\left(\theta_0\right)\right)'E_{\theta_0}\frac{\partial c_n(b(\theta_0))_j}{\partial\beta}s_n'\left(\theta_0\right)\right]_{j=1,\dots,l}$$
$$+\frac{1}{2\sqrt{n}}\left[tr\sqrt{n}\left(\theta_n-\theta_0\right)\sqrt{n}\left(\theta_n-\theta_0\right)'\times\begin{pmatrix}E_{\theta_0}c_n\left(b\left(\theta_0\right)\right)_j s_n\left(\theta_0\right)s_n'\left(\theta_0\right)\\+E_{\theta_0}c_n\left(b\left(\theta_0\right)\right)_j H_n\left(\theta_0\right)\end{pmatrix}\right]_{j=1,\dots,l}$$

Now, we have that

$$\frac{1}{n}\sum_i W^{**}\left(x_i,\theta_n^*\right)=\frac{1}{n}\sum W^{**}\left(x_i,\theta_0\right)+\frac{1}{\sqrt{n}}\left[\frac{\partial}{\partial\theta'}\frac{1}{n}\sum W_{rj}^{**}\left(x_i,\theta_0\right)\sqrt{n}\left(\theta_n^*-\theta_0\right)\right]_{r,j=1,\dots,l}$$

and from the first order conditions we have that

$$\frac{\partial E_{\theta_n}\left(c_n\left(\beta_{i,n}\right)\right)'}{\partial\theta}\frac{1}{n}\sum_i W^{**}\left(x_i,\theta_n^*\right)\sqrt{n}E_{\theta_n}\left(c_n\left(\beta_{i,n}\right)\right)=0_{p\times1}=>$$

$$\begin{bmatrix}-E_{\theta_0}\frac{\partial c_n(b(\theta_0))}{\partial\beta'}\frac{\partial b(\theta_0)}{\partial\theta'}-\frac{1}{\sqrt{n}}\left[k_1'E_{\theta_0}\frac{\partial^2}{\partial\beta\partial\beta'}c_n\left(b\left(\theta_0\right)\right)_j\frac{\partial b(\theta_0)}{\partial\theta'}\right]_{j=1,\dots,l}\\+\frac{1}{\sqrt{n}}\left[\sqrt{n}\left(\theta_n-\theta_0\right)'\begin{pmatrix}\frac{\partial b'(\theta_0)}{\partial\theta}E_{\theta_0}\frac{\partial^2}{\partial\beta\partial\beta'}c_n\left(b\left(\theta_0\right)\right)_j\frac{\partial b(\theta_0)}{\partial\theta'}\\-\left[E_{\theta_0}\frac{\partial c_n(b(\theta_0))_j}{\partial\beta'}\frac{\partial b(\theta_0)}{\partial\theta_r\partial\theta'}\right]_{r=1,\dots p}\end{pmatrix}\right]_{j=1,\dots,l}\end{bmatrix}'$$
$$\times\left(\frac{1}{n}\sum W^{**}\left(x_i,\theta_0\right)+\frac{1}{\sqrt{n}}\left[\frac{\partial}{\partial\theta'}\frac{1}{n}\sum W_{rj}^{**}\left(x_i,\theta_0\right)\sqrt{n}\left(\theta_n^*-\theta_0\right)\right]_{r,j=1,\dots,l}\right)$$
$$\times\begin{bmatrix}E_{\theta_0}\frac{\partial c_n(b(\theta_0))}{\partial\beta'}\sqrt{n}\left(\beta_n-b\left(\theta_0\right)\right)+E_{\theta_0}c_n\left(b\left(\theta_0\right)\right)s_n'\left(\theta_0\right)\sqrt{n}\left(\theta_n-\theta_0\right)\\+\frac{1}{2\sqrt{n}}\left[tr\sqrt{n}\left(\beta_n-b\left(\theta_0\right)\right)\sqrt{n}\left(\beta_n-b\left(\theta_0\right)\right)'E_{\theta_0}\frac{\partial^2 c_n(b(\theta_0))_j}{\partial\beta\partial\beta'}\right]_{j=1,\dots,l}\\+\frac{1}{2\sqrt{n}}\left[tr\sqrt{n}\left(\beta_n-b\left(\theta_0\right)\right)\sqrt{n}\left(\theta-\theta_0\right)'E_{\theta_0}s_n\left(\theta_0\right)\frac{\partial c_n(b(\theta_0))_j}{\partial\beta'}\right]_{j=1,\dots,l}\\+\frac{1}{2\sqrt{n}}\left[tr\sqrt{n}\left(\theta_n-\theta_0\right)\sqrt{n}\left(\beta_n-b\left(\theta_0\right)\right)'E_{\theta_0}\frac{\partial c_n(b(\theta_0))_j}{\partial\beta}s_n'\left(\theta_0\right)\right]_{j=1,\dots,l}\\+\frac{1}{2\sqrt{n}}\left[tr\sqrt{n}\left(\theta_n-\theta_0\right)\sqrt{n}\left(\theta_n-\theta_0\right)'\times\begin{pmatrix}E_{\theta_0}c_n\left(b\left(\theta_0\right)\right)_j s_n\left(\theta_0\right)s_n'\left(\theta_0\right)\\+E_{\theta_0}c_n\left(b\left(\theta_0\right)\right)_j H_n\left(\theta_0\right)\end{pmatrix}\right]_{j=1,\dots,l}\end{bmatrix}=$$
$$0_{p\times1}=>$$

As now $\left[tr\sqrt{n}\left(\beta_n-b\left(\theta_0\right)\right)\sqrt{n}\left(\theta-\theta_0\right)'E_{\theta_0}s_n\left(\theta_0\right)\frac{\partial c_n(b(\theta_0))_j}{\partial\beta'}\right]_{j=1,\dots,l}=$ $\left[tr\sqrt{n}\left(\theta_n-\theta_0\right)\sqrt{n}\left(\beta_n-b\left(\theta_0\right)\right)'E_{\theta_0}\frac{\partial c_n(b(\theta_0))_j}{\partial\beta}s_n'\left(\theta_0\right)\right]_{j=1,\dots,l}$, since $\sqrt{n}\left(\beta_n-b\left(\theta_0\right)\right)\underset{1/2}{\sim}k_1+\frac{1}{\sqrt{n}}k_2$ and $-E_\theta\frac{\partial c_n(b(\theta))}{\partial\beta'}\frac{\partial b(\theta)}{\partial\theta'}=E_{\theta_0}c_n\left(b\left(\theta_0\right)\right)s_n'\left(\theta_0\right)$ we have:

$$\begin{bmatrix}-\frac{\partial b'(\theta_0)}{\partial\theta}E_{\theta_0}\frac{\partial(c_n(b(\theta_0)))'}{\partial\beta}\frac{1}{n}\sum W^{**}\left(x_i,\theta_0\right)\\-\frac{1}{\sqrt{n}}\frac{\partial b'(\theta_0)}{\partial\theta}E_{\theta_0}\frac{\partial(c_n(b(\theta_0)))'}{\partial\beta}\left[\frac{\partial}{\partial\theta'}\frac{1}{n}\sum W_{rj}^{**}\left(x_i,\theta_0\right)\sqrt{n}\left(\theta_n^*-\theta_0\right)\right]_{r,j=1,\dots,l}\\-\frac{1}{\sqrt{n}}\left(\left[k_1'E_{\theta_0}\frac{\partial^2}{\partial\beta\partial\beta'}c_n\left(b\left(\theta_0\right)\right)_j\frac{\partial b(\theta_0)}{\partial\theta'}\right]_{j=1,\dots,l}\right)'\frac{1}{n}\sum W^{**}\left(x_i,\theta_0\right)\\+\frac{1}{\sqrt{n}}\left(\left[\sqrt{n}\left(\theta_n-\theta_0\right)'\begin{pmatrix}\frac{\partial b'(\theta_0)}{\partial\theta}E_{\theta_0}\frac{\partial^2}{\partial\beta\partial\beta'}c_n\left(b\left(\theta_0\right)\right)_j\frac{\partial b(\theta_0)}{\partial\theta'}\\-\left[E_{\theta_0}\frac{\partial c_n(b(\theta_0))_j}{\partial\beta'}\frac{\partial b(\theta_0)}{\partial\theta_r\partial\theta'}\right]_{r=1,\dots p}\end{pmatrix}\right]_{j=1,\dots,l}\right)'\frac{1}{n}\sum W^{**}\left(x_i,\theta_0\right)\end{bmatrix}$$

$$\times \left[\begin{array}{c} E_{\theta_0}\frac{\partial c_n(b(\theta_0))}{\partial \beta'}\left(k_1+\frac{1}{\sqrt{n}}k_2\right)-E_{\theta}\frac{\partial c_n(b(\theta))}{\partial \beta'}\frac{\partial b(\theta)}{\partial \theta'}\sqrt{n}\left(\theta_n-\theta_0\right) \\ +\frac{1}{2\sqrt{n}}\left[trk_1k_1^{/}E_{\theta_0}\frac{\partial^2 c_n(b(\theta_0))_j}{\partial\beta\partial\beta'}\right]_{j=1,\ldots,l} \\ +\frac{1}{\sqrt{n}}\left[trk_1\sqrt{n}\left(\theta-\theta_0\right)'E_{\theta_0}s_n\left(\theta_0\right)\frac{\partial c_n(b(\theta_0))_j}{\partial\beta'}\right]_{j=1,\ldots,l} \\ +\frac{1}{2\sqrt{n}}\left[tr\sqrt{n}\left(\theta_n-\theta_0\right)\sqrt{n}\left(\theta_n-\theta_0\right)'\times\left(\begin{array}{c} E_{\theta_0}c_n\left(b\left(\theta_0\right)\right)_j s_n\left(\theta_0\right)s_n'\left(\theta_0\right) \\ +E_{\theta_0}c_n\left(b\left(\theta_0\right)\right)_j H_n\left(\theta_0\right)\end{array}\right)\right]_{j=1,\ldots,l} \end{array}\right] =$$

$0_{p\times 1} =>$

$$-\frac{\partial b^{/}(\theta_0)}{\partial\theta}E_{\theta_0}\frac{\partial(c_n(b(\theta_0)))^{/}}{\partial\beta}\frac{1}{n}\sum W^{**}\left(x_i,\theta_0\right)E_{\theta_0}\frac{\partial c_n(b(\theta_0))}{\partial\beta'}k_1$$
$$-\frac{1}{\sqrt{n}}\frac{\partial b^{/}(\theta_0)}{\partial\theta}E_{\theta_0}\frac{\partial(c_n(b(\theta_0)))^{/}}{\partial\beta}\frac{1}{n}\sum W^{**}\left(x_i,\theta_0\right)E_{\theta_0}\frac{\partial c_n(b(\theta_0))}{\partial\beta'}k_2$$
$$-\frac{1}{\sqrt{n}}\frac{\partial b^{/}(\theta_0)}{\partial\theta}E_{\theta_0}\frac{\partial(c_n^{*}(b(\theta_0)))^{/}}{\partial\beta}\left[\frac{\partial}{\partial\theta^{/}}\frac{1}{n}\sum W_{rj}^{**}\left(x_i,\theta_0\right)\sqrt{n}\left(\theta_n^{*}-\theta_0\right)\right]_{r,j=1,\ldots,l}E_{\theta_0}\frac{\partial c_n(b(\theta_0))}{\partial\beta'}k_1$$
$$-\frac{1}{\sqrt{n}}\left(\left[k_1'E_{\theta_0}\frac{\partial^2 c_n(b(\theta_0))_j}{\partial\beta\partial\beta'}\frac{\partial b(\theta_0)}{\partial\theta'}\right]_{j=1,\ldots,l}\right)^{/}\frac{1}{n}\sum W^{**}\left(x_i,\theta_0\right)E_{\theta_0}\frac{\partial c_n(b(\theta_0))}{\partial\beta'}k_1$$
$$+\frac{1}{\sqrt{n}}\left(\left[\sqrt{n}\left(\theta_n-\theta_0\right)'\left(\frac{\partial b^{/}(\theta_0)}{\partial\theta}E_{\theta_0}\frac{\partial^2 c_n(b(\theta_0))_j}{\partial\beta\partial\beta'}\frac{\partial b(\theta_0)}{\partial\theta'}-\left[E_{\theta_0}\frac{\partial c_n(b(\theta_0))_j}{\partial\beta'}\frac{\partial b(\theta_0)}{\partial\theta_r\partial\theta'}\right]_{r=1,\ldots p}\right)\right]_{j=1,\ldots,w_i}\right)^{/}\times$$
$$\frac{1}{n}\sum W^{**}\left(x_i,\theta_0\right)E_{\theta_0}\frac{\partial c_n(b(\theta_0))}{\partial\beta'}k_1$$
$$+\frac{\partial b^{/}(\theta_0)}{\partial\theta}E_{\theta_0}\frac{\partial(c_n(b(\theta_0)))^{/}}{\partial\beta}\frac{1}{n}\sum W^{**}\left(x_i,\theta_0\right)E_{\theta_0}\frac{\partial c_n(b(\theta_0))}{\partial\beta'}\frac{\partial b(\theta_0)}{\partial\theta'}\sqrt{n}\left(\theta_n-\theta_0\right)$$
$$+\frac{1}{\sqrt{n}}\frac{\partial b^{/}(\theta_0)}{\partial\theta}E_{\theta_0}\frac{\partial(c_n(b(\theta_0)))^{/}}{\partial\beta}\left[\frac{\partial}{\partial\theta^{/}}\frac{1}{n}\sum W_{rj}^{**}\left(x_i,\theta_0\right)\sqrt{n}\left(\theta_n^{*}-\theta_0\right)\right]_{r,j=1,\ldots,l}\times$$
$$E_{\theta}\frac{\partial c_n(b(\theta_0))}{\partial\beta'}\frac{\partial b(\theta_0)}{\partial\theta'}\sqrt{n}\left(\theta_n-\theta_0\right)$$
$$+\frac{1}{\sqrt{n}}\left(\left[k_1'E_{\theta_0}\frac{\partial^2 c_n(b(\theta_0))_j}{\partial\beta\partial\beta'}\frac{\partial b(\theta_0)}{\partial\theta'}\right]_{j=1,\ldots,l}\right)^{/}\frac{1}{n}\sum W^{**}\left(x_i,\theta_0\right)E_{\theta}\frac{\partial c_n(b(\theta_0))}{\partial\beta'}\frac{\partial b(\theta_0)}{\partial\theta'}\sqrt{n}\left(\theta_n-\theta_0\right)$$
$$-\frac{1}{\sqrt{n}}\left(\left[\sqrt{n}\left(\theta_n-\theta_0\right)'\left(\begin{array}{c}\frac{\partial b^{/}(\theta_0)}{\partial\theta}E_{\theta_0}\frac{\partial^2 c_n(b(\theta_0))_j}{\partial\beta\partial\beta'}\frac{\partial b(\theta_0)}{\partial\theta'} \\ -\left[E_{\theta_0}\frac{\partial c_n(b(\theta_0))_j}{\partial\beta'}\frac{\partial b(\theta_0)}{\partial\theta_r\partial\theta'}\right]_{r=1,\ldots p}\end{array}\right)\right]_{j=1,\ldots,l}\right)^{/}\times$$
$$\frac{1}{n}\sum W^{**}\left(x_i,\theta_0\right)E_{\theta_0}\frac{\partial c_n(b(\theta_0))}{\partial\beta'}\frac{\partial b(\theta_0)}{\partial\theta'}\sqrt{n}\left(\theta_n-\theta_0\right)$$
$$-\frac{1}{2\sqrt{n}}\frac{\partial b^{/}(\theta_0)}{\partial\theta}E_{\theta_0}\frac{\partial(c_n(b(\theta_0)))^{/}}{\partial\beta}\frac{1}{n}\sum W^{**}\left(x_i,\theta_0\right)\left[trk_1k_1^{/}E_{\theta_0}\frac{\partial^2 c_n(b(\theta_0))_j}{\partial\beta\partial\beta'}\right]_{j=1,\ldots,l}$$
$$-\frac{1}{\sqrt{n}}\frac{\partial b^{/}(\theta_0)}{\partial\theta}E_{\theta_0}\frac{\partial(c_n(b(\theta_0)))^{/}}{\partial\beta}\frac{1}{n}\sum W^{**}\left(x_i,\theta_0\right)\left[trk_1\sqrt{n}\left(\theta-\theta_0\right)'E_{\theta_0}s_n\left(\theta_0\right)\frac{\partial c_n(b(\theta_0))_j}{\partial\beta'}\right]_{j=1,\ldots,l}$$
$$-\frac{1}{2\sqrt{n}}\frac{\partial b^{/}(\theta_0)}{\partial\theta}E_{\theta_0}\frac{\partial(c_n(b(\theta_0)))^{/}}{\partial\beta}\frac{1}{n}\sum W^{**}\left(x_i,\theta_0\right)\times$$
$$\left[tr\sqrt{n}\left(\theta_n-\theta_0\right)\sqrt{n}\left(\theta_n-\theta_0\right)'\times\left(\begin{array}{c}E_{\theta_0}c_n\left(b\left(\theta_0\right)\right)_j s_n\left(\theta_0\right)s_n'\left(\theta_0\right) \\ +E_{\theta_0}c_n\left(b\left(\theta_0\right)\right)_j H_n\left(\theta_0\right)\end{array}\right)\right]_{j=1,\ldots,l}=0_{p\times 1} =>$$

As now $\frac{\partial b^{/}(\theta_0)}{\partial\theta}E_{\theta_0}\frac{\partial^2}{\partial\beta\partial\beta'}c_n\left(b\left(\theta_0\right)\right)_j\frac{\partial b(\theta_0)}{\partial\theta'}-\left[E_{\theta_0}\frac{\partial c_n(b(\theta_0))_j}{\partial\beta'}\frac{\partial^2 b(\theta_0)}{\partial\theta_r\partial\theta'}\right]_{r=1,\ldots p}=$
$E_{\theta_0}c_n\left(b\left(\theta_0\right)\right)_j s_n\left(\theta_0\right)s_n'\left(\theta_0\right)+E_{\theta_0}c_n\left(b\left(\theta_0\right)\right)_j H_n\left(\theta_0\right)$, $E_{\theta_0}s_n\left(\theta_0\right)\frac{\partial c_n(b(\theta_0))_j}{\partial\beta'}=$

$-\frac{\partial b'(\theta_0)}{\partial \theta} E_{\theta_0} \frac{\partial^2}{\partial\beta\partial\beta'} c_n\left(b\left(\theta_0\right)\right)_j$ and

$\frac{1}{n}\sum_i W^{**}\left(x_i,\theta_0\right) = W^{**}\left(b\left(\theta_0\right)\right) + \frac{1}{\sqrt{n}} w^{**}\left(z,\theta_0\right)$, where $z \sim N\left(0,\Sigma\right)$, we have

$$-\frac{\partial b'(\theta_0)}{\partial \theta} E_{\theta_0} \frac{\partial (c_n(b(\theta_0)))'}{\partial \beta} W^{**}\left(b\left(\theta_0\right)\right) E_{\theta_0} \frac{\partial c_n(b(\theta_0))}{\partial \beta'} k_1$$
$$-\frac{1}{\sqrt{n}} \frac{\partial b'(\theta_0)}{\partial \theta} E_{\theta_0} \frac{\partial (c_n(b(\theta_0)))'}{\partial \beta} w^{**}\left(z,\theta_0\right) E_{\theta_0} \frac{\partial c_n(b(\theta_0))}{\partial \beta'} k_1$$
$$-\frac{1}{\sqrt{n}} \frac{\partial b'(\theta_0)}{\partial \theta} E_{\theta_0} \frac{\partial (c_n(b(\theta_0)))'}{\partial \beta} W^{**}\left(b\left(\theta_0\right)\right) E_{\theta_0} \frac{\partial c_n(b(\theta_0))}{\partial \beta'} k_2$$
$$-\frac{1}{\sqrt{n}} \frac{\partial b'(\theta_0)}{\partial \theta} E_{\theta_0} \frac{\partial (c_n(b(\theta_0)))'}{\partial \beta} \left[\frac{\partial}{\partial \theta'} W^{**}_{rj}\left(b\left(\theta_0\right)\right)\sqrt{n}\left(\theta^*_n-\theta_0\right)\right]_{r,j=1,\ldots,l} E_{\theta_0} \frac{\partial c_n(b(\theta_0))}{\partial \beta'} k_1$$
$$-\frac{1}{\sqrt{n}} \left(\left[k_1' E_{\theta_0} \frac{\partial^2 c_n(b(\theta_0))_j}{\partial\beta\partial\beta'} \frac{\partial b(\theta_0)}{\partial \theta'}\right]_{j=1,\ldots,l}\right)' W^{**}\left(b\left(\theta_0\right)\right) E_{\theta_0} \frac{\partial c_n(b(\theta_0))}{\partial \beta'} k_1$$
$$+\frac{1}{\sqrt{n}} \left(\left[\sqrt{n}\left(\theta_n-\theta_0\right)' \begin{pmatrix} \frac{\partial b'(\theta_0)}{\partial \theta} E_{\theta_0} \frac{\partial^2 c_n(b(\theta_0))_j}{\partial\beta\partial\beta'} \frac{\partial b(\theta_0)}{\partial \theta'} \\ -\left[E_{\theta_0} \frac{\partial c_n(b(\theta_0))_j}{\partial \beta'} \frac{\partial b(\theta_0)}{\partial \theta_r \partial \theta'}\right]_{r=1,\ldots p} \end{pmatrix}\right]_{j=1,\ldots,l}\right)' \times$$
$$W^{**}\left(b\left(\theta_0\right)\right) E_{\theta_0} \frac{\partial c_n(b(\theta_0))}{\partial \beta'} k_1$$
$$+\frac{\partial b'(\theta_0)}{\partial \theta} E_{\theta_0} \frac{\partial (c_n(b(\theta_0)))'}{\partial \beta} W^{**}\left(b\left(\theta_0\right)\right) E_{\theta_0} \frac{\partial c_n(b(\theta_0))}{\partial \beta'} \frac{\partial b(\theta_0)}{\partial \theta'} \sqrt{n}\left(\theta_n-\theta_0\right)$$
$$+\frac{1}{\sqrt{n}} \frac{\partial b'(\theta_0)}{\partial \theta} E_{\theta_0} \frac{\partial (c_n(b(\theta_0)))'}{\partial \beta} w^{**}\left(z,\theta_0\right) E_{\theta} \frac{\partial c_n(b(\theta))}{\partial \beta'} \frac{\partial b(\theta)}{\partial \theta'} \sqrt{n}\left(\theta_n-\theta_0\right)$$
$$+\frac{1}{\sqrt{n}} \frac{\partial b'(\theta_0)}{\partial \theta} E_{\theta_0} \frac{\partial (c_n(b(\theta_0)))'}{\partial \beta} \left[\frac{\partial}{\partial \theta'} W^{**}_{rj}\left(b\left(\theta_0\right)\right)\sqrt{n}\left(\theta^*_n-\theta_0\right)\right]_{r,j=1,\ldots,l} \times$$
$$E_{\theta} \frac{\partial c_n(b(\theta))}{\partial \beta'} \frac{\partial b(\theta)}{\partial \theta'} \sqrt{n}\left(\theta_n-\theta_0\right)$$
$$+\frac{1}{\sqrt{n}} \left(\left[k_1' E_{\theta_0} \frac{\partial^2 c_n(b(\theta_0))_j}{\partial\beta\partial\beta'} \frac{\partial b(\theta_0)}{\partial \theta'}\right]_{j=1,\ldots,l}\right)' W^{**}\left(b\left(\theta_0\right)\right) E_{\theta} \frac{\partial c_n(b(\theta_0))}{\partial \beta'} \frac{\partial b(\theta_0)}{\partial \theta'} \sqrt{n}\left(\theta_n-\theta_0\right)$$
$$-\frac{1}{\sqrt{n}} \left(\left[\sqrt{n}\left(\theta_n-\theta_0\right)' \begin{pmatrix} \frac{\partial b'(\theta_0)}{\partial \theta} E_{\theta_0} \frac{\partial^2 c_n(b(\theta_0))_j}{\partial\beta\partial\beta'} \frac{\partial b(\theta_0)}{\partial \theta'} \\ -\left[E_{\theta_0} \frac{\partial c_n(b(\theta_0))_j}{\partial \beta'} \frac{\partial b(\theta_0)}{\partial \theta_r \partial \theta'}\right]_{r=1,\ldots p} \end{pmatrix}\right]_{j=1,\ldots,l}\right)' \times$$
$$W^{**}\left(b\left(\theta_0\right)\right) E_{\theta} \frac{\partial c_n(b(\theta))}{\partial \beta'} \frac{\partial b(\theta)}{\partial \theta'} \sqrt{n}\left(\theta_n-\theta_0\right)$$
$$-\frac{1}{2\sqrt{n}} \frac{\partial b'(\theta_0)}{\partial \theta} E_{\theta_0} \frac{\partial (c_n(b(\theta_0)))'}{\partial \beta} W^{**}\left(b\left(\theta_0\right)\right) \left[tr k_1 k_1' E_{\theta_0} \frac{\partial^2 c_n(b(\theta_0))_j}{\partial\beta\partial\beta'}\right]_{j=1,\ldots,l}$$
$$+\frac{1}{\sqrt{n}} \frac{\partial b'(\theta_0)}{\partial \theta} E_{\theta_0} \frac{\partial (c_n(b(\theta_0)))'}{\partial \beta} W^{**}\left(b\left(\theta_0\right)\right) \left[tr k_1 \sqrt{n}\left(\theta-\theta_0\right)' \frac{\partial b'(\theta_0)}{\partial \theta} E_{\theta_0} \frac{\partial^2}{\partial\beta\partial\beta'} c_n\left(b\left(\theta_0\right)\right)_j\right]_{j=1,\ldots,l}$$
$$-\frac{1}{2\sqrt{n}} \frac{\partial b'(\theta_0)}{\partial \theta} E_{\theta_0} \frac{\partial (c_n(b(\theta_0)))'}{\partial \beta} W^{**}\left(b\left(\theta_0\right)\right) \begin{bmatrix} tr\sqrt{n}\left(\theta_n-\theta_0\right)\sqrt{n}\left(\theta_n-\theta_0\right)' \times \\ \begin{pmatrix} \frac{\partial b'(\theta_0)}{\partial \theta} E_{\theta_0} \frac{\partial^2}{\partial\beta\partial\beta'} c_n\left(b\left(\theta_0\right)\right)_j \frac{\partial b(\theta_0)}{\partial \theta'} \\ -\left[E_{\theta_0} \frac{\partial c_n(b(\theta_0))_j}{\partial \beta'} \frac{\partial^2 b(\theta_0)}{\partial \theta_r \partial \theta'}\right]_{r=1,\ldots p} \end{pmatrix} \end{bmatrix}_{j=1,\ldots,l} =$$
$$0_{p\times 1} \Rightarrow$$

Now for $J = \frac{\partial b'(\theta_0)}{\partial \theta} E_{\theta_0} \frac{\partial (c_n(b(\theta_0)))'}{\partial \beta} W^{**}\left(b\left(\theta_0\right)\right) E_{\theta_0} \frac{\partial c_n(b(\theta_0))}{\partial \beta'} \frac{\partial b(\theta_0)}{\partial \theta'}$ we get

$$\sqrt{n}\left(\theta_n-\theta_0\right) = J^{-1} \frac{\partial b'(\theta_0)}{\partial \theta} E_{\theta_0} \frac{\partial (c_n(b(\theta_0)))'}{\partial \beta} W^{**}\left(b\left(\theta_0\right)\right) E_{\theta_0} \frac{\partial c_n(b(\theta_0))}{\partial \beta'} k_1$$
$$+\frac{1}{\sqrt{n}} J^{-1} \frac{\partial b'(\theta_0)}{\partial \theta} E_{\theta_0} \frac{\partial (c_n(b(\theta_0)))'}{\partial \beta} w^{**}\left(z,\theta_0\right) E_{\theta_0} \frac{\partial c_n(b(\theta_0))}{\partial \beta'} k_1$$

$$+\frac{1}{\sqrt{n}}J^{-1}\frac{\partial b^{/}(\theta_0)}{\partial\theta}E_{\theta_0}\frac{\partial(c_n(b(\theta_0)))^{/}}{\partial\beta}W^{**}(b(\theta_0))E_{\theta_0}\frac{\partial c_n(b(\theta_0))}{\partial\beta'}k_2$$

$$+\frac{1}{\sqrt{n}}J^{-1}\frac{\partial b^{/}(\theta_0)}{\partial\theta}E_{\theta_0}\frac{\partial(c_n(b(\theta_0)))^{/}}{\partial\beta}\left[\frac{\partial}{\partial\theta^{/}}W^{**}_{rj}(b(\theta_0))\sqrt{n}(\theta^*_n-\theta_0)\right]_{r,j=1,\dots,l}E_{\theta_0}\frac{\partial c_n(b(\theta_0))}{\partial\beta'}k_1$$

$$+\frac{1}{\sqrt{n}}J^{-1}\left(\left[k_1'E_{\theta_0}\frac{\partial^2c_n(b(\theta_0))_j}{\partial\beta\partial\beta'}\frac{\partial b(\theta_0)}{\partial\theta'}\right]_{j=1,\dots,l}\right)^{/}W^{**}(b(\theta_0))E_{\theta_0}\frac{\partial c_n(b(\theta_0))}{\partial\beta'}k_1$$

$$-\frac{1}{\sqrt{n}}J^{-1}\left(\left[\sqrt{n}(\theta_n-\theta_0)'\begin{pmatrix}\frac{\partial b^{/}(\theta_0)}{\partial\theta}E_{\theta_0}\frac{\partial^2c_n(b(\theta_0))_j}{\partial\beta\partial\beta'}\frac{\partial b(\theta_0)}{\partial\theta'}\\-\left[E_{\theta_0}\frac{\partial c_n(b(\theta_0))_j}{\partial\beta'}\frac{\partial b(\theta_0)}{\partial\theta_r\partial\theta'}\right]_{r=1,\dots p}\end{pmatrix}\right]_{j=1,\dots,l}\right)^{/}$$

$$\times W^{**}(b(\theta_0))E_{\theta_0}\frac{\partial c_n(b(\theta_0))}{\partial\beta'}k_1$$

$$-\frac{1}{\sqrt{n}}J^{-1}\frac{\partial b^{/}(\theta_0)}{\partial\theta}E_{\theta_0}\frac{\partial(c_n(b(\theta_0)))^{/}}{\partial\beta}w^{**}(z,\theta_0)E_{\theta_0}\frac{\partial c_n(b(\theta_0))}{\partial\beta'}\frac{\partial b(\theta_0)}{\partial\theta'}J^{-1}$$

$$\times\frac{\partial b^{/}(\theta_0)}{\partial\theta}E_{\theta_0}\frac{\partial(c_n(b(\theta_0)))^{/}}{\partial\beta}W^{**}(b(\theta_0))E_{\theta_0}\frac{\partial c_n(b(\theta_0))}{\partial\beta'}k_1$$

$$-\frac{1}{\sqrt{n}}J^{-1}\frac{\partial b^{/}(\theta_0)}{\partial\theta}E_{\theta_0}\frac{\partial(c_n(b(\theta_0)))^{/}}{\partial\beta}\left[\frac{\partial}{\partial\theta^{/}}W^{**}_{rj}(b(\theta_0))\sqrt{n}(\theta^*_n-\theta_0)\right]_{r,j=1,\dots,l}$$

$$\times E_{\theta}\frac{\partial c_n(b(\theta))}{\partial\beta'}\frac{\partial b(\theta)}{\partial\theta'}\sqrt{n}(\theta_n-\theta_0)$$

$$-\frac{1}{\sqrt{n}}J^{-1}\left(\left[k_1'E_{\theta_0}\frac{\partial^2c_n(b(\theta_0))_j}{\partial\beta\partial\beta'}\frac{\partial b(\theta_0)}{\partial\theta'}\right]_{j=1,\dots,l}\right)^{/}W^{**}(b(\theta_0))\times$$

$$E_{\theta_0}\frac{\partial c_n(b(\theta_0))}{\partial\beta'}\frac{\partial b(\theta_0)}{\partial\theta'}J^{-1}\frac{\partial b^{/}(\theta_0)}{\partial\theta}E_{\theta_0}\frac{\partial(c_n(b(\theta_0)))^{/}}{\partial\beta}W^{**}(b(\theta_0))E_{\theta_0}\frac{\partial c_n(b(\theta_0))}{\partial\beta'}k_1$$

$$+\frac{1}{\sqrt{n}}J^{-1}\left(\left[\sqrt{n}(\theta_n-\theta_0)'\begin{pmatrix}\frac{\partial b^{/}(\theta_0)}{\partial\theta}E_{\theta_0}\frac{\partial^2c_n(b(\theta_0))_j}{\partial\beta\partial\beta'}\frac{\partial b(\theta_0)}{\partial\theta'}\\-\left[E_{\theta_0}\frac{\partial c_n(b(\theta_0))_j}{\partial\beta'}\frac{\partial b(\theta_0)}{\partial\theta_r\partial\theta'}\right]_{r=1,\dots p}\end{pmatrix}\right]_{j=1,\dots,l}\right)^{/}$$

$$\times W^{**}(b(\theta_0))E_{\theta_0}\frac{\partial c_n(b(\theta_0))}{\partial\beta'}\frac{\partial b(\theta_0)}{\partial\theta'}\sqrt{n}(\theta_n-\theta_0)$$

$$+\frac{1}{2\sqrt{n}}J^{-1}\frac{\partial b^{/}(\theta_0)}{\partial\theta}E_{\theta_0}\frac{\partial(c_n(b(\theta_0)))^{/}}{\partial\beta}W^{**}(b(\theta_0))\left[trk_1k_1^{/}E_{\theta_0}\frac{\partial^2c_n(b(\theta_0))_j}{\partial\beta\partial\beta'}\right]_{j=1,\dots,l}$$

$$-\frac{1}{\sqrt{n}}J^{-1}\frac{\partial b^{/}(\theta_0)}{\partial\theta}E_{\theta_0}\frac{\partial\left(c^*_{i,n}(b(\theta_0))\right)^{/}}{\partial\beta}W^{**}(b(\theta_0))\left[trk_1\sqrt{n}(\theta-\theta_0)'\frac{\partial b'(\theta_0)}{\partial\theta}E_{\theta_0}\frac{\partial^2}{\partial\beta\partial\beta'}c_n(b(\theta_0))_j\right]_{j=1,\dots,l}$$

$$+\frac{1}{2\sqrt{n}}J^{-1}\frac{\partial b^{/}(\theta_0)}{\partial\theta}E_{\theta_0}\frac{\partial(c_n(b(\theta_0)))^{/}}{\partial\beta}W^{**}(b(\theta_0))\left[\begin{array}{c}tr\sqrt{n}(\theta_n-\theta_0)\sqrt{n}(\theta_n-\theta_0)'\times\\\begin{pmatrix}\frac{\partial b^{/}(\theta_0)}{\partial\theta}E_{\theta_0}\frac{\partial^2}{\partial\beta\partial\beta'}c_n(b(\theta_0))_j\frac{\partial b(\theta_0)}{\partial\theta'}\\-\left[E_{\theta_0}\frac{\partial c_n(b(\theta_0))_j}{\partial\beta'}\frac{\partial^2b(\theta_0)}{\partial\theta_r\partial\theta'}\right]_{r=1,\dots p}\end{pmatrix}\end{array}\right]_{j=1,\dots,l}\Rightarrow$$

Hence $\sqrt{n}(\theta_n-\theta_0)=J^{-1}\frac{\partial b^{/}(\theta_0)}{\partial\theta}E_{\theta_0}\frac{\partial(c_n(b(\theta_0)))^{/}}{\partial\beta}W^{**}(b(\theta_0))E_{\theta_0}\frac{\partial c_n(b(\theta_0))}{\partial\beta'}k_1$

$$+\frac{1}{\sqrt{n}}J^{-1}\frac{\partial b^{/}(\theta_0)}{\partial\theta}E_{\theta_0}\frac{\partial(c_n(b(\theta_0)))^{/}}{\partial\beta}W^{**}(b(\theta_0))E_{\theta_0}\frac{\partial c_n(b(\theta_0))}{\partial\beta'}k_2$$

$$+\frac{1}{\sqrt{n}}J^{-1}\frac{\partial b^{/}(\theta_0)}{\partial\theta}E_{\theta_0}\frac{\partial(c_n(b(\theta_0)))^{/}}{\partial\beta}w^{**}(z,\theta_0)E_{\theta_0}\frac{\partial c_n(b(\theta_0))}{\partial\beta'}$$

$$\times\left(Id_{q\times q}-\frac{\partial b(\theta_0)}{\partial\theta'}J^{-1}\frac{\partial b^{/}(\theta_0)}{\partial\theta}E_{\theta_0}\frac{\partial(c_n(b(\theta_0)))^{/}}{\partial\beta}W^{**}(b(\theta_0))E_{\theta_0}\frac{\partial c_n(b(\theta_0))}{\partial\beta'}\right)k_1$$

$$+\frac{1}{\sqrt{n}}J^{-1}\frac{\partial b^{/}(\theta_0)}{\partial\theta}E_{\theta_0}\frac{\partial(c_n(b(\theta_0)))^{/}}{\partial\beta}\left[\frac{\partial}{\partial\theta^{/}}W^{**}_{rj}(b(\theta_0))\sqrt{n}(\theta^*_n-\theta_0)\right]_{r,j=1,\dots,l}\times$$

$$E_{\theta_0}\frac{\partial c_n(b(\theta_0))}{\partial\beta'}\left(k_1-\frac{\partial b(\theta_0)}{\partial\theta'}\sqrt{n}(\theta_n-\theta_0)\right)$$

$$
\begin{aligned}
&+\tfrac{1}{\sqrt{n}}J^{-1}\left(\left[k_1^{\prime}E_{\theta_0}\tfrac{\partial^2 c_n(b(\theta_0))_j}{\partial\beta\partial\beta^{\prime}}\tfrac{\partial b(\theta_0)}{\partial\theta^{\prime}}\right]_{j=1,\ldots,l}\right)^{\prime}W^{**}\left(b\left(\theta_0\right)\right)E_{\theta_0}\tfrac{\partial c_n(b(\theta_0))}{\partial\beta^{\prime}}\\
&\times\left(Id_{q\times q}-\tfrac{\partial b(\theta_0)}{\partial\theta^{\prime}}J^{-1}\tfrac{\partial b^{\prime}(\theta_0)}{\partial\theta}E_{\theta_0}\tfrac{\partial(c_n(b(\theta_0)))^{\prime}}{\partial\beta}W^{**}\left(b\left(\theta_0\right)\right)E_{\theta_0}\tfrac{\partial c_n(b(\theta_0))}{\partial\beta^{\prime}}\right)k_1\\
&-\tfrac{1}{\sqrt{n}}J^{-1}\left(\left[\sqrt{n}\left(\theta_n-\theta_0\right)^{\prime}\begin{pmatrix}\tfrac{\partial b^{\prime}(\theta_0)}{\partial\theta}E_{\theta_0}\tfrac{\partial^2 c_n(b(\theta_0))_j}{\partial\beta\partial\beta^{\prime}}\tfrac{\partial b(\theta_0)}{\partial\theta^{\prime}}\\-\left[E_{\theta_0}\tfrac{\partial c_n(b(\theta_0))_j}{\partial\beta^{\prime}}\tfrac{\partial b(\theta_0)}{\partial\theta_r\partial\theta^{\prime}}\right]_{r=1,\ldots p}\end{pmatrix}\right]_{j=1,\ldots,l}\right)^{\prime}\\
&\times W^{**}\left(b\left(\theta_0\right)\right)E_{\theta_0}\tfrac{\partial c_n(b(\theta_0))}{\partial\beta^{\prime}}\left(k_1-\tfrac{\partial b(\theta_0)}{\partial\theta^{\prime}}\sqrt{n}\left(\theta_n-\theta_0\right)\right)\\
&+\tfrac{1}{2\sqrt{n}}J^{-1}\tfrac{\partial b^{\prime}(\theta_0)}{\partial\theta}E_{\theta_0}\tfrac{\partial(c_n(b(\theta_0)))^{\prime}}{\partial\beta}W^{**}\left(b\left(\theta_0\right)\right)\\
&\times\begin{pmatrix}\left[trk_1k_1^{\prime}E_{\theta_0}\tfrac{\partial^2 c_n(b(\theta_0))_j}{\partial\beta\partial\beta^{\prime}}\right]_{j=1,\ldots,l}\\-\left[tr\sqrt{n}\left(\theta_n-\theta_0\right)\sqrt{n}\left(\theta_n-\theta_0\right)^{\prime}\left[E_{\theta_0}\tfrac{\partial c_n(b(\theta_0))_j}{\partial\beta^{\prime}}\tfrac{\partial^2 b(\theta_0)}{\partial\theta_r\partial\theta^{\prime}}\right]_{r=1,\ldots p}\right]_{j=1,\ldots,l}\end{pmatrix}\\
&+\tfrac{1}{2\sqrt{n}}J^{-1}\tfrac{\partial b^{\prime}(\theta_0)}{\partial\theta}E_{\theta_0}\tfrac{\partial(c_n(b(\theta_0)))^{\prime}}{\partial\beta}W^{**}\left(b\left(\theta_0\right)\right)\\
&\times\begin{pmatrix}\left[tr\sqrt{n}\left(\theta_n-\theta_0\right)\sqrt{n}\left(\theta_n-\theta_0\right)^{\prime}\tfrac{\partial b^{\prime}(\theta_0)}{\partial\theta}E_{\theta_0}\tfrac{\partial^2}{\partial\beta\partial\beta^{\prime}}c_n\left(b\left(\theta_0\right)\right)_j\tfrac{\partial b(\theta_0)}{\partial\theta^{\prime}}\right]_{j=1,\ldots,l}\\-2\left[trk_1\sqrt{n}\left(\theta-\theta_0\right)^{\prime}\tfrac{\partial b^{\prime}(\theta_0)}{\partial\theta}E_{\theta_0}\tfrac{\partial^2 c_n(b(\theta_0))_j}{\partial\beta\partial\beta^{\prime}}\right]_{j=1,\ldots,l}\end{pmatrix}.\ \blacksquare
\end{aligned}
$$

Remark R.40 *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 GMR 1 counterpart.*

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

Corollary 4 *GT estimator $\underset{0}{\sim}$ (GMR 1 estimator $\underset{0}{\sim}$ GMR 2 estimator) iff the weighting matrix for GMR 1 and the GMR 2 estimators is chosen as*

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

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

Proof. Trivial. ■

Lemma 3.9 *For $A, M\in\mathcal{M}^{q\times q}$, and M invertible*

$$\left[tr\left(A\frac{\partial^2 b_i\left(\theta_0\right)}{\partial\theta\partial\theta^{\prime}}\right)\right]_{i=1,\ldots,q}=M^{-1}\left[\sum_{j=1}^{q}M_{i,j}tr\left(A\frac{\partial^2 b_j\left(\theta_0\right)}{\partial\theta\partial\theta^{\prime}}\right)\right]_{i=1,\ldots,q}$$

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 $v_i = \sum_{j=1}^{q}\sum_{m=1}^{q} M^{i,m}M_{m,j}tr\left(A\frac{\partial^2 b_j(\theta_0)}{\partial\theta\partial\theta'}\right) = \sum_{j=1}^{q}\delta_{i,j}tr\left(A\frac{\partial^2 b_j(\theta_0)}{\partial\theta\partial\theta'}\right) = tr\left(A\frac{\partial^2 b_i(\theta_0)}{\partial\theta\partial\theta'}\right) = u_i, \forall i = 1,\ldots q.$ ■

In the special case of equality between the involved dimensions we obtain the following corollary which is proven with the help of the following lemma.

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

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

and

$$q_2 = \left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1} k_2 - \frac{1}{2}\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\left[trq_1'q_1\frac{\partial^2 b_j(\theta_0)}{\partial\theta\partial\theta'}\right]_{j=1,\ldots,l}$$

Proof. By direct substitution we obtain that $q_1 = \left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1} k_1$, and $q_2 = \left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1} k_2 - \frac{1}{2\sqrt{n}}\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\left(E_{\theta_0}\frac{\partial c_n(b(\theta_0))}{\partial\beta'}\right)^{-1}\left[trq_1'q_1\left[E_{\theta_0}\frac{\partial c_n(b(\theta_0))_j}{\partial\beta'}\frac{\partial^2 b(\theta_0)}{\partial\theta_r\partial\theta'}\right]_{r=1,\ldots p}\right]_{j=1,\ldots,l} = \left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1} k_2 - \frac{1}{2\sqrt{n}}\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\left(E_{\theta_0}\frac{\partial c_n(b(\theta_0))}{\partial\beta'}\right)^{-1}\left[\sum_r E_{\theta_0}\frac{\partial c_n(b(\theta_0))_j}{\partial\beta'}trq_1'q_1\left[\frac{\partial^2 b_r(\theta_0)}{\partial\theta\partial\theta'}\right]_{r=1,\ldots p}\right]_{j=1,\ldots,l}$.
and by the above Lemma we have $q_2 = \left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1} k_2 - \frac{1}{2\sqrt{n}}\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\left[trq_1'q_1\frac{\partial^2 b_j(\theta_0)}{\partial\theta\partial\theta'}\right]_{j=1,\ldots,l}$.
■

Remark R.41 *This corollary is in accordance with lemma 1.2. It shows **neither the GT estimator is second order unbiased** under the framework imposed by corollary R.41 or any relevant framework.*

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.8 enables the possibility that $c(x,\beta)$ lies on a particular bundle (Hilbert bundle, see among others [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 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 [17] (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 $b^{**}|_{\mathbf{par}^{-1}(\mathcal{O}_{P_\Omega})} = \mathbf{par}_{**}^* \circ f|_{\mathcal{O}_{P_\Omega}} \circ \mathbf{par}^{-1} = \left(\theta_1, \theta_2, \ldots, \theta_p, \underbrace{0, \ldots, 0}_{q-p}\right)$. This representation is called canonical immersion around P_Ω. Hence due to the aforementioned theorem and the assumed properties of the binding function the following is true.

Lemma 3.10 *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 $b^{**}(\theta) = \begin{pmatrix} \theta \\ \mathbf{0}_{q-p} \end{pmatrix}$ $\forall \theta \in \mathbf{par}^{-1}(\mathcal{O}_{P_\Omega})$.*

Proof. See the proof of theorem 10.2 of [17] and note 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. ■

Remark R.42 *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 hereafter we denote it by $b(\theta)$ hence from this point and until the end of the present paragraph.*[12] *It is easily seen that when $b(\theta)$ is*

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

on the relevant form the aforementioned expansions simplify in some extend. We explore some interesting cases. In every one of these we assume that $W^*(x,\theta_0) = W^* = \begin{pmatrix} W_{1,p\times p} & W_{3,p\times q-p} \\ W'_{3,p\times q-p} & W_{2,q-p\times q-p} \end{pmatrix}$ *where* W_1, W_2, W_3 *are non stochastic matrices, independent of* θ*, of the relevant dimensions. Consider first the expansion of the GMR 1 estimator.*

Corollary 6 *Consider lemma 3.4, suppose that* b *is in local canonical form and* $W^*(x,\theta_0) = W^* = \begin{pmatrix} W_{1,p\times p} & W_{3,p\times q-p} \\ W'_3 & W_{2,q-p\times q-p} \end{pmatrix}$ *then*

$$q_1 = \begin{pmatrix} Id_{p\times p} & W^{-1}_{1,p\times p} W_{3,p\times q-p} \end{pmatrix} k_1$$

and

$$q_2 = \begin{pmatrix} Id_{p\times p} & W^{-1}_{1,p\times p} W_{3,p\times q-p} \end{pmatrix} k_2$$

Proof. Follows from direct substitutions on the results of lemma 3.4 by noting first that $\frac{\partial b(\theta_0)}{\partial \theta'} = \begin{pmatrix} Id_{p\times p} \\ \mathbf{0}_{q-p\times p} \end{pmatrix}$, $\frac{\partial b^2(\theta_0)_j}{\partial\theta\partial\theta'} = \mathbf{0}_{p\times p}$, $\forall j = 1,\ldots,q$, $w^* = \mathbf{0}_p$. ■

Remark R.43 *It is evident that* $\min_{W_{3,p\times q-p}} \|E_{\theta_0} q_2\| = \left\| \begin{pmatrix} (E_{\theta_0} k_2)_1 \\ \vdots \\ (E_{\theta_0} k_2)_p \end{pmatrix} \right\|$ *for* $W_{3,p\times q-p} = \mathbf{0}_{p\times q-p}$ *where* u_i *denotes the* i^{th} *element of the particular vector.*

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 GMR 2 estimator.

Corollary 7 *Consider lemma 3.6, suppose that* b *is in local canonical form and* $W^*(x,\theta_0) = W^* = \begin{pmatrix} W_{1,p\times p} & W_{3,p\times q-p} \\ W'_3 & W_{2,q-p\times q-p} \end{pmatrix}$ *then*

$$q_{1,n} = \begin{pmatrix} Id_{p\times p} & W^{-1}_{1,p\times p} W_{3,p\times q-p} \end{pmatrix} k_{1,n}$$

and

$$q_{2,n} = \begin{pmatrix} Id_{p\times p} & W^{-1}_{1,p\times p} W_{3,p\times q-p} \end{pmatrix} (k_{2,n} - E_{\theta_0} k_{2,n})$$

Proof. Follows from direct substitutions on the results of lemma 3.6 by noting that $\frac{\partial b(\theta_0)}{\partial \theta'} = \begin{pmatrix} Id_{p\times p} \\ \mathbf{0}_{q-p\times p} \end{pmatrix}$, $\frac{\partial b^2(\theta_0)_j}{\partial\theta\partial\theta'} = \mathbf{0}_{p\times p}$, $\forall j = 1,\ldots,q$. ■

Remark R.44 *The GMR 2 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 and ρ, 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.*

Example

We continue with an example. In this lemma 1.2 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 $p = q = l = 1$, $c(x_i, \beta) = \frac{\partial \rho(x_i, \beta)}{\partial \beta} = x_i x_{i-1} - \beta x_{i-1}^2$, and $b : (-1, 1) \to \left(-\frac{1}{2}, \frac{1}{2}\right)$ is globally invertible. We obtain from [5]

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

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

In the case of the GMR 1 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

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

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

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

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

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

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 GMR 2 estimator derived by this is second order unbiased by lemma 7. The particular reparametrization and the employment of GMR2 on it, coincides (see remark R.45) with the defined below 1−GMR2. The analogous expansion of the auxiliary estimator (or equivalently of $\theta_n^{(0}$ in the language of the next section) coincides with the one of the GMR1 presented above. For the bias corrector GMR2 (or equivalently $\theta_n^{(1)}$) we have that

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

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

establishing the second order unbiaseness. Notice that due to the global (hence local) nature of the moment approximation of [5], 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 [5].

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 $\sqrt{n}\left(\theta_n - \theta_0\right) \underset{\frac{(s-1)}{2}}{\sim} E\left(g\left(z, \frac{1}{\sqrt{n}}, \theta_0\right)\right)$. Analogously it will be termed s^{th}-order unbiased locally around θ_0, if the relevant expansion is valid, and $\sqrt{n}\left(\theta_n - \theta\right) \underset{\frac{(s-1)}{2}}{\sim} E\left(g\left(z, \frac{1}{\sqrt{n}}, \theta\right)\right)$ 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 $\sqrt{n}\left(\theta_n - \theta\right) \underset{\frac{(s-1)}{2}}{\sim} E\left(g\left(z, \frac{1}{\sqrt{n}}, \theta\right)\right)$ everywhere. Notice that up to the previous section we were essentially concerned with the first notion.

Now, the set up enabling lemma 3.10, 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 parametrizations, 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 $\forall\theta \in B(\theta_0, \varepsilon)$ we have that $b_n(\theta) = E_\theta \beta_n^* = \begin{pmatrix} \theta \\ 0_{q-p} \end{pmatrix}$, and 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.

Instead in an indirect answer to aforementioned question of result generalization, 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* $k > 1$ *and* $i \leq r$, $\theta_n^{(k} = \arg\min_\theta \left\| \theta_n^{(k-1} - E_\theta \theta_n^{(k-1} \right\|$.

Remark R.45 *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.7 and A.10.

Assumption A.17 $E(k_i(\theta, z))$ *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.46 *The assumption above is satisfied if* $E_\theta \beta_n = b(\theta) + \sum_{i=1}^{\infty} \frac{1}{n^{i/2}} E(k_i(\theta, z))$, $\forall \theta \in B(\theta_0, \varepsilon_5)$*, 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 3.6 or lemma 3.8 hold* ***locally*** *around* θ_0 *, then the* $r-$*GMR2 estimator, is of order* $2r + 1$ *unbiased at* θ_0*.*

Proof. Let the i^{th} element of vector x be denoted by x_i. Then we have that in the assumed neighborhood of θ_0

$$\sqrt{n}\left(\theta_n^{(0}-\theta\right)_i=(k_1)_i+\frac{1}{n^{\frac{1}{2}}}(k_2)_i+\frac{1}{n}(k_3)_i+\frac{1}{n^{\frac{3}{2}}}(k_4)_i+o\left(n^{-\frac{3}{2}}\right),\quad for\ i=1,..p$$

with $E_\theta(k_1)_i=0$. Now the GMR2 is defined as $\theta_n^{(1}=\arg\min_\theta\left(\theta_n^{(0}-E_\theta\theta_n^{(0)}\right)^2$. Hence we have that $\theta_n^{(0}-E_{\theta_n^{(1}}\theta_n^{(0}=0$. Expanding, the i^{th} element of $E_{\theta_n^{(1}}\theta_n^{(0}$, $\left(E_{\theta_n^{(1}}\theta_n^{(0)}\right)_i$ say, around θ_0 we get: $\left(E_{\theta_n^{(1}}\theta_n^{(0)}\right)_i=\left(E_{\theta_0}\theta_n^{(0)}\right)_i+\sum_{j=1}^p\frac{\partial\left(E_{\theta_0}\theta_n^{(0)}\right)_i}{\partial\theta_j}\left(\theta_n^{(1}-\theta_0\right)_j$
$+\frac{1}{2}\sum_{j=1}^p\sum_{m=1}^p\frac{\partial^2\partial\left(E_{\theta_0}\theta_n^{(0)}\right)_i}{\partial\theta_m\partial\theta_j}\left(\theta_n^{(1}-\theta_0\right)_m\left(\theta_n^{(1}-\theta_0\right)_j$
$+\frac{1}{3!}\sum_{j=1}^p\sum_{m=1}^p\sum_{l=1}^p\frac{\partial^3\partial\left(E_{\theta_0}\theta_n^{(0)}\right)_i}{\partial\theta_l\partial\theta_m\partial\theta_j}\left(\theta_n^{(1}-\theta_0\right)_l\left(\theta_n^{(1}-\theta_0\right)_m\left(\theta_n^{(1}-\theta_0\right)_j+....$ Hence as $\theta_n^{(0}-E_{\theta_n^{(1}}\theta_n^{(0}=0\Rightarrow$

$$\sqrt{n}\left(\theta_n^{(0}-E_{\theta_0}\theta_n^{(0)}\right)_i=\sum_{j=1}^p\frac{\partial\left(E_{\theta_0}\theta_n^{(0)}\right)_i}{\partial\theta_j}\sqrt{n}\left(\theta_n^{(1}-\theta_0\right)_j$$
$$+\frac{1}{2\sqrt{n}}\sum_{j=1}^p\sum_{m=1}^p\frac{\partial^2\partial\left(E_{\theta_0}\theta_n^{(0)}\right)_i}{\partial\theta_m\partial\theta_j}\sqrt{n}\left(\theta_n^{(1}-\theta_0\right)_m\sqrt{n}\left(\theta_n^{(1}-\theta_0\right)_j$$
$$+\frac{1}{3!n}\sum_{j=1}^p\sum_{m=1}^p\sum_{l=1}^p\frac{\partial^3\partial\left(E_{\theta_0}\theta_n^{(0)}\right)_i}{\partial\theta_l\partial\theta_m\partial\theta_j}\sqrt{n}\left(\theta_n^{(1}-\theta_0\right)_l\sqrt{n}\left(\theta_n^{(1}-\theta_0\right)_m\sqrt{n}\left(\theta_n^{(1}-\theta_0\right)_j+$$
$$o\left(n^{-1}\right)$$

Now for any θ in the assumed open ball at θ_0, and any $i,j,m,l,r=1,...p$, we have that due to assumption A.17

$$\begin{aligned}
\left(E_\theta\theta_n^{(0)}\right)_i &= \theta_i+\frac{1}{n}E_\theta k_{2,i}+\frac{1}{n^{\frac{3}{2}}}E_\theta k_{3,i}+\frac{1}{n^2}E_\theta k_{4,i}+o\left(n^{-2}\right),\\
\frac{\partial\left(E_\theta\theta_n^{(0)}\right)_i}{\partial\theta_j} &= \delta_{ij}+\frac{1}{n}\frac{\partial\left(E_\theta k_2\right)_i}{\partial\theta_j}+\frac{1}{n^{\frac{3}{2}}}\frac{\partial\left(E_\theta k_3\right)_i}{\partial\theta_j}+\frac{1}{n^2}\frac{\partial\left(E_\theta k_4\right)_i}{\partial\theta_j}+o\left(n^{-2}\right)\\
\frac{\partial^2\left(E_\theta\theta_n^{(0)}\right)_i}{\partial\theta_m\partial\theta_j} &= \frac{1}{n}\frac{\partial^2\left(E_\theta k_2\right)_i}{\partial\theta_m\partial\theta_j}+\frac{1}{n^{\frac{3}{2}}}\frac{\partial^2\left(E_\theta k_3\right)_i}{\partial\theta_m\partial\theta_j}+\frac{1}{n^2}\frac{\partial^2\left(E_\theta k_4\right)_i}{\partial\theta_m\partial\theta_j}+o\left(n^{-2}\right)\\
\frac{\partial^3\left(E_\theta\theta_n^{(0)}\right)_i}{\partial\theta_l\partial\theta_m\partial\theta_j} &= \frac{1}{n}\frac{\partial^3\left(E_\theta k_2\right)_i}{\partial\theta_l\partial\theta_m\partial\theta_j}+\frac{1}{n^{\frac{3}{2}}}\frac{\partial^3\left(E_\theta k_3\right)_i}{\partial\theta_l\partial\theta_m\partial\theta_j}+\frac{1}{n^2}\frac{\partial^3\left(E_\theta k_4\right)_i}{\partial\theta_l\partial\theta_m\partial\theta_j}+o\left(n^{-2}\right)
\end{aligned}$$

Hence

$$\sqrt{n}\left(\theta_n^{(0}-E_{\theta_0}\theta_n^{(0)}\right)_i=(k_1)_i+\frac{\left(k_2-E_{\theta_0}k_2\right)_i}{n^{\frac{1}{2}}}+\frac{\left(k_3-E_{\theta_0}k_3\right)_i}{n}+\frac{\left(k_4-E_{\theta_0}k_4\right)_i}{n^{\frac{3}{2}}}+o\left(n^{-\frac{3}{2}}\right)$$

Now, for $i=j$, let $\frac{\partial\left(E_\theta\theta_n^{(0}\right)_i}{\partial\theta_i}=1+\frac{1}{n}A+\frac{1}{n^{\frac{3}{2}}}B+\frac{1}{n^2}C$ then $\left(\frac{\partial\left(E_\theta\theta_n^{(0}\right)_i}{\partial\theta_j}\right)^{-1}=\frac{1}{1+\frac{1}{n}A+\frac{1}{n^{\frac{3}{2}}}B+\frac{1}{n^2}C}=\frac{1}{1+Ax^2+Bx^3+Cx^4}=f(x)$, with $f(0)=1$, $f^{/}(0)=0$, $f^{//}(0)=-2A$, and $f^{///}(0)=-6B$. Expanding $f(x)$ around $x=0$ we get $f(x)=1-Ax^2-Bx^3+o\left(n^{-\frac{3}{2}}\right)$ and consequently $\left(\frac{\partial\left(E_\theta\theta_n^{(0}\right)_i}{\partial\theta_i}\right)^{-1}=1-\frac{1}{n}\frac{\partial(E_\theta k_2)_i}{\partial\theta_i}-\frac{1}{n^{\frac{3}{2}}}\frac{\partial(E_\theta k_3)_i}{\partial\theta_i}+o\left(n^{-\frac{3}{2}}\right)$. Hence

$$
\begin{aligned}
&(k_1)_i+\frac{1}{n^{\frac{1}{2}}}(k_2-E_{\theta_0}k_2)_i+\frac{1}{n}(k_3-E_{\theta_0}k_3)_i-\frac{1}{n}\frac{\partial(E_\theta k_2)_i}{\partial\theta_i}(k_1)_i\\
&+\frac{1}{n^{\frac{3}{2}}}(k_4-E_{\theta_0}k_4)_i-\frac{1}{n^{\frac{3}{2}}}\frac{\partial(E_\theta k_2)_i}{\partial\theta_i}(k_2-E_{\theta_0}k_2)_i-\frac{1}{n^{\frac{3}{2}}}\frac{\partial(E_\theta k_3)_i}{\partial\theta_i}(k_1)_i\\
=\;&\sqrt{n}\left(\theta_n^{(1}-\theta_0\right)_i+\sum_{j\neq i=1}^{p}\frac{\partial\left(E_{\theta_0}\theta_n^{(0}\right)_i}{\partial\theta_j}\sqrt{n}\left(\theta_n^{(1}-\theta_0\right)_j\\
&-\frac{1}{n}\frac{\partial(E_\theta k_2)_i}{\partial\theta_i}\sum_{j\neq i=1}^{p}\frac{\partial\left(E_{\theta_0}\theta_n^{(0}\right)_i}{\partial\theta_j}\sqrt{n}\left(\theta_n^{(1}-\theta_0\right)_j\\
&+\frac{1}{2\sqrt{n}}\sum_{j=1}^{p}\sum_{m=1}^{p}\frac{\partial^2\partial\left(E_{\theta_0}\theta_n^{(0}\right)_i}{\partial\theta_m\partial\theta_j}\sqrt{n}\left(\theta_n^{(1}-\theta_0\right)_m\sqrt{n}\left(\theta_n^{(1}-\theta_0\right)_j\\
&+\frac{1}{3!n}\sum_{j=1}^{p}\sum_{m=1}^{p}\sum_{l=1}^{p}\frac{\partial^3\partial\left(E_{\theta_0}\theta_n^{(0}\right)_i}{\partial\theta_l\partial\theta_m\partial\theta_j}\sqrt{n}\left(\theta_n^{(1}-\theta_0\right)_l\sqrt{n}\left(\theta_n^{(1}-\theta_0\right)_m\sqrt{n}\left(\theta_n^{(1}-\theta_0\right)_j+o\left(n^{-\frac{3}{2}}\right)
\end{aligned}
$$

Now notice that first all the higher order derivatives are of order $O(n^{-1})$, i.e. $\frac{\partial^2\left(E_\theta\theta_n^{(0}\right)_i}{\partial\theta_m\partial\theta_j}=O(n^{-1})$, $\frac{\partial^3\left(E_\theta\theta_n^{(0}\right)_i}{\partial\theta_l\partial\theta_m\partial\theta_j}=O(n^{-1})$ and $\frac{\partial^4\left(E_\theta\theta_n^{(0}\right)_i}{\partial\theta_r\partial\theta_l\partial\theta_m\partial\theta_j}=O(n^{-1})$.

Further the same is true for $i \neq j$, i.e. $\frac{\partial\left(E_{\theta_0}\theta_n^{(0}\right)_i}{\partial\theta_j} = O\left(n^{-1}\right)$. Hence we get

$$\begin{aligned}
&(k_1)_i + \frac{1}{n^{\frac{1}{2}}}\left(k_2 - E_{\theta_0}k_2\right)_i + \frac{1}{n}\left(k_3 - E_{\theta_0}k_3\right)_i - \frac{1}{n}\frac{\partial\left(E_\theta k_2\right)_i}{\partial\theta_i}(k_1)_i \\
&+\frac{1}{n^{\frac{3}{2}}}\left(k_4 - E_{\theta_0}k_4\right)_i - \frac{1}{n^{\frac{3}{2}}}\frac{\partial\left(E_\theta k_2\right)_i}{\partial\theta_i}\left(k_2 - E_{\theta_0}k_2\right)_i - \frac{1}{n^{\frac{3}{2}}}\frac{\partial\left(E_\theta k_3\right)_i}{\partial\theta_i}(k_1)_i \\
= \;& \sqrt{n}\left(\theta_n^{(1} - \theta_0\right)_i + \sum_{j\neq i=1}^{p}\left(\frac{1}{n}\frac{\partial\left(E_{\theta_0}k_2\right)_i}{\partial\theta_j} + \frac{1}{n^{\frac{3}{2}}}\frac{\partial\left(E_{\theta_0}k_3\right)_i}{\partial\theta_j}\right)\sqrt{n}\left(\theta_n^{(1} - \theta_0\right)_j \\
&+\frac{1}{2n^{\frac{3}{2}}}\sum_{j=1}^{p}\sum_{m=1}^{p}\frac{\partial^2\left(E_{\theta_0}k_2\right)_i}{\partial\theta_m\partial\theta_j}\sqrt{n}\left(\theta_n^{(1} - \theta_0\right)_m\sqrt{n}\left(\theta_n^{(1} - \theta_0\right)_j + o\left(n^{-\frac{3}{2}}\right)
\end{aligned}$$

Inverting we get:

$$\begin{aligned}
\sqrt{n}\left(\theta_n^{(1} - \theta_0\right)_i \;=\;& (k_1)_i + \frac{1}{n^{\frac{1}{2}}}\left(k_2 - E_{\theta_0}k_2\right)_i + \frac{1}{n}\left(k_3 - E_{\theta_0}k_3\right)_i \\
&-\frac{1}{n}\sum_{j=1}^{p}\frac{\partial\left(E_{\theta_0}k_2\right)_i}{\partial\theta_j}(k_1)_j + \frac{1}{n^{\frac{3}{2}}}\left(k_4 - E_{\theta_0}k_4\right)_i \\
&-\frac{1}{n^{\frac{3}{2}}}\sum_{j=1}^{p}\frac{\partial\left(E_{\theta_0}k_3\right)_i}{\partial\theta_j}(k_1)_j - \frac{1}{n^{\frac{3}{2}}}\sum_{j=1}^{p}\frac{\partial\left(E_{\theta_0}k_2\right)_i}{\partial\theta_j}\left(k_2 - E_{\theta_0}k_2\right)_j \\
&-\frac{1}{2n^{\frac{3}{2}}}\sum_{j=1}^{p}\sum_{m=1}^{p}\frac{\partial^2\left(E_{\theta_0}k_2\right)_i}{\partial\theta_m\partial\theta_j}(k_1)_m(k_1)_j + o\left(n^{-\frac{3}{2}}\right)
\end{aligned}$$

Notice that $E_{\theta_0}\left(\theta_n^{(1}\right)_i = (\theta_0)_i - \frac{1}{2n^2}\sum_{j=1}^{p}\sum_{m=1}^{p}\frac{\partial^2\left(E_{\theta_0}k_2\right)_i}{\partial\theta_m\partial\theta_j}E_{\theta_0}\left((k_1)_m(k_1)_j\right)$ which is of order $O\left(n^{-2}\right)$, i.e. $\theta_n^{(1}$ is $O\left(n^{-\frac{3}{2}}\right)$ unbiased. Hence the proposition is true for $r = 1$.
Assume now that it is true for $r = h$, i.e. assume that, for $i = 1, ..p$ we have:

$$\begin{aligned}
\sqrt{n}\left(\theta_n^{(h} - \theta\right)_i \;=\;& (k_1)_i + \frac{1}{n^{\frac{1}{2}}}(k_2)_i + \frac{1}{n}(k_3)_i + ... + \frac{1}{n^{\frac{2h}{2}}}(k_{2h+1})_i \\
&+\frac{1}{n^{\frac{2h+1}{2}}}(k_{2h+2})_i + \frac{1}{n^{\frac{2h+2}{2}}}(k_{2h+3})_i + \frac{1}{n^{\frac{2h+3}{2}}}(k_{2h+4})_i + o\left(n^{-\frac{2h+3}{2}}\right),
\end{aligned}$$

with $E_\theta(k_1)_i = E_\theta(k_2)_i = .. = E_\theta(k_{2h+1})_i = 0$, i.e. $\theta_n^{(h}$ is $O\left(n^{-\frac{2h+1}{2}}\right)$ $\theta_0 -$ *unbiased*. Now for any $\theta \in B(\theta_0, \varepsilon)$, and any $i, j, m, l, r = 1, ...p$, we have that
$\left(E_\theta\theta_n^{(h}\right)_i = \theta_i + \frac{1}{n^{\frac{2h+2}{2}}}E_\theta(k_{2h+2})_i + \frac{1}{n^{\frac{2h+3}{2}}}E_\theta(k_{2h+3})_i + \frac{1}{n^{\frac{2h+4}{2}}}E_\theta(k_{2h+4})_i +$

$o\left(n^{-\frac{2h+4}{2}}\right)$,

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

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

$\frac{\partial^3\left(E_\theta\theta_n^{(h}\right)_i}{\partial\theta_l\partial\theta_m\partial\theta_j} = \frac{1}{n^{\frac{2h+2}{2}}}\frac{\partial^3 E_\theta (k_{2h+2})_i}{\partial\theta_l\partial\theta_m\partial\theta_j} + \frac{1}{n^{\frac{2h+3}{2}}}\frac{\partial^3 E_\theta (k_{2h+3})_i}{\partial\theta_l\partial\theta_m\partial\theta_j} + \frac{1}{n^{\frac{2h+4}{2}}}\frac{\partial^3 E_\theta (k_{2h+4})_i}{\partial\theta_l\partial\theta_m\partial\theta_j} + o\left(n^{-\frac{2h+4}{2}}\right)$ ...*etc.*

Notice that all higher order derivatives are of $O\left(n^{-\frac{2h+2}{2}}\right)$ and 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 $\theta_n^{(h} - E_{\theta_n^{(h+1}}\theta_n^{(h} = 0$. Expanding, the i^{th} element of $E_{\theta_n^{(h+1}}\theta_n^{(h}$, $\left(E_{\theta_n^{(h+1}}\theta_n^{(h}\right)_i$ say, around θ_0 we get: $\left(E_{\theta_n^{(h+1}}\theta_n^{(h}\right)_i = \left(E_{\theta_0}\theta_n^{(h}\right)_i + \sum_{j=1}^p \frac{\partial\left(E_{\theta_0}\theta_n^{(h}\right)_i}{\partial\theta_j}\left(\theta_n^{(h+1} - \theta_0\right)_j + \frac{1}{2}\sum_{j=1}^p\sum_{m=1}^p \frac{\partial^2\partial\left(E_{\theta_0}\theta_n^{(h}\right)_i}{\partial\theta_m\partial\theta_j}\left(\theta_n^{(h+1} - \theta_0\right)_m\left(\theta_n^{(h+1} - \theta_0\right)_j +$ R, where R includes 3^{rd} and higher order derivatives. Hence as $\theta_n^{(h} - E_{\theta_n^{(h+1}}\theta_n^{(h} = 0$ we get

$\sqrt{n}\left(\theta_n^{(h} - E_{\theta_0}\theta_n^{(h}\right)_i = \sum_{j=1}^p \frac{\partial\left(E_{\theta_0}\theta_n^{(h}\right)_i}{\partial\theta_j}\sqrt{n}\left(\theta_n^{(h+1} - \theta_0\right)_j$

$+\frac{1}{2\sqrt{n}}\sum_{j=1}^p\sum_{m=1}^p \frac{\partial^2\left(E_{\theta_0}\theta_n^{(h}\right)_i}{\partial\theta_m\partial\theta_j}\sqrt{n}\left(\theta_n^{(h+1} - \theta_0\right)_m\sqrt{n}\left(\theta_n^{(h+1} - \theta_0\right)_j + o\left(n^{-\frac{2h+3}{2}}\right)$

as the 3^{rd} and higher order derivatives are $O\left(n^{-\frac{2h+2}{2}}\right)$.

Further,

$$\begin{aligned}\sqrt{n}\left(\theta_n^{(h} - \theta_0\right)_i &= (k_1)_i + \frac{1}{n^{\frac{1}{2}}}(k_2)_i + \frac{1}{n}(k_3)_i + ... + \frac{1}{n^{\frac{2h}{2}}}(k_{2h+1})_i \\ &+ \frac{1}{n^{\frac{2h+1}{2}}}(k_{2h+2} - E_{\theta_0}k_{2h+2})_i + \frac{1}{n^{\frac{2h+2}{2}}}(k_{2h+3} - E_{\theta_0}k_{2h+3})_i \\ &+ \frac{1}{n^{\frac{2h+3}{2}}}(k_{2h+4} - E_{\theta_0}k_{2h+4})_i + o\left(n^{-\frac{2h+3}{2}}\right)\end{aligned}$$

and it follows that

$$\left(k_1\right)_i+\frac{1}{n^{\frac{1}{2}}}\left(k_2\right)_i+\frac{1}{n}\left(k_3\right)_i ...+\frac{1}{n^{\frac{2h}{2}}}\left(k_{2h+1}\right)_i$$
$$+\frac{1}{n^{\frac{2h+1}{2}}}\left(k_{2h+2}-E_{\theta_0}k_{2h+2}\right)_i+\frac{1}{n^{\frac{2h+2}{2}}}\left(k_{2h+3}-E_{\theta_0}k_{2h+3}\right)_i+\frac{1}{n^{\frac{2h+3}{2}}}\left(k_{2h+4}-E_{\theta_0}k_{2h+4}\right)_i$$
$$=\frac{\partial\left(E_{\theta_0}\theta_n^{(h}\right)_i}{\partial\theta_i}\sqrt{n}\left(\theta_n^{(h+1}-\theta_0\right)_i+\sum_{j\neq i=1}^{p}\frac{\partial\left(E_{\theta_0}\theta_n^{(h}\right)_i}{\partial\theta_j}\sqrt{n}\left(\theta_n^{(h+1}-\theta_0\right)_j$$
$$+\frac{1}{2\sqrt{n}}\sum_{j=1}^{p}\sum_{m=1}^{p}\frac{\partial^2\left(E_{\theta_0}\theta_n^{(h}\right)_i}{\partial\theta_m\partial\theta_j}\sqrt{n}\left(\theta_n^{(h+1}-\theta_0\right)_m\sqrt{n}\left(\theta_n^{(h+1}-\theta_0\right)_j+o\left(n^{-\frac{2h+3}{2}}\right)$$

Expanding $\left(\frac{\partial\left(E_\theta\theta_n^{(h}\right)_i}{\partial\theta_i}\right)^{-1}$ as before we get $\left(\frac{\partial\left(E_\theta\theta_n^{(h}\right)_i}{\partial\theta_i}\right)^{-1}=1-\frac{1}{n^{\frac{2h+2}{2}}}\frac{\partial E(k_{h+2})_i}{\partial\theta_j}-\frac{1}{n^{\frac{2h+3}{2}}}\frac{\partial E_\theta(k_{h+3})_i}{\partial\theta_j}-\frac{1}{n^{\frac{2h+4}{2}}}\frac{\partial E_\theta(k_{h+4})_i}{\partial\theta_j}+o\left(n^{-\frac{2h+4}{2}}\right)$. Hence solving for $\sqrt{n}\left(\theta_n^{(h+1}-\theta_0\right)_i$ we get:

$\sqrt{n}\left(\theta_n^{(h+1}-\theta_0\right)_i=\left(k_1\right)_i+\frac{1}{n^{\frac{1}{2}}}\left(k_2\right)_i+\frac{1}{n}\left(k_3\right)_i ...+\frac{1}{n^{\frac{2h}{2}}}\left(k_{2h+1}\right)_i$
$+\frac{1}{n^{\frac{2h+1}{2}}}\left(k_{2h+2}-E_{\theta_0}k_{2h+2}\right)_i+\frac{1}{n^{\frac{2h+2}{2}}}\left(k_{2h+3}-E_{\theta_0}k_{2h+3}\right)_i$
$-\frac{1}{n^{\frac{2h+2}{2}}}\sum_{j=1}^{p}\frac{\partial E_{\theta_0}(k_{2h+2})_i}{\partial\theta_j}\left(k_1\right)_j+\frac{1}{n^{\frac{2h+3}{2}}}\left(k_{2h+4}-E_{\theta_0}k_{2h+4}\right)_i$
$-\frac{1}{n^{\frac{2h+3}{2}}}\sum_{j=1}^{p}\frac{\partial E_{\theta_0}(k_{2h+2})_i}{\partial\theta_j}\left(k_2\right)_j-\frac{1}{n^{\frac{2h+3}{2}}}\sum_{j=1}^{p}\frac{\partial E_{\theta_0}(k_{2h+3})_i}{\partial\theta_j}\left(k_1\right)_j$
$-\frac{1}{2}\frac{1}{n^{\frac{2h+3}{2}}}\sum_{j=1}^{p}\sum_{m=1}^{p}\frac{\partial^2 E_{\theta_0}(k_{2h+2})_i}{\partial\theta_m\partial\theta_j}\left(k_1\right)_j\left(k_1\right)_j+o\left(n^{-\frac{2h+3}{2}}\right),$
and it follows that

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

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

Remark R.47 *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.48 *Proposition 8 essentially holds* ***locally*** *at θ_0 due to the properties of open balls as basic sets of neighborhoods (see also the example of the previous section).*

4 Conclusions

In this section we first provide a brief review of our results. These can be summarized to:

- We provide conditions that ensure the validity of the formal Edgeworth approximation of the auxiliary and the three indirect estimators for any finite order. The aforementioned validation was previously unattained by the relevant literature. We massively rely on lemma AL.3.
- Given the previous, we provide integrability conditions that validate moment approximations of the aforementioned estimators. These conditions validate the partial results of the relevant literature. We identify the approximations up to the second order. In this respect we are able to provide information on the bias structure of the estimator sequences up to the second order, in a quite general setup that incorporates frameworks of random weighting schemes.
- 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.
- We provide new results on the issue of second order properties of the three indirect estimators. First the expansions of GMR1 and GT estimators are new and reveal a higher order asymptotic inequivalence with the GMR2.
- We massively generalize the GMR2 expansion. We are able to generalize the conditions under which the GMR is second order unbiased (at θ_0) even in this set up.
- 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 deterministic weighting.
- In response to the issue of higher order bias correction, we define indirect estimators that emerge from multistep optimization procedures. If we strengthen 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.

We conclude with some possible future extensions:

- The derivation of the actual Edgeworth approximations, in the sense that the coefficients of the relevant polynomials are expresses as functions of the approximations of the auxiliary estimators, could be useful for the derivation of analogous properties of indirect testing procedures.

- The extension of the previous results in the semiparametric case.

- An interesting case lies in the possibility that $b(\theta_0)$ is in the boundary of B, even if θ_0 is in the interior of Θ, due to the fact that the binding function is not a local homeomorphism. In this case even the first order distribution of the estimator will be non standard.

- The determination of invariant parts of the expansions with respect to reparametrizations.

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] 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.

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

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

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

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

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

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

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

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

[12] Gallant A. Ronald, and Tauchen G.E., "Which Moments to Match," Working Papers 95-20, 1995, Duke University, Department of Economics.

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

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

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

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

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

Appendices

The following are a collection of helpful lemmas that are frequently referenced in the proofs of the main results.

The following lemma concerns weighting matrices and initial estimators in general, hence it is directly connected to assumptions A.9 and A.11. It provides a result useful in almost every step of the derivation of validity of the analogous Edgeworth expansion for any of the estimators examined.

Lemma AL.1 *Suppose that $W_n(\omega,\theta_n^*)$, $W(\theta_0)$, θ_n^* are defined as in assumptions A.9 and A.13, then*

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

and

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

Proof. Under assumptions A.9 and A.13, Lemmas 3 and 5 of [2], and due the triangle inequality we have that

$$\begin{aligned}
& P_\Omega\left(\|W_n(\omega,\theta_n^*)-W(\theta_0)\|>\varepsilon\right)\\
\leq\ & P_\Omega\left(\|W_n(\omega,\theta_0)-W(\theta_0)\|+\|W_n(\omega,\theta_n^*)-W_n(\omega,\theta_0)\|>\varepsilon\right)\\
\leq\ & P_\Omega\left(\|W_n(\omega,\theta_0)-W(\theta_0)\|>\frac{\varepsilon}{2}\right)+P_\Omega\left(\|W_n(\omega,\theta_n^*)-W_n(\omega,\theta_0)\|>\frac{\varepsilon}{2}\right)\\
\leq\ & o\left(n^{-a}\right)+P_\Omega\left(u_n\|\theta_n^*-\theta_0\|>\frac{\varepsilon}{2}\right)=o\left(n^{-a}\right)
\end{aligned}$$

and

$$\begin{aligned}
& P_\Omega\left(\left\|D^rW_n(\omega,\theta)|_{\theta=\theta_n^*}-D^rW(\theta)|_{\theta=\theta_0}\right\|>\varepsilon\right)\\
\leq\ & P_\Omega\left(\|D^rW_n(\omega,\theta)|_{\theta=\theta_0}-D^rW(\theta)|_{\theta=\theta_0}\|+\left\|D^rW_n(\omega,\theta)|_{\theta=\theta_n^*}-D^rW_n(\omega,\theta)|_{\theta=\theta_0}\right\|>\varepsilon\right)\\
\leq\ & P_\Omega\left(\left\|D^rW_n(\omega,\theta)|_{\theta=\theta_n^*}-D^rW_n(\omega,\theta)|_{\theta=\theta_0}\right\|>\frac{\varepsilon}{2}\right)\\
& +P_\Omega\left(\|D^rW_n(\omega,\theta)|_{\theta=\theta_0}-D^rW(\theta)|_{\theta=\theta_0}\|>\frac{\varepsilon}{2}\right)\\
\leq\ & P_\Omega\left(u_{Dn}\|\theta_n^*-\theta_0\|>\frac{\varepsilon}{2}\right)+o\left(n^{-a}\right)=o\left(n^{-a}\right).
\end{aligned}$$

■

In the following we 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 J its probability limit. Remember also that $d = \max(2a+2, 3)$. Our next lemma concerns the derivation of the validity of the Edgeworth expansion in any of the 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.2 *1.* $P_{\theta_0}\left(\left\|n^{\frac{1}{2}}(\theta_n - \theta_0)\right\| > C \ln^{1/2} n\right) = o(n^{-a})$, $P_{\theta_0}\left(\left\|n^{\frac{1}{2}}(\varphi_n - \varphi_0)\right\| > C^* \ln^{1/2} n\right) = o(n^{-a})$ *for* $C, C^* > 0$,
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* $1 - o(n^{-a})$, *and* $\frac{\partial J^2(\theta_0,\varphi_0)}{\partial\theta\partial\theta'}$ *is positive definite,*
3. $P_{\theta_0}\left(\left\|n^{\frac{1}{2}}(\varphi_n - \varphi_0) - n^{\frac{1}{2}}\pi(R_n)\right\| > \omega_n^*\right) = o(n^{-a})$ *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* $\omega_n = o(n^{-a})$ *deterministic.*

Proof. By the previous remark, 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

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

where the remainder
$r_n^* = \sum_{j_1=0}^{d-1} \frac{1}{(d-1)!}\binom{d-1}{j_1} \left(D^{j_1,d-1-j_1}\left(\frac{\partial J_n\left(\theta_n^+,\varphi_n^+\right)}{\partial\theta}\right) - D^{j_1,d-1-j_1}\left(\frac{\partial J_n((\theta_0,\varphi_0))}{\partial\theta}\right)\right)$
$\times \begin{pmatrix} \underbrace{\theta_n-\theta_0,\cdots,\theta_n-\theta_0}_{j_1 \text{ times}}, \\ \underbrace{\varphi_n-\varphi_0,\cdots,\varphi_n-\varphi_0}_{j_2 \text{ times}} \end{pmatrix}$ and each $D^{i_1,i_2}\left(\frac{\partial J_n(\theta_0,\theta_0)}{\partial\theta}\right)$ 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 $\frac{\partial J_n(\theta,\varphi)}{\partial\varphi}$ with respect to the components of the final φ at (θ_0,φ_0). Hence the previous can be rewritten as ([2], lemma 8)

$$v^*\left(\theta_n-\theta_0,\varphi_n-\varphi_0,R_n^*+\epsilon_n^*\right)=\mathbf{0}_p$$

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 $v\left(\boldsymbol{\pi}^*(x,y),x,y\right)=\mathbf{0}_p\ \forall(x,y)\in V_{(\mathbf{0}_q,\mathbf{R}^*)}$. But we know that $P_{\theta_0}\left(\theta_n-\theta_0\in U_{\mathbf{0}_p}\right)=1-o\left(n^{-a}\right)$, $P_{\theta_0}\left(\varphi_n-\varphi_0\in U_{\mathbf{0}_q}\right)=1-o\left(n^{-a}\right)$ $P_{\theta_0}\left(R_n^*+\epsilon_n^*-R^*\in U_{\mathbf{0}_m}\right)=1-o\left(n^{-a}\right)$ and $\varphi_n-\varphi_0=\pi\left(R_n\right)+\epsilon_n$ where $P_{\theta_0}\left(\sqrt{n}\left\|\epsilon_n\right\|>\omega_n\right)=o\left(n^{-a}\right)$, $\omega_n=o\left(n^{-a}\right)$ and $\pi(.)$ is a smooth we obtain that for large enough n

$$\theta_n-\theta_0=\pi^*\left(\varphi_n-\varphi_0,R_n^*+\epsilon_n^*\right)=\pi^*\left(\pi\left(R_n\right)+\epsilon_n,R_n^*+\epsilon_n^*\right)=\pi^*(x,y)$$

with $\pi^*\left(\mathbf{0}_q,R^*\right)=\mathbf{0}_p$. Now due to the smoothness of π^*, we have that

$$\begin{aligned} & \left\|\pi^*\left(\pi\left(R_n\right)+\epsilon_n,R_n^*+\epsilon_n^*\right)-\pi^*\left(\pi\left(R_n\right),R_n^*\right)\right\| \\ \leq & \left\|\frac{\partial\pi^*\left(\pi\left(R_n\right)+\epsilon_n^+,R_n^*+\epsilon_n^{*+}\right)}{\partial x'}\epsilon_n+\frac{\partial\pi^*\left(\pi\left(R_n\right)+\epsilon_n^+,R_n^*+\epsilon_n^{*+}\right)}{\partial y'}\epsilon_n^*\right\| \\ \leq & \left\|\frac{\partial\pi^*\left(\pi\left(R_n\right)+\epsilon_n^+,R_n^*+\epsilon_n^{*+}\right)}{\partial x'}\epsilon_n\right\|+\left\|\frac{\partial\pi^*\left(\pi\left(R_n\right)+\epsilon_n^+,R_n^*+\epsilon_n^{*+}\right)}{\partial y'}\epsilon_n^*\right\| \end{aligned}$$

where ϵ_n^+ is in the line segment between $\mathbf{0}_k$ and ϵ_n and ϵ_n^{*+} is in the line segment between $\mathbf{0}_m$ and ϵ_n^*, we have that $P_{\theta_0}\left(\left\|\epsilon_n\right\|>\varepsilon\right)=o\left(n^{-a}\right)$ and from continuous mapping theorem (applicable due to smoothness of π^* and π)

$P_{\theta_0}\left(\left\|\frac{\partial\pi^*\left(\pi(R_n)+\epsilon_n^+,R_n^*+\epsilon_n^{*+}\right)}{\partial x'}\right\|>K_1\right)=o\left(n^{-a}\right)$ for some $K_1>0$ and $P_{\theta_0}\left(\left\|\frac{\partial\pi^*\left(\pi(R_n)+\epsilon_n^+,R_n^*+\epsilon_n^{*+}\right)}{\partial y'}\right\|>K_2\right)=o\left(n^{-a}\right)$ for some $K_2>0$. Also we know

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

$$P_{\theta_0}\left(n^{1/2}\left\|\varphi_n-\varphi_0\right\|>C_*\ln^{\frac{1}{2}}n\right)=o\left(n^{-a}\right)$$

and we choose $\omega_n=Cn^{\frac{1}{2}-\frac{d}{2}}\ln^{\frac{d}{2}}n=o\left(n^{-a}\right)$, where $C>0$, to be determined, since $d=\max\left(2a+2,3\right)$. Now, from above

$$\begin{aligned}
& P_{\theta_0}\left(n^{\frac{1}{2}}\left\|\pi^*\left(\pi\left(R_n\right)+\epsilon_n,R_n^*+\epsilon_n^*\right)-\pi^*\left(\pi\left(R_n\right),R_n^*\right)\right\|>Cn^{\frac{1}{2}-\frac{d}{2}}\ln^{\frac{d}{2}}n\right)\\
\leq\; & P_{\theta_0}\left(n^{\frac{1}{2}}\left\|\frac{\partial\pi^*\left(\pi\left(R_n\right)+\epsilon_n^+,R_n^*+\epsilon_n^{*+}\right)}{\partial x'}\epsilon_n\right\|>C_1n^{\frac{1}{2}-\frac{d}{2}}\ln^{\frac{d}{2}}n\right)\\
& +P_{\theta_0}\left(n^{\frac{1}{2}}\left\|\frac{\partial\pi^*\left(\pi\left(R_n\right)+\epsilon_n^+,R_n^*+\epsilon_n^{*+}\right)}{\partial y'}\epsilon_n^*\right\|>C_2n^{\frac{1}{2}-\frac{d}{2}}\ln^{\frac{d}{2}}n\right)
\end{aligned}$$

with C_1 and C_2 due to 3, given C. Now let $\varepsilon\left(n\right)=C_1n^{\frac{1}{2}-\frac{d}{2}}\ln^{\frac{d}{2}}n$ then

$P_{\theta_0}\left(n^{\frac{1}{2}}\left\|\frac{\partial\pi^*\left(\pi(R_n)+\epsilon_n^+,R_n^*+\epsilon_n^{*+}\right)}{\partial y'}\epsilon_n^*\right\|>\varepsilon\left(n\right)\right)\leq$
$P_{\theta_0}\left(\left\|\frac{\partial\pi^*\left(\pi(R_n)+\epsilon_n^+,R_n^*+\epsilon_n^{*+}\right)}{\partial y'}\right\|n^{\frac{1}{2}}\left\|\epsilon_n^*\right\|>\varepsilon\left(n\right)\right)=$
$P_{\theta_0}\left(\left\|\frac{\partial\pi^*\left(\pi(R_n)+\epsilon_n^+,R_n^*+\epsilon_n^{*+}\right)}{\partial y'}\right\|n^{\frac{1}{2}}\left\|r_n^*\right\|>\varepsilon\left(n\right)\right)=$
$P_{\theta_0}\left(\left\|\frac{\partial\pi^*\left(\pi(R_n)+\epsilon_n^+,R_n^*+\epsilon_n^{*+}\right)}{\partial y'}\right\|n^{\frac{1}{2}}\left\|r_n^*\right\|>\varepsilon\left(n\right)\cap\left\|\frac{\partial\pi^*\left(\pi(R_n)+\epsilon_n^+,R_n^*+\epsilon_n^{*+}\right)}{\partial y'}\right\|>K_2\right)$
$+P_{\theta_0}\left(\left\|\frac{\partial\pi^*\left(\pi(R_n)+\epsilon_n^+,R_n^*+\epsilon_n^{*+}\right)}{\partial y'}\right\|n^{\frac{1}{2}}\left\|r_n^*\right\|>\varepsilon\left(n\right)\cap\left\|\frac{\partial\pi^*\left(\pi(R_n)+\epsilon_n^+,R_n^*+\epsilon_n^{*+}\right)}{\partial y'}\right\|\leq K_2\right)\leq$
$P_{\theta_0}\left(\left\|\frac{\partial\pi^*\left(\pi(R_n)+\epsilon_n^+,R_n^*+\epsilon_n^{*+}\right)}{\partial y'}\right\|>K_2\right)+P_{\theta_0}\left(K_2n^{\frac{1}{2}}\left\|r_n^*\right\|>\varepsilon\left(n\right)\right)\leq$
$o\left(n^{-a}\right)+$

$$P_{\theta_0}\left(K_2n^{\frac{1}{2}}\sum_{j_1=0}^{d-1}\frac{\binom{d-1}{j_1}}{(d-1)!}\left\|\begin{pmatrix}D^{j_1,d-1-j_1}\left(\frac{\partial J_n\left(\theta_n^+,\varphi_n^+\right)}{\partial\theta}\right)\\-D^{j_1,d-1-j_1}\left(\frac{\partial J_n\left(\left(\theta_0,\varphi_0\right)\right)}{\partial\theta}\right)\end{pmatrix}\times\begin{pmatrix}\underbrace{\theta_n-\theta_0,\cdots,\theta_n-\theta_0}_{j_1\text{ times}},\\\underbrace{\varphi_n-\varphi_0,\cdots,\varphi_n-\varphi_0}_{j_2\text{ times}}\end{pmatrix}\right\|>\varepsilon\left(n\right)\right)\leq$$

$$o\left(n^{-a}\right)+P_{\theta_0}\left(\begin{array}{c} K_2 n^{\frac{1}{2}}\sum_{j_1=0}^{d-1}\frac{1}{(d-1)!}\binom{d-1}{j_1} \\ \times\left[\left\|\left(\begin{array}{c} D^{j_1,d-1-j_1}\left(\frac{\partial J_n\left(\theta_n^+,\varphi_n^+\right)}{\partial\theta}\right) \\ -D^{j_1,d-1-j_1}\left(\frac{\partial J_n((\theta_0,\varphi_0))}{\partial\theta}\right)\end{array}\right)\right\|\right. \\ \left.\times\left\|\theta_n-\theta_0\right\|^{j_1}\left\|\varphi_n-\varphi_0\right\|^{d-1-j_1}\right]\end{array} > \varepsilon(n)\right)$$

$$\leq o\left(n^{-a}\right)+P_{\theta_0}\left(K_2 n^{\frac{1}{2}}\sum_{j_1=0}^{d-1}\left\{\begin{array}{c}\frac{\binom{d-1}{j_1}}{(d-1)!}\left|L_{n,j_1}\right|\left[\left\|\theta_n^+-\theta_0\right\|+\left\|\varphi_n^+-\varphi_0\right\|\right] \\ \times\left\|\theta_n-\theta_0\right\|^{j_1}\left\|\varphi_n-\varphi_0\right\|^{d-1-j_1}\end{array}\right\} > \varepsilon(n)\right)$$

$$\leq o\left(n^{-a}\right)+P_{\theta_0}\left(K_2 n^{\frac{1}{2}}\sum_{j_1=0}^{d-1}\left\{\begin{array}{c}\frac{\binom{d-1}{j_1}}{(d-1)!}\left|L_{n,j_1}\right|\left[\left\|\theta_n^+-\theta_0\right\|+\left\|\varphi_n^+-\varphi_0\right\|\right] \\ \times\left\|\theta_n-\theta_0\right\|^{j_1}\left\|\varphi_n-\varphi_0\right\|^{d-1-j_1}\end{array}\right\} > \varepsilon(n)\right)$$

Assuming there are $K_{j_1} > 0$ such that $P_{\theta_0}\left(\left|L_{n,j_1}\right| > K_{j_1}\right) = o\left(n^{-a}\right)$, something true in our case due to 2,we can find $K > 0$ such $P_{\theta_0}\left(\max_{j_1}\left|L_{n,j_1}\right| > K\right) = o\left(n^{-a}\right)$. But $P_{\theta_0}\left(\max_{j_1}\left|L_{n,j_1}\right| > K\right) = P_{\theta_0}\left(\cup_{j_1}\left|L_{n,j_1}\right| > K\right) \leq \sum_{j_1} P_{\theta_0}\left(\left|L_{n,j_1}\right| > K\right) =$

$\sum_{j_1} o\left(n^{-a}\right)$ for $K = \max_{j_1}\left\{K_{n,j_1}\right\}$. Hence

$$\begin{aligned}
& P_{\theta_0}\left(n^{\frac{1}{2}}\left\|\frac{\partial \pi^*\left(\pi\left(R_n\right)+\epsilon_n^+, R_n^*+\epsilon_n^{*+}\right)}{\partial y'}\epsilon_n^*\right\| > C_2 n^{\frac{1}{2}-\frac{d}{2}}\ln^{\frac{d}{2}} n\right) \\
\leq\ & o\left(n^{-a}\right)+P_{\theta_0}\left(K_2 n^{\frac{1}{2}}\sum_{j_1=0}^{d-1}\left\{\begin{array}{c}\frac{1}{(d-1)!}\binom{d-1}{j_1}\left|L_{n,j_1}\right|\left[\left\|\theta_n^+-\theta_0\right\|+\left\|\varphi_n^+-\varphi_0\right\|\right] \\ \times\left\|\theta_n-\theta_0\right\|^{j_1}\left\|\varphi_n-\varphi_0\right\|^{d-1-j_1}\end{array}\right\} > \varepsilon(n)\right) \\
\leq\ & o\left(n^{-a}\right)+P_{\theta_0}\left(K_2 n^{\frac{1}{2}}\max_{j_1}\left|L_{n,j_1}\right|\sum_{j_1=0}^{d-1}\left\{\begin{array}{c}\frac{1}{(d-1)!}\binom{d-1}{j_1}\left[\left\|\theta_n^+-\theta_0\right\|+\left\|\varphi_n^+-\varphi_0\right\|\right] \\ \times\left\|\theta_n-\theta_0\right\|^{j_1}\left\|\varphi_n-\varphi_0\right\|^{d-1-j_1}\end{array}\right\} > \varepsilon(n)\right) \\
=\ & o\left(n^{-a}\right)+P_{\theta_0}\left(\begin{array}{c}K_2 n^{\frac{1}{2}}\max_{j_1}\left|L_{n,j_1}\right|\sum_{j_1=0}^{d-1}\left\{\begin{array}{c}\frac{1}{(d-1)!}\binom{d-1}{j_1}\left[\left\|\theta_n^+-\theta_0\right\|+\left\|\varphi_n^+-\varphi_0\right\|\right] \\ \times\left\|\theta_n-\theta_0\right\|^{j_1}\left\|\varphi_n-\varphi_0\right\|^{d-1-j_1}\end{array}\right\} > \varepsilon(n) \\ \cap\max_{j_1}\left|L_{n,j_1}\right| > K\end{array}\right) \\
& +P_{\theta_0}\left(\begin{array}{c}K_2 n^{\frac{1}{2}}\max_{j_1}\left|L_{n,j_1}\right|\sum_{j_1=0}^{d-1}\left\{\begin{array}{c}\frac{1}{(d-1)!}\binom{d-1}{j_1}\left[\left\|\theta_n^+-\theta_0\right\|+\left\|\varphi_n^+-\varphi_0\right\|\right] \\ \times\left\|\theta_n-\theta_0\right\|^{j_1}\left\|\varphi_n-\varphi_0\right\|^{d-1-j_1}\end{array}\right\} > \varepsilon(n) \\ \cap\max_{j_1}\left|L_{n,j_1}\right| \leq K\end{array}\right) \\
\leq\ & o\left(n^{-a}\right)+P_{\theta_0}\left(\max_{j_1}\left|L_{n,j_1}\right| > K\right) \\
& +P_{\theta_0}\left(K_2 n^{\frac{1}{2}}K\sum_{j_1=0}^{d-1}\left\{\begin{array}{c}\frac{1}{(d-1)!}\binom{d-1}{j_1}\left[\left\|\theta_n^+-\theta_0\right\|+\left\|\varphi_n^+-\varphi_0\right\|\right] \\ \times\left\|\theta_n-\theta_0\right\|^{j_1}\left\|\varphi_n-\varphi_0\right\|^{d-1-j_1}\end{array}\right\} > \varepsilon(n)\right) \\
=\ & o\left(n^{-a}\right)+P_{\theta_0}\left(\begin{array}{c}K_2 K\sum_{j_1=0}^{d-1}\left\{\begin{array}{c}\frac{1}{(d-1)!}\binom{d-1}{j_1}\left[n^{\frac{1}{2}}\left\|\theta_n^+-\theta_0\right\|+n^{\frac{1}{2}}\left\|\varphi_n^+-\varphi_0\right\|\right] \\ \times\left\|\theta_n-\theta_0\right\|^{j_1}\left\|\varphi_n-\varphi_0\right\|^{d-1-j_1}\end{array}\right\} > \varepsilon(n) \\ \cap\left\|\varphi_n-\varphi_0\right\| > C_* n^{-\frac{1}{2}}\ln^{\frac{1}{2}} n\end{array}\right) \\
& +P_{\theta_0}\left(\begin{array}{c}K_2 K\sum_{j_1=0}^{d-1}\left\{\begin{array}{c}\frac{1}{(d-1)!}\binom{d-1}{j_1}\left[n^{\frac{1}{2}}\left\|\theta_n^+-\theta_0\right\|+n^{\frac{1}{2}}\left\|\varphi_n^+-\varphi_0\right\|\right] \\ \times\left\|\theta_n-\theta_0\right\|^{j_1}\left\|\varphi_n-\varphi_0\right\|^{d-1-j_1}\end{array}\right\} > \varepsilon(n) \\ \cap\left\|\varphi_n-\varphi_0\right\| \leq C_* n^{-\frac{1}{2}}\ln^{\frac{1}{2}} n\end{array}\right)
\end{aligned}$$

$$
\begin{aligned}
&\leq \; o\left(n^{-a}\right)+P_{\theta_0}\left(\left\|\varphi_n-\varphi_0\right\|>C_* n^{-\frac{1}{2}}\ln^{\frac{1}{2}} n\right) \\
&\quad +P_{\theta_0}\left(K_2 K\sum_{j_1=0}^{d-1}\left\{\begin{array}{c}\frac{1}{(d-1)!}\binom{d-1}{j_1}\left[n^{\frac{1}{2}}\left\|\theta_n^{+}-\theta_0\right\|+C_*\ln^{\frac{1}{2}} n\right] \\ \times\left\|\theta_n-\theta_0\right\|^{j_1} C_*^{d-1-j_1} n^{-\frac{d-1-j_1}{2}}\ln^{\frac{d-1-j_1}{2}} n\end{array}\right\}>\varepsilon(n)\right) \\
&= \; o\left(n^{-a}\right)+P_{\theta_0}\left(\begin{array}{c} K_2 K\sum_{j_1=0}^{d-1}\left\{\begin{array}{c}\frac{1}{(d-1)!}\binom{d-1}{j_1}\left[n^{\frac{1}{2}}\left\|\theta_n^{+}-\theta_0\right\|+C_*\ln^{\frac{1}{2}} n\right] \\ \times\left\|\theta_n-\theta_0\right\|^{j_1} C_*^{d-1-j_1} n^{-\frac{d-1-j_1}{2}}\ln^{\frac{d-1-j_1}{2}} n\end{array}\right\}>\varepsilon(n) \\ \cap\left\|\theta_n-\theta_0\right\|>n^{-\frac{1}{2}} Q\ln^{\frac{1}{2}} n\end{array}\right) \\
&\quad +P_{\theta_0}\left(\begin{array}{c} K_2 K\sum_{j_1=0}^{d-1}\left\{\begin{array}{c}\frac{1}{(d-1)!}\binom{d-1}{j_1}\left[n^{\frac{1}{2}}\left\|\theta_n^{+}-\theta_0\right\|+C_*\ln^{\frac{1}{2}} n\right] \\ \times\left\|\theta_n-\theta_0\right\|^{j_1} C_*^{d-1-j_1} n^{-\frac{d-1-j_1}{2}}\ln^{\frac{d-1-j_1}{2}} n\end{array}\right\}>\varepsilon(n) \\ \cap\left\|\theta_n-\theta_0\right\|\leq n^{-\frac{1}{2}} Q\ln^{\frac{1}{2}} n\end{array}\right) \\
&\leq \; o\left(n^{-a}\right)+P_{\theta_0}\left(\left\|\theta_n-\theta_0\right\|>n^{-\frac{1}{2}} Q\ln^{\frac{1}{2}} n\right) \\
&\quad +P_{\theta_0}\left(K_2 K\sum_{j_1=0}^{d-1}\left\{\frac{\binom{d-1}{j_1}}{(d-1)!}\left[Q\ln^{\frac{1}{2}} n+C_*\ln^{\frac{1}{2}} n\right] Q^{j_1} C_*^{d-1-j_1} n^{-\frac{d-1}{2}}\ln^{\frac{d-1}{2}} n\right\}>\varepsilon(n)\right) \\
&= \; o\left(n^{-a}\right)
\end{aligned}
$$

for $C_1>K_2 K\sum_{j_1=0}^{d-1}\left\{\frac{1}{(d-1)!}\binom{d-1}{j_1}\left[Q+C_*\right] Q^{j_1} C_*^{d-1-j_1}\right\}$.
When now $J_n(\theta,\varphi)=Q_n^{/}(\theta,\varphi) W(\varphi) Q_n(\theta,\varphi)$ and consequently $\frac{\partial J_n(\theta,\varphi)}{\partial\theta}=\frac{\partial Q_n^{/}(\theta,\varphi)}{\partial\theta} W(\varphi) Q_n(\theta,\varphi)$ then

$$
D^{j_1,j_2}\left(\frac{\partial J_n((\theta_0,\varphi_0))}{\partial\theta}\right)\left(\begin{array}{c}\underbrace{\theta_n-\theta_0,\cdots,\theta_n-\theta_0,}_{j_1\text{ times}} \\ \underbrace{\varphi_n-\varphi_0,\cdots,\varphi_n-\varphi_0}_{j_2\text{ times}}\end{array}\right)=
$$

$$
=D^{j_1,j_2}\left(\frac{\partial Q_n^{/}(\theta_0,\varphi_0)}{\partial\theta} W(\varphi_0)\right)\left(\begin{array}{c}\underbrace{\theta_n-\theta_0,\cdots,\theta_n-\theta_0,}_{j_1\text{ times}} \\ \underbrace{\varphi_n-\varphi_0,\cdots,\varphi_n-\varphi_0}_{j_2\text{ times}}\end{array}\right) Q_n(\theta_0,\varphi_0)
$$

$$
+\frac{\partial Q_n^{/}(\theta_0,\varphi_0)}{\partial\theta} W(\varphi_0) D^{j_1,j_2}\left(Q_n(\theta_0,\varphi_0)\right)\left(\begin{array}{c}\underbrace{\theta_n-\theta_0,\cdots,\theta_n-\theta_0,}_{j_1\text{ times}} \\ \underbrace{\varphi_n-\varphi_0,\cdots,\varphi_n-\varphi_0}_{j_2\text{ times}}\end{array}\right).
$$

Then when $j_1+j_2=d-1$, then

$$P\left(\sqrt{n}\left\|D^{j_1,j_2}\left(\frac{\partial Q_n^{/}(\theta_0,\varphi_0)}{\partial\theta}W(\varphi_0)\right)\begin{pmatrix}\underbrace{\theta_n-\theta_0,\cdots,\theta_n-\theta_0}_{j_1\text{ times}},\\\underbrace{\varphi_n-\varphi_0,\cdots,\varphi_n-\varphi_0}_{j_2\text{ times}}\end{pmatrix}Q_n(\theta_0,\varphi_0)\right\|>\omega_n\right)=$$

$o(n^{-a})$ where $\omega_n=o(n^{-a})$ (to be determined) are not included in $\pi^*(R_n^*)$. Consequently, they are included in ϵ_n^*. The determination of ω_n is achieved since:

$$P\left(\sqrt{n}\left\|D^{j_1,j_2}\left(\frac{\partial Q_n^{/}(\theta_0,\varphi_0)}{\partial\theta}W(\varphi_0)\right)\begin{pmatrix}\underbrace{\theta_n-\theta_0,\cdots,\theta_n-\theta_0}_{j_1\text{ times}},\\\underbrace{\varphi_n-\varphi_0,\cdots,\varphi_n-\varphi_0}_{j_2\text{ times}}\end{pmatrix}Q_n(\theta_0,\varphi_0)\right\|>\omega_n\right)=$$

$$=P\left((\sqrt{n})^{-(d-1)}\left\|\begin{array}{c}D^{j_1,j_2}\left(\frac{\partial Q_n^{/}(\theta_0,\varphi_0)}{\partial\theta}W(\varphi_0)\right)\begin{pmatrix}\underbrace{\sqrt{n}(\theta_n-\theta_0),\cdots,\sqrt{n}(\theta_n-\theta_0)}_{j_1\text{ times}},\\\underbrace{\sqrt{n}(\varphi_n-\varphi_0),\cdots,\sqrt{n}(\varphi_n-\varphi_0)}_{j_2\text{ times}}\end{pmatrix}\\\times\sqrt{n}Q_n(\theta_0,\varphi_0)\end{array}\right\|>\omega_n\right)$$

$$\leq P\left(\begin{array}{c}(\sqrt{n})^{-(d-1)}\left\|D^{j_1,j_2}\left(\frac{\partial Q_n^{/}(\theta_0,\varphi_0)}{\partial\theta}W(\varphi_0)\right)\right\|\\\times\|\sqrt{n}(\varphi_n-\varphi_0)\|^{j_2}\|\sqrt{n}(\theta_n-\theta_0)\|^{j_1}\|\sqrt{n}Q_n(\theta_0,\varphi_0)\|>\omega_n\end{array}\right)$$

$$\leq o(n^{-a})+P\left((\sqrt{n})^{-(d-1)}K\|\sqrt{n}(\varphi_n-\varphi_0)\|^{j_2}\|\sqrt{n}(\theta_n-\theta_0)\|^{j_1}\|\sqrt{n}Q_n(\theta_0,\varphi_0)\|>\omega_n\right)$$

$$\leq o(n^{-a})+P\left((\sqrt{n})^{-(d-1)}K(C_\varphi)^{j_2}(C_\theta)^{j_1}C_Q\ln^{\frac{d}{2}}n>\omega_n\right)$$

which is true for $\omega_n=Cn^{-\frac{d-1}{2}}\ln^{\frac{d}{2}}n$ where $C>K(C_\varphi)^{j_2}(C_\theta)^{j_1}C_Q$. ■

The next two results are of great importance in both the validity of Edgeworth expansions as well as in the validation of moment approximations.

Lemma AL.3 *Suppose that $\sqrt{n}(\beta_n-b(\theta_0))$ 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 $a_n=o(n^{-\varepsilon})$ and $P(\sqrt{n}\|x_n\|>a_n)=o(n^{-a})$. Then $\sqrt{n}(\beta_n-b(\theta_0)+x_n)$ admits a valid Edgeworth expansion of the same order.*

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

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

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

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

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

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

DEPARTMENT OF ECONOMICS

D I S C U S S I O N   P A P E R S

(An electronic version of the paper may be downloaded from: www.econ.aueb.gr)

191. **Nicholas C. Baltas**. AN ANALYSIS OF INVESTMENT ACTIVITY IN THE GREEK AGRICULTURAL PRODUCTS AND FOOD MANUFACTURING SECTOR. Forthcoming in: *Journal of Economic Asymmetries*, 2008.
192. **George C. Bitros**, AUEB. THE PROPORTIONALITY HYPOTHESIS IN CAPITAL THEORY: AN ASSESSMENT OF THE LITERATURE. (March 2008).
193. **George C. Bitros**, AUEB. AGGREGATION OF PRODUCER DURABLES WITH EXOGENOUS TECHNICAL CHANGE AND ENDOGENOUS USEFUL LIVES. Published in: *Journal of Economic and Social Measurement*, 34 (2009), 133-158.
194. **Efthymios Tsionas,** AUEB, **Nicholas C. Baltas**, AUEB and **Dionysios P. Chionis**, Democritus Univ. of Thrace. COST STRUCTURE, EFFICIENCY AND PRODUCTIVITY IN HELLENIC RAILWAYS. Published in: *Journal of Economic Asymmetries*, Vol. 5, No.1, pp. 39-52, 2008.
195. **George C. Bitros**, AUEB. THE THEOREM OF PROPORTIONALITY IN MAINSTREAM CAPITAL THEORY: AN ASSESSMENT OF ITS CONCEPTUAL FOUNDATIONS. (This paper has evolved from Discussion Paper No.192 under the title "The Hypothesis of Proportionality in Capital Theory: An Assessment of the Literature". The present covers only the theoretical literature and it will be supplemented by another paper covering the empirical literature.) [Sept. 2009].
196. **George C. Bitros,** AUEB. THE THEOREM OF PROPORTIONALITY IN MAINSTREAM CAPITAL THEORY: AN ASSESSMENT OF ITS APPLICABILITY. (This paper has evolved from Discussion Paper No.192 under the title "The Hypothesis of Proportionality in Capital Theory: An Assessment of the Literature". It surveys the empirical literature and accompanies Discussion Paper No. 195 under the title "The theorem of Proportionality in Mainstream Capital theory: An Assessment of its Conceptual Foundations.) Forthcoming in a slightly revised version in the *Journal of Economic and Social Measurement*.
197. **George D. Demopoulos,** European Chair Jean Monnet and AUEB, **Nicholas A. Yannacopoulos**, University of Piraeus, **and Athanassios N. Yannacopoulos,** AUEB. THEORY AND POLICY IN MONETARY UNIONS: INDETERMINACY AND OPTIMAL CONTROL. (Feb. 2010).
198. **Stelios Arvanitis** and **Antonis Demos**, AUEB. STOCHASTIC EXPANSIONS AND MOMENT APPROXIMATIONS FOR THREE INDIRECT ESTIMATORS. (June 2010).