SIMULATED ASYMPTOTIC LEAST SQUARES THEORY*

by

Ramdan Dridi
London School of Economics and Political Science

Contents:

The Suntory Centre
Suntory and Toyota International Centres
for Economics and Related Disciplines
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
Tel.: 020-7405 7686

Discussion Paper
No. EM/00/396
June 2000

---

* I thank Russell Davidson, Vassilis Hajivassiliou, Lars Hansen, Javier Hidalgo, Alain Monfort, Whitney Newey, Eric Renault, Peter Robinson, Enrique Santana, and seminar audiences at LSE, ESRC Bristol, ISI Helsinki, ESEM Compostella, GREQAM Marseille and CEMFI Madrid for helpful comments All remaining errors are my own.

Abstract

We develop in this paper a general econometric methodology referred to as the Simulated Asymptotic Least Squares (SALS). It is shown that this approach provides a unifying theory for “approximation-based” or simulation-based inference methods and nests the Simulated Nonlinear Least Squares (SSNLS), the Simulated Pseudo Maximum Likelihood (SPML), the Simulated Method of Moments (SMM) in both parametric and semiparametric settings, the Indirect Inference (II) and the Efficient Method of Moments (EMM).

We produce a new notion of Efficiency Bounds in Direction and provide a general study of the efficiency in the SALS framework.

In the particular case of the II and the EMM methods and when the instrumental model is of a GMM type, we characterise a new weighting matrix for a more efficient estimation about the structural parameters of interest θ^0. This new weighting matrix does no longer correspond, in the general case, to the classical one as characterised by Hansen (1982). Generalized global specification tests extending the previous existing ones are also proposed.

Keywords: Simulated Asymptotic Least Squares; Approximation-based and Simulation-based estimation; Efficiency Bounds in Direction; GMM; SNLS; SPML; SMM; II; GII; EMM.

JEL Nos.: C13, C14, C15, C52.

1 Introduction

Econometric models often lead nowadays to a complex formulation of either conditional probability distribution function (p.d.f. hereafter) of the endogenous variables given the exogenous and predetermined ones in the fully parametric setting or more generally to a complex formulation of the so-called **estimating equations** in the semiparametric setting. Optimizing behavior and optimizing subroutines are such that one cannot directly write the functional forms associated with a given parametrization.

In this context, several approaches depending on the primitive setting and circumventing such difficulties have been introduced in the literature: the Simulated Pseudo Maximum Likelihood (SPML hereafter) by Laroque and Salanié (1989), the Simulated Method of Moments (SMM hereafter) by McFadden (1989), Pakes and Pollard (1989), Ingram and Lee (1991), Duffie and Singleton (1993), the Simulated Nonlinear Least Squares (SNLS hereafter) by Laffont, Ossard and Vuong (1995) and more recently the Indirect Inference (II hereafter) and the Efficient Method of Moments (EMM hereafter) respectively by Gouriéroux, Monfort and Renault (1993) and Gallant and Tauchen (1996).

However, it turns out, on the one hand, that so far no general econometric theory enabling the unification of all the aforementioned methodologies has been proposed. On the other hand, there are now cases arising from the macroeconometric as well as the econometric literature where the sole application of such approaches does not fully exploit the information brought about by the available estimating equations. This occurs, for instance, when one has at his disposal overidentifying moment restrictions defining a set of instrumental parameters in conjunction with a fully specified parametric structural model.

In this respect, we propose in this paper a general econometric theory referred to as the Simulated Asymptotic Least Squares (SALS hereafter). It is shown that this approach provides a unifying theory for simulation-based or more generally approximation-based inference methods and nests the SPML, the SMM both in parametric and semiparametric settings, the II, the EMM and the SNLS approaches.

It can indeed be regarded, on the one hand, as a simulated or approximate extension of the earlier ALS theory introduced by Gouriéroux, Monfort and Trognon (1985) to the case where the estimating equations are intractable but can be approximated in some sense either by simulations or more generally by approximation methods such as for instance the quadrature-based methods or Marcet parametrized expectations type procedures.

But on the other hand, it can also be regarded as a generalization of the ALS theory to the case where the number of estimating equations r (say) is bigger than the number of auxiliary parameters q (say), that is $r > q$. Indeed, we stress in this paper that while in Gouriéroux, Monfort and Trognon (1985) the ALS theory is developed in the particular case where $r \leq q$, there are now numerous examples where $r > q$. We provide here a general study dealing with both issues. In this respect we think that the (S)ALS theory should enjoy some renewal especially in light of the now increasing literature in macroeconometrics and more generally in econometrics often leading to restrictions or estimating equations that are poorly handled by the common simulation-based methods.

Besides, it enables the exact and precise characterization of what is now abusively referred to as the "matching" in the approximation methods through the use of the estimating equations. Each existing methodology (SPML, SMM, II, EMM, SNLS) is thus characterized by particular "matching characteristics" or particular estimating equations.

The paper is organized as follows. We first recall in section 2 and prove under weaker stochastic equicontinuity conditions the available results from the II. In section 3, we develop the Generalized Indirect Inference (GII) seen as a particular and introductory illustration to the SALS theory. While analyzing the efficiency gains brought about by additional constraints (such as Euler conditions) on the instrumental criterion in the II framework, we show that Hansen (1982) theory of efficient overidentified moments estimation is no longer available here and characterize the new weighting matrix for performing a more efficient indirect estimation about the structural parameters of interest. In section 4, we develop the SALS theory and provide the general efficiency study in the SALS framework. We are thus led to introduce a new notion of **Efficiency Bounds in Direction**. Section 5 proposes a battery of generalized global specification tests extending the previous existing ones. We state some concluding remarks in section 6.

2 Asymptotic Properties of Indirect Inference

Extending Gouriéroux, Monfort and Renault (1993) and Gouriéroux and Monfort (1995b), we consider the parametric nonlinear simultaneous equations model defined by:

$$r(y_t, y_{t-1}, x_t, u_t, \theta) = 0, \tag{2.1}$$

$$\varphi(u_t, u_{t-1}, \varepsilon_t, \theta) = 0, \tag{2.2}$$

$$\theta \in \Theta \text{ a compact subset of } \mathbb{R}^p,$$

where the process $\{y_t, t \in \mathbb{Z}\}$ corresponds to the dependent variables and $\{x_t, t \in \mathbb{Z}\}$ is the vector of exogenous observable variables. The variables $\{u_t, t \in \mathbb{Z}\}$ and $\{\varepsilon_t, t \in \mathbb{Z}\}$ are not observed.[1]
$\{x_t, t \in \mathbb{Z}\}$ is independent of $\{\varepsilon_t, t \in \mathbb{Z}\}$ (and $\{u_t, t \in \mathbb{Z}\}$). The process $\{\varepsilon_t, t \in \mathbb{Z}\}$ is a white noise whose distribution $G_\circ$ is known.
The data consist in the observations of a stochastic process $\{(y_t, x_t), t \in \mathbb{Z}\}$ at dates $t = 1, \ldots, T$. The range of x_t and y_t are respectively $\mathcal{X} \subset \mathbb{R}^{p(x)}$ and $\mathcal{Y} \subset \mathbb{R}^{p(y)}$. We denote by $P_\circ$ the true unknown probability distribution (as characterized by Kolmogorov's theorem) of $\{(y_t, x_t), t \in \mathbb{Z}\}$ and $F_\circ$ denotes the p.d.f. of $\{x_t, t \in \mathbb{Z}\}$.
Assumption (A1):
$P_\circ$ belongs to the family $\{P_\theta, \theta \in \Theta\}$ of probability distributions on $(\mathcal{X} \times \mathcal{Y})^{\mathbb{Z}}$ delineated by the model $(2.1) - (2.2)$.
In this case there exists a true unknown value of the structural parameters θ° such that $P_\circ = P_{\theta^\circ}$, we assume that $\theta^\circ \in \overset{\circ}{\Theta}$. We also denote the probability distribution function of the joint process $\{(\varepsilon_t, x_t), t \in \mathbb{Z}\}$: $\pi_\circ$. With a slight abuse of notations, we will also write $\pi_\circ$ as the product of $F_\circ$ and $G_\circ$: $\pi_\circ = F_\circ \otimes G_\circ$. We can then rewrite $P_\circ$ and P_θ as follows:

$$P_\circ = \rho(\pi_\circ, \theta^\circ),$$

$$P_\theta = \rho(\pi_\circ, \theta),$$

for each $\theta \in \Theta$ and where $\rho(\cdot, \cdot)$ is implicitly defined through $(2.1) - (2.2)$.
As pointed out by Gouriéroux, Monfort and Renault (1993) the knowledge of the distribution of $\{\varepsilon_t, t \in \mathbb{Z}\}$ is not a real assumption, in the parametric case, since $\{\varepsilon_t, t \in \mathbb{Z}\}$ can always be considered as a function of a white noise with a known distribution function and of additional parameters which can be incorporated into θ. DGP with more than one lag in y, x, u can be included in this framework by increasing the dimension of the processes.
Note also, that it is not required that r and φ are known in a closed form. The only requirement is that, r and φ are computable at each point. This, for instance, can be achieved through optimization routine reflecting optimizing behavior.
This formulation encompasses the one proposed in Gouriéroux, Monfort and Renault (1993). This allows indeed the treatment of a broader class of models as for instance the models stemming from the stochastic

[1] Note that by construction of $(2.1) - (2.2)$ the variables $\{x_t, t \in \mathbb{Z}\}$ correspond here to strictly exogenous ones. However the case of weakly exogenous variables is not ruled out here and in that case one has to specify the conditional p.d.f. of u given X. Throughout the paper we just maintain the weak exogeneity assumption and when it is required we will stress the strong exogeneity assumption.

growth literature (see JBES January 1990, vol 8, $n^0 1$ for an in-depth discussion of such models).
However, this theory turns out to be relevant as soon as one has at his disposal a family of p.d.f. $\{P_\theta, \theta \in \Theta\}$ on $(\mathcal{X} \times \mathcal{Y})^{\mathbb{Z}}$ for which expectations of nonlinear functions are easily computed by simulation, by quadrature or by analytic expressions. In this respect the results proposed in this paper straightforwardly extend to the case of quadrature computation or analytic expressions. We just focus here on the simulation-based inference for sake of presentational convenience and consistency with the earlier papers.
For each given value of the parameters θ, it is possible to simulate a path $\{\tilde{y}_1(\theta, z_\circ), \ldots, \tilde{y}_T(\theta, z_\circ)\}$ conditionally on the observed path of the exogenous variables $\{x_1, \ldots, x_T\}$ and for given initial conditions $z_\circ = (y_\circ, u_\circ)$. This is done by simulating values $\{\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_T\}$ from $G_\circ$. Then by repeatedly solving equation (2.2) in the unknown variables $\tilde{u}_t(\theta, u_\circ)$:

$$\begin{cases} \varphi(\tilde{u}_t(\theta, u_\circ), \tilde{u}_{t-1}(\theta, u_\circ), \tilde{\varepsilon}_t, \theta) = 0, \quad t = 1, \ldots, T, \\ u_\circ, \end{cases}$$

we get $\tilde{u}_1(\theta, u_\circ), \ldots, \tilde{u}_T(\theta, u_\circ)$. Finally by solving equation (2.1) in the unknown variables $\tilde{y}_t(\theta, z_\circ)$:

$$\begin{cases} r(\tilde{y}_t(\theta, z_\circ), \tilde{y}_{t-1}(\theta, z_\circ), x_t, \tilde{u}_t(\theta, u_\circ), \theta) = 0, \quad t = 1, \ldots, T, \\ y_\circ, \end{cases}$$

we obtain a simulated path $\{\tilde{y}_1(\theta, z_\circ), \ldots, \tilde{y}_T(\theta, z_\circ)\}$. This implicitly assumes that, for each value of the parameters θ, for the observed exogenous variables $\{x_1, \ldots, x_T\}$ and for the initial conditions $z_\circ$, equations $(2.1) - (2.2)$ uniquely define the process $\{(y_t, u_t), t \in \mathbb{Z}\}$.
A direct estimation of the true unknown value of the structural parameters θ° is in practice impossible since the conditional p.d.f of $\{y_1, \ldots, y_T\}$ given $\{z_\circ, x_1, \ldots, x_T\}$ is computationally intractable. The idea is then to replace the intractable log-likelihood function of the structural model:

$$L_T(\theta) = \sum_{t=1}^{T} \log f_t(y_t / \underline{y}_{t-1}, \underline{x}_T, \theta), \tag{2.3}$$

by an instrumental criterion which involves a vector β of q instrumental parameters:

$$Q_T\left(\underline{y}_T, \underline{x}_T, \beta\right), \qquad \beta \in \mathcal{B} \text{ a compact set of } \mathbb{R}^q. \tag{2.4}$$

[2]

Assumption (A2):

(A2) [3]

1) $\forall \theta \in \Theta, \beta \in \mathcal{B}, \ \underset{T \to +\infty}{\pi_\circ \lim} \left| Q_T\left(\underline{\tilde{y}}_T^s(\theta, z_\circ^s), \underline{x}_T, \beta\right) - q(\theta, \beta) \right| = 0,$

2) $\forall \theta \in \Theta, \varepsilon > 0, \eta > 0, \ \exists \ \tilde{\Gamma}_T^s(\theta, \varepsilon, \eta), \tilde{\tau}^s(\theta, \varepsilon, \eta)$ *such that* $\forall T \geq \tilde{\tau}^s(\theta, \varepsilon, \eta)$:

- $\pi_\circ\left(\tilde{\Gamma}_T^s(\theta, \varepsilon, \eta) > \varepsilon\right) < \eta,$
- $\forall \beta \in \mathcal{B}, \ \exists \ \mathcal{O}_{\beta,\theta,\varepsilon,\eta}^s$ *an open set containing* β *with* :
$\underset{\overline{\beta} \in \mathcal{O}_{\beta,\theta,\varepsilon,\eta}^s}{Sup} \left| Q_T\left(\underline{\tilde{y}}_T^s(\theta, z_\circ^s), \underline{x}_T, \overline{\beta}\right) - Q_T\left(\underline{\tilde{y}}_T^s(\theta, z_\circ^s), \underline{x}_T, \beta\right) \right| \leq \tilde{\Gamma}_T^s(\theta, \varepsilon, \eta),$

for $s = 1, \ldots, S$ *and* $z_\circ^s$.

[2] We focus here on compact sets Θ and $\mathcal{B}$. However, other assumptions can be formulated so as to avoid such compactness hypotheses. See Andrews (1994) and Newey and McFadden (1994) for an in-depth discussion.
[3] We denote by $\underset{T \to +\infty}{\pi_\circ \lim}$ the limit in probability (with respect to $\pi_\circ$) when T goes to infinity.

$\{\tilde{y}_1^s(\theta, z_\circ^s), \ldots, \tilde{y}_T^s(\theta, z_\circ^s)\}$ correspond to simulated paths of the dependent variable according to the model $(2.1) - (2.2)$ conditionally on $\{x_1, \ldots, x_T\}$ and $z_\circ^s$, for $s = 1, \ldots, S$.

Assumption $(A2.2)$ expresses a stochastic equicontinuity property about the instrumental criterion $Q_T\left(\underline{\tilde{y}}_T^s(\theta, z_\circ^s), \underline{x}_T, \beta\right)$ computed for the simulated paths $\underline{\tilde{y}}_T^s(\theta, z_\circ^s)$.[4]

Proposition 2.1 : *Under assumptions $(A1) - (A2)$, we have:*

$$\pi_\circ \lim_{T\to+\infty} \mathop{Sup}_{\beta\in\mathcal{B}} \left| Q_T\left(\underline{y}_T, \underline{x}_T, \beta\right) - q\left(\theta^\circ, \beta\right) \right| = 0,$$

[5]

$$\pi_\circ \lim_{T\to+\infty} \mathop{Sup}_{\beta\in\mathcal{B}} \left| Q_T\left(\underline{\tilde{y}}_T^s(\theta, z_\circ^s), \underline{x}_T, \beta\right) - q\left(\theta, \beta\right) \right| = 0\cdot$$

Proof : *This result is obtained by simply applying Newey (1991) theorem 2.1. to the simulated paths $\{\tilde{y}_1^s(\theta, z_\circ^s), \ldots, \tilde{y}_T^s(\theta, z_\circ^s)\}$.*

Assumption (A3):

$q(\theta, \beta)$ is a non stochastic twice differentiable function not depending on the initial condition $z_\circ^s$ and with a unique minimum with respect to β for each value of $\theta \in \Theta$. Let $\beta^\circ = \tilde{\beta}(\theta^\circ)$ and $\tilde{\beta}(\theta)$ be respectively the minimum of $q(\theta^\circ, \beta)$ and $q(\theta, \beta)$, that is:

$$\begin{aligned} \beta^\circ = \tilde{\beta}(\theta^\circ) &= \mathop{Argmin}_{\beta\in\mathcal{B}} q(\theta^\circ, \beta), \\ \tilde{\beta}(\theta) &= \mathop{Argmin}_{\beta\in\mathcal{B}} q(\theta, \beta)\cdot \end{aligned}$$

We also assume that $\forall\theta \in \Theta,\ \tilde{\beta}(\theta) \in \mathring{\mathcal{B}}$.

Let us introduce the following estimators:

$$\begin{aligned} \hat{\beta}_T &= \mathop{Argmin}_{\beta\in\mathcal{B}} Q_T\left(\underline{y}_T, \underline{x}_T, \beta\right), \\ \tilde{\beta}_T^s(\theta) &= \mathop{Argmin}_{\beta\in\mathcal{B}} Q_T\left(\underline{\tilde{y}}_T^s(\theta, z_\circ^s), \underline{x}_T, \beta\right), \\ \tilde{\beta}_{TS}(\theta) &= \frac{1}{S}\sum_{s=1}^{S} \tilde{\beta}_T^s(\theta)\cdot \end{aligned} \tag{2.5}$$

[4] Note that the observed path $\underline{y}_T$ can always be regarded under correct specification as a simulated one for the value $\theta = \theta^\circ$. Thus, it is also implicitly assumed that the instrumental criterion obeys a stochastic equicontinuity property when computed on the observed path $\underline{y}_T$.

[5] Note that:

$$P_\circ \lim_{T\to+\infty} \mathop{Sup}_{\beta\in\mathcal{B}} \left| Q_T\left(\underline{y}_T, \underline{x}_T, \beta\right) - q\left(\theta^\circ, \beta\right) \right| = \pi_\circ \lim_{T\to+\infty} \mathop{Sup}_{\beta\in\mathcal{B}} \left| Q_T\left(\underline{y}_T, \underline{x}_T, \beta\right) - q\left(\theta^\circ, \beta\right) \right| = 0,$$

$$P_\theta \lim_{T\to+\infty} \mathop{Sup}_{\beta\in\mathcal{B}} \left| Q_T\left(\underline{\tilde{y}}_T^s(\theta, z_\circ^s), \underline{x}_T, \beta\right) - q\left(\theta, \beta\right) \right| = \pi_\circ \lim_{T\to+\infty} \mathop{Sup}_{\beta\in\mathcal{B}} \left| Q_T\left(\underline{\tilde{y}}_T^s(\theta, z_\circ^s), \underline{x}_T, \beta\right) - q\left(\theta, \beta\right) \right| = 0\cdot$$

Proposition 2.2 : *Under assumptions* $(A1)-(A3)$*, these estimators converge to:*

$$\pi_\circ \lim_{T\to+\infty} \widehat{\beta}_T = \widetilde{\beta}\left(\theta^\circ\right) = \beta^\circ,$$

$$\pi_\circ \lim_{T\to+\infty} \widetilde{\beta}_T^s(\theta) = \pi_\circ \lim_{T\to+\infty} \widetilde{\beta}_{TS}(\theta) = \widetilde{\beta}(\theta)\cdot$$

Proof : *See appendix A.1.*

Assumption (A4):

$$\forall\varepsilon>0,\eta>0,\exists\ \widetilde{\Delta}_T^s\left(\varepsilon,\eta\right),\widetilde{\tau}^s\left(\varepsilon,\eta\right)\ \textit{such that}\ \forall T\geq\widetilde{\tau}^s\left(\varepsilon,\eta\right):$$

- $\pi_\circ\left(\widetilde{\Delta}_T^s\left(\varepsilon,\eta\right)>\varepsilon\right)<\eta,$

$(A4)$

- $\exists\ \mathcal{N}_{\theta,\varepsilon,\eta}^s$ *an open set containing* θ *with* : $\underset{\overline{\theta}\in\mathcal{N}_{\theta,\varepsilon,\eta}^s}{Sup}\left\|\widetilde{\beta}_T^s(\overline{\theta})-\widetilde{\beta}_T^s(\theta)\right\|_q\leq\widetilde{\Delta}_T^s\left(\varepsilon,\eta\right),$

$for\ s=1,\ldots,S\ and\ z_\circ^s\cdot$

Assumption $(A4)$ expresses a stochastic equicontinuity property about $\widetilde{\beta}_T^s(\cdot)$.

Proposition 2.3 : *Under assumptions* $(A1)-(A4)$*, we have:*

$$\pi_\circ \lim_{T\to+\infty}\ \underset{\theta\in\Theta}{Sup}\left\|\widetilde{\beta}_T^s(\theta)-\widetilde{\beta}(\theta)\right\|_q=0\cdot$$

Proof : *This result is obtained by simply applying Newey (1991) theorem 2.1. to* $\widetilde{\beta}_T^s(\theta)$*.*

Assumption (A5): $\widetilde{\beta}(\cdot)$ is one-to-one.

The class of indirect estimators is thus indexed by a choice of a positive[6] weighting matrix Ω of size $q\times q$. For a given Ω, the indirect inference estimator is defined by:

$$\widehat{\theta}_{TS}(\Omega)=\underset{\theta\in\Theta}{Argmin}\left[\widehat{\beta}_T-\widetilde{\beta}_{TS}(\theta)\right]'\Omega\left[\widehat{\beta}_T-\widetilde{\beta}_{TS}(\theta)\right]\cdot \tag{2.6}$$

Under assumptions $(A1)-(A5)$, $\widehat{\theta}_{TS}(\Omega)$ is a consistent estimator of θ°. The same kind of proof as sketched for proposition 2.2 can be developed. We make in addition the following assumptions:

$(A6)$ $$\sqrt{T}\frac{\partial Q_T}{\partial\beta}\left(\underline{y}_T,\underline{x}_T,\beta^\circ\right),$$

[6]We will refer in all the paper to a positive matrix A of size $q\times q$ as a symmetric matrix such that $Sp\left(A\right)\in\left(\mathbb{R}_+^*\right)^q$. This, of course, implies that A is non singular. A is referred to as a non negative matrix if $Sp\left(A\right)\in\left(\mathbb{R}_+\right)^q$.

is asymptotically normally distributed with mean zero and with an asymptotic covariance matrix $I_\circ$ of full rank q.

(A7) $$J_\circ = \pi_\circ \lim_{T\to+\infty} \frac{\partial^2 Q_T}{\partial\beta\partial\beta'}\left(\underline{y}_T, \underline{x}_T, \beta^\circ\right) \textit{ is of full rank } q.$$

(A8) $$\lim_{T\to+\infty} \underset{\circ}{Cov}\left\{\sqrt{T}\frac{\partial Q_T}{\partial\beta}\left(\underline{\tilde{y}}_T^s(\theta^\circ, z_\circ^s), \underline{x}_T, \beta^\circ\right), \sqrt{T}\frac{\partial Q_T}{\partial\beta}\left(\underline{\tilde{y}}_T^\ell(\theta^\circ, z_\circ^\ell), \underline{x}_T, \beta^\circ\right)\right\} = K_\circ,$$

independent of the initial values $z_\circ^s$ and $z_\circ^\ell$, for $s \neq \ell$.

(A9) $$\pi_\circ \lim_{T\to+\infty} \frac{\partial\tilde{\beta}_T^s}{\partial\theta'}(\theta^\circ) = \frac{\partial\tilde{\beta}}{\partial\theta'}(\theta^\circ),$$
$$\frac{\partial\tilde{\beta}}{\partial\theta'}(\theta^\circ) \textit{ is of full column rank } p.$$

As usual, the indirect inference estimator $\widehat{\theta}_{TS}(\Omega)$ is computed while replacing Ω by a consistent estimator $\widehat{\Omega}_T$ of Ω but the asymptotic normal probability distribution of $\widehat{\theta}_{TS}(\Omega)$ will not depend on the choice of this estimator. In order to minimize the asymptotic covariance matrix of $\widehat{\theta}_{TS}(\Omega)$, an optimal choice of Ω as characterized by Gouriéroux, Monfort and Renault (1993) and Gouriéroux and Monfort (1995b) is:

$$\Omega^* = J_\circ {\Phi_\circ}^{-1} J_\circ, \tag{2.7}$$

where:

$$J_\circ = \pi_\circ \lim_{T\to+\infty} \frac{\partial^2 Q_T}{\partial\beta\partial\beta'}\left(\underline{y}_T, \underline{x}_T, \beta^\circ\right),$$
and:
$$\Phi_\circ = \underset{\circ}{V}ar_{as}\left\{\sqrt{T}\frac{\partial Q_T}{\partial\beta}\left(\underline{y}_T, \underline{x}_T, \beta^\circ\right) - \underset{\circ}{E}\left[\sqrt{T}\frac{\partial Q_T}{\partial\beta}\left(\underline{y}_T, \underline{x}_T, \beta^\circ\right)/\underline{x}_T\right]\right\} = I_\circ - K_\circ. \tag{2.8}$$

The corresponding asymptotic distribution of the efficient II estimator $\widehat{\theta}_{TS}(J_\circ {\Phi_\circ}^{-1} J_\circ) = \widehat{\theta}_{TS}^*$ is then:

$$\sqrt{T}\left(\widehat{\theta}_{TS}^* - \theta^\circ\right) \xrightarrow[T\to+\infty]{D} \mathcal{N}\left(O, W_S\right),$$
$$W_S = \left(1 + \frac{1}{S}\right)\left[\frac{\partial\tilde{\beta}'}{\partial\theta}(\theta^\circ) J_\circ {\Phi_\circ}^{-1} J_\circ \frac{\partial\tilde{\beta}}{\partial\theta'}(\theta^\circ)\right]^{-1}. \tag{2.9}$$

3 Generalized Indirect Inference

3.1 Indirect Inference and matching moments

The first point we want to stress here is that, in spite of Gallant and Tauchen (1996) paper title, a new theory is needed when one is interested in an instrumental criterion which matches general moment conditions:

Assumption (A10):

$$\underset{\circ}{E}\left[g(w_t,\beta^{\circ})\right]=0, \tag{3.1}$$

where $w_t=(y_t,y_{t-1},x_{t-1},\dots,y_{t-K},x_{t-K})$ (for a fixed number of K lags) and β° is the true unknown value of a vector $\beta\in\overset{\circ}{\mathcal{B}}$ a compact subset of $I\!R^q$ of instrumental parameters. To the best of our knowledge, the only case considered until now in the literature is the just-identified and separable one where the dimension of g (the number of moment conditions) is exactly equal to q and the instrumental model $Q_T\left(\underline{y}_T,\underline{x}_T,\beta\right)$ corresponds to:

$$Q_T\left(\underline{y}_T,\underline{x}_T,\beta\right)=\frac{1}{2}\left[\frac{1}{T}\sum_{t=1}^{T}g(w_t)-\beta\right]'\left[\frac{1}{T}\sum_{t=1}^{T}g(w_t)-\beta\right]^{7}.$$

Indeed, the original papers by Gouriéroux, Monfort and Renault (1993) and Gallant and Tauchen (1996) were only interested in M-estimation of the instrumental model in such a way that, first order conditions could be interpreted as just-identifying moment conditions. But we want to argue that:

- On the one hand, it turns out that in many circumstances, one wants to use an instrumental model (3.1) which is defined by overidentifying moment restrictions ($dim(g)=r>q$) (See subsection 3.2).
- On the other hand, the classical Hansen's (1982) theory of efficient overidentified GMM does no longer apply when one is interested in indirect efficient estimation of θ° and not in direct efficient estimation of β°. In light of this, we will argue that this result does differ from the ones obtained by Kodde, Palm and Pfann (1990) (See section 4: Simulated Asymptotic Least Squares, for more details).

With respect to these two arguments, we will develop in subsection 3.3 a general asymptotic theory for this setting, which provides answers to the two following related issues:

- First, what is the optimal weighting matrix Λ when one considers as an instrumental criterion for indirect inference:

$$Q_T\left(\underline{y}_T,\underline{x}_T,\beta\right)=\frac{1}{2}\left[\frac{1}{T}\sum_{t=1}^{T}g(w_t,\beta)\right]'\Lambda\left[\frac{1}{T}\sum_{t=1}^{T}g(w_t,\beta)\right],^{8} \tag{3.2}$$

that is an overidentifying GMM type criterion.

Of course the "optimality" of the weighting matrix refers to the asymptotic covariance matrix of the deduced indirect estimator of θ°. Moreover, it is clear that as usual, this covariance matrix will not be modified if Λ is replaced by a consistent estimator $\widehat{\Lambda}_T$.

- Second, how can we interpret this optimal choice as a selection of a just-identifying set of moment conditions? However we will argue in subsection 3.4 that this does not prevent us from being interested in a set of overidentifying instrumental moment conditions.

[7] The case of just-identifying moment conditions, whether separable or not, is extensively studied in subsection 3.4.

[8] For sake of computational convenience, we have rescaled the instrumental GMM criterion by the factor $\frac{1}{2}$.

Finally, it is worth noticing that the theory of such "a Generalized Indirect Inference" (GII hereafter) goes further in generalizing the Simulated Method of Moments (SMM) as proposed by McFadden (1989) for i.i.d. environments, and by Ingram and Lee (1991) and Duffie and Singleton (1993) for a time series environment.

While in Gouriéroux, Monfort and Renault (1993), it is shown that SMM with "separable" moment restrictions:

$$g(w_t, \beta) = h(w_t) - \beta, \tag{3.3}$$

is tantamount to Indirect Inference with an instrumental criterion:

$$Q_T\left(\underline{y}_T, \underline{x}_T, \beta\right) = \frac{1}{T}\sum_{t=1}^{T}\left\|h(w_t) - \beta\right\|_q^2, \tag{3.4}$$

where $\left\|\cdot\right\|_q$ is any norm on $I\!R^q$, GII opens the door to simulated moment restrictions which are overidentified with respect to β.

Indeed since the moment restrictions defined by (3.3) are just-identifying and separable, they can be interpreted as first order conditions corresponding to the M-estimator defined by the criterion (3.4). But, the general case of overidentifying instrumental moment restrictions, is not nested within the standard Indirect Inference, based on an instrumental M-estimator. However as Gouriéroux, Monfort and Renault (1993) we do not include within the GII framework the Simulated Method of Moments as developed in the semiparametric setting by Pakes and Pollard (1989) (See however section 4: Simulated Asymptotic Least Squares).

3.2 Examples

We propose in this subsection to motivate the GII approach through two examples based on Stochastic Volatility (SV hereafter) models estimation and on Asset Pricing models estimation introducing overidentifying moment restrictions as defined by (3.1).

3.2.1 Stochastic volatility models

Empirical financial studies have found strong evidence that the stock market returns present strong conditional heteroskedasticity, asymmetry, leptokurtosis patterns at the high frequency data level. In order to provide appropriate valuation of financial equities, it is essential to answer the question about the modelling of such patterns.

In this respect, the SV model has been introduced by Clark (1973), Tauchen and Pitts (1983), Taylor (1986-1994) among many other authors. These models appear as an alternative specification to the Autoregressive Conditionally Heteroskedastic (ARCH) model as introduced by Engle (1982) and Bollerslev (1986).

The SV models turn out to be more appealing for many reasons: broad general features of the data can be reproduced (persistent volatility, volatility clustering effect, leverage effect, asymmetries and leptokurtosis), less parameters have to be estimated, and SV models (3.5) are closed under temporal aggregation.

We focus in this subsection on SV models $\{y_t, t \in \mathbb{Z}\}$ defined by Meddahi and Renault (1997) as follows:

$$\begin{cases} y_t &= \sigma_{t-1}\varepsilon_t, \\ \sigma_t^2 &= \omega + \gamma\sigma_{t-1}^2 + \nu_t, \end{cases} \quad ^{9} \tag{3.5}$$

where we take for stationarity and positivity considerations on the volatility process the following assumptions: $0 < \gamma < 1$ and $0 < \omega$. The range of y_t is $\mathcal{Y} \subset \mathbb{R}$.
In order to complete the previous semiparametric specification (3.5), the innovation processes $\{\varepsilon_t, t \in \mathbb{Z}\}$ and $\{\nu_t, t \in \mathbb{Z}\}$ are assumed to share the following properties:

$$\begin{array}{lll} \underset{\circ}{E}\left[\varepsilon_t / I_{t-1}\right] = 0, & \underset{\circ}{E}\left[\varepsilon_t^2 / I_{t-1}\right] = 1, & \underset{\circ}{E}\left[\varepsilon_t^3 / I_{t-1}\right] = \mu_3^\circ, \\ \underset{\circ}{E}\left[\varepsilon_t^4 / I_{t-1}\right] = \mu_4^\circ, & \underset{\circ}{E}\left[\nu_t / I_{t-1}\right] = 0, & \underset{\circ}{E}\left[\nu_t^2 / I_{t-1}\right] = \eta^{\circ 2}, \ ^{10} \\ \underset{\circ}{E}\left[\varepsilon_t \nu_t / I_{t-1}\right] = \rho^\circ \eta^\circ, & \underset{\circ}{E}\left[\varepsilon_t^2 \nu_t / I_{t-1}\right] = 0, & \end{array} \tag{3.6}$$

where the information set $I_t = \sigma\left(\varepsilon_t, \varepsilon_\tau, \nu_\tau, \tau < t\right)$ is the σ-field generated by $(\varepsilon_t, \varepsilon_\tau, \nu_\tau, \tau < t)$.
Moreover, the empirical financial studies have laid the emphasis on the important asymmetric behavior of the stock market returns. Within the framework delineated by $(3.5)-(3.6)$, this stylized fact is explained by the skewness of the standardized innovation process $\{\varepsilon_t, t \in \mathbb{Z}\}$ ($\mu_3^\circ \neq 0$) and also by the so-called leverage effect ($\rho^\circ < 0$) (see Dridi and Renault (2000) for more details).
In order to estimate the previous SV model, one can use the following set of moment restrictions:

$$\begin{aligned} &\underset{\circ}{E}\left(y_t^2\right) = \frac{\omega^\circ}{1-\gamma^\circ}, \\ &\underset{\circ}{E}\left(y_t^2 y_{t-k}^2\right) = \frac{\omega^{\circ 2}}{\left(1-\gamma^\circ\right)^2} + \gamma^{\circ k}\frac{\eta^{\circ 2}}{1-\gamma^{\circ 2}}, \quad k \in \mathbb{N}^*, \\ &\underset{\circ}{E}\left(y_t^4\right) = \mu_4^\circ\left(\frac{\omega^{\circ 2}}{\left(1-\gamma^\circ\right)^2} + \frac{\eta^{\circ 2}}{1-\gamma^{\circ 2}}\right), \\ &\underset{\circ}{E}\left[y_t^2 - \omega^\circ - \gamma^\circ y_{t-1}^2 / I_{t-2}\right] = 0\cdot \end{aligned} \tag{3.7}$$

However, this set of moment restrictions enables only the identification and thus the estimation of the true unknown value $\left(\omega^\circ, \gamma^{\circ 2}, \eta^{\circ 2}, \mu_4^\circ\right)'$ of the parameters of interest ; and as already pointed out by Drost and Meddahi (1998), within the semiparametric SV specification $(3.5)-(3.6)$ and without further assumptions on the p.d.f. of the innovation process $\{\nu_t, t \in \mathbb{Z}\}$ one cannot identify the asymmetry parameters (ρ, μ_3).
In this context, Dridi and Renault (2000) suggest, first, to choose a specification for the p.d.f. of the joint process $\{(\varepsilon_t, \nu_t), t \in \mathbb{Z}\}$ and second to perform an indirect inference on the parameters of interest $\left(\omega^\circ, \gamma^{\circ 2}, \eta^{\circ 2}, \mu_4^\circ, \mu_3^\circ, \rho^\circ, \theta_2^{\circ\prime}\right)'$, where θ_2° correspond to the additional parameters introduced in the fully

[9] For sake of notational simplicity, we have written σ_t^2 rather than $\log\sigma_t^2$. However all the results extend if one considers $\log\sigma_t^2$ (to ensure the positivity of the conditional variance process).

[10] The symmetry assumption $\underset{\circ}{E}\left[\varepsilon_t^2\nu_t / I_{t-1}\right] = 0$ is made for sake of computational simplicity and can easily be fulfilled by setting:

$$\nu_t = \rho\eta\varepsilon_t + \xi_t,$$

where the process $\{\xi_t, t \in \mathbb{Z}\}$ is such that $\underset{\circ}{E}\left[\xi_t / I_{t-1}\right] = 0$ and $\{\xi_t, t \in \mathbb{Z}\} \perp\!\!\!\perp \{\varepsilon_t, t \in \mathbb{Z}\}$.

parametric SV model (deduced from the semiparametric SV model and the additional assumptions)[11]. We have, thus to specify an instrumental model based for example on an ARCH SNP expansion or alternatively on an ARCH(q_1) specification introducing instrumental parameters β_1. But here we also have at our disposal a set of moment restrictions as defined by (3.1):

$$\begin{aligned}
&\underset{\circ}{E}\left(g(w_t,\beta^\circ)\right)=0,\\
&g\left(\cdot,\cdot\right)=\left({g_1}'\left(\cdot,\cdot\right),{g_2}'\left(\cdot,\cdot\right)\right)',\\
&\beta=\left({\beta_1}',{\beta_2}'\right)',\quad \beta_2=\left(\beta_{21},\ldots,\beta_{24}\right)',\\
&g_2\left(w_t,\beta_2\right)=\begin{bmatrix}
y_t^2-\beta_{21}\\
y_t^2y_{t-1}^2-\beta_{21}^2-\beta_{23}\beta_{22}\\
\cdot\\
\cdot\\
\cdot\\
y_t^2y_{t-k}^2-\beta_{21}^2-\beta_{23}^k\beta_{22}\\
\cdot\\
\cdot\\
\cdot\\
y_t^2y_{t-K}^2-\beta_{21}^2-\beta_{23}^K\beta_{22}\\
y_t^4-\beta_{24}\left(\beta_{21}^2+\beta_{22}\right)\\
\left[y_t^2-\beta_{21}\left(1-\beta_{23}\right)-\beta_{23}y_{t-1}^2\right]\otimes Z_{t-2}
\end{bmatrix},
\end{aligned}\tag{3.8}$$

where K is a given integer and $Z_{t-2}\in I_{t-2}$ corresponds to any set of instrumental variables belonging to I_{t-2} and where we have also defined the one-to-one mapping:

$$\begin{aligned}
\mathbb{R}_+^*\times]0,1[\times\mathbb{R}_+^{*^2} \quad &\longrightarrow \quad \mathbb{R}_+^{*^2}\times]0,1[\times\mathbb{R}_+^*,\\
\left(\omega,\gamma,\mu_4,\eta^2\right)' \quad &\longrightarrow \quad \beta_2=\left(\beta_{21},\ldots,\beta_{24},\right)',\\
&\qquad \beta_{21}=\frac{\omega}{1-\gamma},\\
&\qquad \beta_{22}=\frac{\eta^2}{1-\gamma^2},\\
&\qquad \beta_{23}=\gamma,\\
&\qquad \beta_{24}=\mu_4.
\end{aligned}\tag{3.9}$$

[11]Note that Dridi and Renault (2000) have focused on **robust** indirect estimation of the parameters of interest $\left(\omega^\circ,{\gamma^\circ}^2,{\eta^\circ}^2,\mu_4^\circ,\mu_3^\circ,\rho^\circ\right)'$ in the presence of misspecification both in the asymmetry parameters and in the additional assumptions on the joint process $\{(\varepsilon_t,\nu_t)\,,t\in\mathbb{Z}\}$: $\overline{\theta}_2$ (say). Here we focus on more efficient indirect estimation in the context of correct specification.

3.2.2 Asset pricing models

There is nowadays a large literature focusing on equilibrium theory based asset pricing models. This dates back to the first work by Lucas (1978). The basic ingredients for all the various asset pricing models start with an agent who is a utility maximizer under her budget constraint and across time. The equilibrium path corresponding to such an economy leads to first order conditions or Euler equations, which can generally be expressed in terms of moment conditions:

$$\underset{\circ}{E}\left(g_2(w_t, \beta_2^\circ)\right) = 0, \tag{3.10}$$

where β_2° corresponds in general to taste or preference parameters[12]. The moment restrictions (3.10) provides also the pricing of any financial equity.

In the context of the estimation and implementation of a nonlinear equilibrium model of exchange rates and interest rates at the weekly frequency, Bansal, Gallant, Hussey and Tauchen (1995) suggests making a "complete specification of the model, including the **law of motion** of the latent driving processes, so that simulated realizations can be generated given candidate parameter settings. The criterion of fit involves comparing the time series properties of simulated versus observed realizations on exchange rates and interest rates".

The instrumental criterion is built upon a seminonparametric score generator while Marcet's parametrized expectations procedure is used for imposing the Euler constraints. They advocate this strategy rather than focusing directly on the Euler equations as done in the usual GMM settings, because reliable data on consumption, endowments and the money supply are not available at the weekly frequency.

Basically, we have on the one hand a fully parametric specified structural model from which it is possible to simulate path of the endogenous variables and on the other hand an instrumental model g introducing instrumental parameters β_1 through the sub-instrumental criterion g_1 which corresponds here to the SNP score generator and to which we have added the restrictions g_2 introducing in turns additional parameters β_2 as delineated by (3.10). So that again we can write:

$$\begin{aligned} &\underset{\circ}{E}\left(g(w_t, \beta^\circ)\right) = 0, \\ &g(\cdot,\cdot) = (g_1(\cdot,\cdot)', g_2(\cdot,\cdot)')' \cdot \end{aligned} \tag{3.11}$$

3.3 Asymptotic theory for Generalized Indirect Inference

We have to reconsider the asymptotic theory sketched in section 2 in the case where the instrumental criterion is of a GMM-type:

$$Q_T\left(\underline{y}_T, \underline{x}_T, \beta\right) = \frac{1}{2}\left[\frac{1}{T}\sum_{t=1}^{T} g(w_t, \beta)\right]' \widehat{\Lambda}_T \left[\frac{1}{T}\sum_{t=1}^{T} g(w_t, \beta)\right], \tag{3.12}$$

where $\widehat{\Lambda}_T$ is a positive $r \times r$ $(r > q)$ matrix converging in $\pi_\circ$-probability to a positive matrix Λ: $\pi_\circ \lim_{T\to+\infty} \widehat{\Lambda}_T = \Lambda$.

We maintain assumptions $(A1) - (A9)$ and naturally impose here that:

$$q(\theta, \beta) = \underset{\theta}{E}\left[g(\widetilde{w}_t(\theta), \beta)\right]' \Lambda \underset{\theta}{E}\left[g(\widetilde{w}_t(\theta), \beta)\right] \cdot$$

[12] See Kocherlakota (1996) for an extensive review of the state of the art.

Under assumptions $(A1)-(A10)$ we have:
$\Lambda^{\frac{1}{2}} \underset{\circ}{E}[g(w_t,\beta)] = 0 \Longrightarrow \beta = \beta^\circ$ and $\Lambda^{\frac{1}{2}} \underset{\circ}{E}\left[\frac{\partial g}{\partial \beta'}(w_t,\beta^\circ)\right]$ is of full column rank q.
We have now to consider matrices $\Phi_\circ(\Lambda)$ and $J_\circ(\Lambda)$ according to the general definitions (2.8):

$$\Phi_\circ(\Lambda) = \underset{\circ}{V}ar_{as}\left\{\sqrt{T}\frac{\partial Q_T}{\partial \beta}\left(\underline{y}_T, \underline{x}_T, \beta^\circ\right) - \underset{\circ}{E}\left[\sqrt{T}\frac{\partial Q_T}{\partial \beta}\left(\underline{y}_T, \underline{x}_T, \beta^\circ\right)/\underline{x}_T\right]\right\},$$

and:

$$J_\circ(\Lambda) = \pi_\circ \lim_{T\to+\infty} \frac{\partial^2 Q_T}{\partial \beta \partial \beta'}\left(\underline{y}_T, \underline{x}_T, \beta^\circ\right).$$

But, it is important to notice that the matrices $\Phi_\circ(\Lambda)$, $J_\circ(\Lambda)$ and the binding function $\widetilde{\beta}(\cdot,\cdot)$ generally depend on the weighting matrix Λ chosen for the instrumental GMM model (3.12). Of course, under correct specification, the true value β° will, not depend on Λ but the derivatives of the binding function which appear in the asymptotic covariance do depend on Λ. Without expliciting this dependence at this stage, we have under assumptions $(A1)-(A10)$ and $(A11)$:

Assumption (A11):

$$(A11) \qquad \pi_\circ \lim_{T\to+\infty} \frac{1}{T}\sum_{t=1}^{T} \frac{\partial g'}{\partial \beta}(w_t,\beta^\circ) = \underset{\circ}{E}\left[\frac{\partial g'}{\partial \beta}(w_t,\beta^\circ)\right],$$

that:

$$\Phi_\circ(\Lambda) = \underset{\circ}{E}\left[\frac{\partial g'}{\partial \beta}(w_t,\beta^\circ)\right]\Lambda \underset{\circ}{V}ar_{as}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t^*(w_t,\beta^\circ)\right]\Lambda \underset{\circ}{E}\left[\frac{\partial g}{\partial \beta'}(w_t,\beta^\circ)\right], \tag{3.13}$$

where:

$$g_t^*(w_t,\beta^\circ) = g(w_t,\beta^\circ) - \underset{\circ}{E}[g(w_t,\beta^\circ)/\underline{x}_T],$$

and:

$$J_\circ(\Lambda) = \underset{\circ}{E}\left[\frac{\partial g'}{\partial \beta}(w_t,\beta^\circ)\right]\Lambda \underset{\circ}{E}\left[\frac{\partial g}{\partial \beta'}(w_t,\beta^\circ)\right]. \tag{3.14}$$

The corresponding asymptotic covariance matrix of the efficient II estimator $\widehat{\theta}^*_{TS}(\Lambda)$ is then:

$$W_S(\Lambda) = \left(1+\frac{1}{S}\right)\left[\frac{\partial \widetilde{\beta}'}{\partial \theta}(\theta^\circ,\Lambda) J_\circ(\Lambda)\Phi_\circ(\Lambda)^{-1} J_\circ(\Lambda)\frac{\partial \widetilde{\beta}}{\partial \theta'}(\theta^\circ,\Lambda)\right]^{-1}. \tag{3.15}$$

As already noticed, the accuracy of $\widehat{\theta}^*_{TS}(\Lambda)$ depends through $\Phi_\circ(\Lambda)$, $J_\circ(\Lambda)$ and $\frac{\partial \widetilde{\beta}}{\partial \theta'}(\theta^\circ,\Lambda)$ on the initial choice Λ of the weighting matrix in the instrumental moment conditions (3.12). We call **GII estimator** any II efficient estimator $\widehat{\theta}^*_{TS}(\Lambda^*)$ associated with a weighting matrix Λ^* such that, for any Λ, $W_S(\Lambda) - W_S(\Lambda^*)$ is a non negative matrix. The main contribution of this subsection is to prove the existence of such optimal weighting matrices Λ^* and to characterize them. For such a characterization we have to explicit the dependence of $\frac{\partial \widetilde{\beta}}{\partial \theta'}(\theta^\circ,\Lambda)$ on the choice of Λ. Under assumptions $(A1)-(A3)$, the binding function is characterized by:

$$\widetilde{\beta}(\theta,\Lambda) = \pi_\circ \lim_{T\to+\infty} \widetilde{\beta}_{TS}(\theta,\Lambda),$$

$$\widetilde{\beta}(\theta,\Lambda) = \underset{\beta\in\mathcal{B}}{Argmin}\, z(\theta,\beta)'\Lambda z(\theta,\beta), \tag{3.16}$$

$$z(\theta,\beta) = \underset{\theta}{E}\left[g(\widetilde{w}_t(\theta),\beta)\right],$$

where the notation $z(\theta,\beta) = \underset{\theta}{E}\left[g(\widetilde{w}_t(\theta),\beta)\right]$ means that the value of the structural parameters governing $\widetilde{w}_t(\theta)$ is θ. Under assumption $(A3)$, $\widetilde{\beta}(\theta,\Lambda)$ is defined as the unique solution to the first order conditions:

$$\frac{\partial z'}{\partial \beta}\left(\theta,\widetilde{\beta}(\theta,\Lambda)\right)\Lambda z\left(\theta,\widetilde{\beta}(\theta,\Lambda)\right) = 0\cdot \tag{3.17}$$

[13]

By differentiating (3.17) with respect to θ and taking into account that:

$$z(\theta^\circ,\beta^\circ) = 0,$$

we get:

$$\frac{\partial z'}{\partial \beta}(\theta^\circ,\beta^\circ)\Lambda\left[\frac{\partial z}{\partial \theta'}(\theta^\circ,\beta^\circ) + \frac{\partial z}{\partial \beta'}(\theta^\circ,\beta^\circ)\frac{\partial \widetilde{\beta}}{\partial \theta'}(\theta^\circ,\Lambda)\right] = 0\cdot$$

We have then proved[14]:

Lemma 3.1 : *Under assumptions* $(A1)-(A5)$, $(A7)$, *and* $(A9)-(A11)$, *we have:*

$$\frac{\partial \widetilde{\beta}}{\partial \theta'}(\theta^\circ,\Lambda) = -J_\circ(\Lambda)^{-1}\frac{\partial z'}{\partial \beta}(\theta^\circ,\beta^\circ)\Lambda\frac{\partial z}{\partial \theta'}(\theta^\circ,\beta^\circ)\cdot$$

By replacing the result of lemma 3.1 into the general expression (3.15) of $W_S(\Lambda)$, we get the asymptotic covariance matrix of $\widehat{\theta}^*_{TS}(\Lambda)$ as:

$$W_S(\Lambda) = \left(1+\frac{1}{S}\right)\left[\frac{\partial z'}{\partial \theta}(\theta^\circ,\beta^\circ)\Lambda\frac{\partial z}{\partial \beta'}(\theta^\circ,\beta^\circ)\Phi_\circ(\Lambda)^{-1}\frac{\partial z'}{\partial \beta}(\theta^\circ,\beta^\circ)\Lambda\frac{\partial z}{\partial \theta'}(\theta^\circ,\beta^\circ)\right]^{-1}, \tag{3.18}$$

The key point is then to notice that $W_S(\Lambda)$ can be written:

$$W_S(\Lambda) = \left(1+\frac{1}{S}\right)\left[\frac{\partial z'}{\partial \theta}(\theta^\circ,\beta^\circ)V^{-\frac{1}{2}}X(\Lambda)\left(X(\Lambda)'X(\Lambda)\right)^{-1}X(\Lambda)'V^{-\frac{1}{2}}\frac{\partial z}{\partial \theta'}(\theta^\circ,\beta^\circ)\right]^{-1}, \tag{3.19}$$

where $V = \underset{\circ}{V}ar_{as}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t^*(w_t,\beta^\circ)\right]$ and $X(\Lambda)$ is a square root of $\Phi_\circ(\Lambda)$: $(\Phi_\circ(\Lambda) = X(\Lambda)'X(\Lambda))$:

$$X(\Lambda) = V^{\frac{1}{2}}\Lambda\frac{\partial z}{\partial \beta'}(\theta^\circ,\beta^\circ)\cdot$$

Therefore for any Λ, $W_S(\Lambda) - W_S^*$ is a non negative matrix where:

$$W_S^* = \left(1+\frac{1}{S}\right)\left[\frac{\partial z'}{\partial \theta}(\theta^\circ,\beta^\circ)V^{-1}\frac{\partial z}{\partial \theta'}(\theta^\circ,\beta^\circ)\right]^{-1}\cdot \tag{3.20}$$

Moreover, the lower bound W_S^* will be reached if and only if the weighting matrix Λ (and thus X itself) is chosen so that the columns of $V^{-\frac{1}{2}}\frac{\partial z}{\partial \theta'}(\theta^\circ,\beta^\circ)$ belong to the vectorial space $ImX(\Lambda)$ spanned by the columns of $X(\Lambda)$.

[13] Note that it is not assumed or required that $z\left(\theta,\widetilde{\beta}(\theta,\Lambda)\right) = 0$.

[14] Or equivalently by applying the implicit function theorem.

Proof :

$$\left(W_S^{*^{-1}} - W_S(\Lambda)^{-1}\right)\left(1+\frac{1}{S}\right) = \frac{\partial z'}{\partial \theta}(\theta^\circ,\beta^\circ)V^{-1}\frac{\partial z}{\partial \theta'}(\theta^\circ,\beta^\circ) -$$
$$\frac{\partial z'}{\partial \theta}(\theta^\circ,\beta^\circ)V^{-\frac{1}{2}}X(\Lambda)(X(\Lambda)'X(\Lambda))^{-1}X(\Lambda)'V^{-\frac{1}{2}}\frac{\partial z}{\partial \theta'}(\theta^\circ,\beta^\circ),$$

$$= \frac{\partial z'}{\partial \theta}(\theta^\circ,\beta^\circ)V^{-\frac{1}{2}}\left[I_r - X(\Lambda)(X(\Lambda)'X(\Lambda))^{-1}X(\Lambda)'\right]V^{-\frac{1}{2}}\frac{\partial z}{\partial \theta'}(\theta^\circ,\beta^\circ) >> 0,$$

$$W_S^* - W_S(\Lambda) = 0,$$

$$\iff \quad \frac{\partial z'}{\partial \theta}(\theta^\circ,\beta^\circ)V^{-\frac{1}{2}}\left[I_r - X(\Lambda)(X(\Lambda)'X(\Lambda))^{-1}X(\Lambda)'\right]V^{-\frac{1}{2}}\frac{\partial z}{\partial \theta'}(\theta^\circ,\beta^\circ) = 0,$$ [15]

$$\iff \forall x \in I\!R^p, \quad \left\|\left[I_r - X(\Lambda)(X(\Lambda)'X(\Lambda))^{-1}X(\Lambda)'\right]V^{-\frac{1}{2}}\frac{\partial z}{\partial \theta'}(\theta^\circ,\beta^\circ)x\right\|_2^2 = 0,$$

$$\iff \forall x \in I\!R^p, \quad \left[I_r - X(\Lambda)(X(\Lambda)'X(\Lambda))^{-1}X(\Lambda)'\right]V^{-\frac{1}{2}}\frac{\partial z}{\partial \theta'}(\theta^\circ,\beta^\circ)x = 0,$$

$$\iff \quad \left[I_r - X(\Lambda)(X(\Lambda)'X(\Lambda))^{-1}X(\Lambda)'\right]V^{-\frac{1}{2}}\frac{\partial z}{\partial \theta'}(\theta^\circ,\beta^\circ) = 0,$$

$$\iff \quad V^{-\frac{1}{2}}\frac{\partial z}{\partial \theta'}(\theta^\circ,\beta^\circ) \in ImX(\Lambda)\cdot$$

With a slight abuse of notations, we shall write this condition as:

$$V^{-1}\frac{\partial z}{\partial \theta'}(\theta^\circ,\beta^\circ) \in Im\Lambda\frac{\partial z}{\partial \beta'}(\theta^\circ,\beta^\circ)\cdot \tag{3.21}$$

Two main conclusions from (3.21) are worth noticing:

• First, as already announced, there is no reason why the optimal choice of Λ should be in any case $\left\{\underset{\circ}{Var}_{as}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g(w_t,\beta^\circ)\right]\right\}^{-1}$ (which in turns coincides with V^{-1} in the case without exogenous variables).

In other words, the **efficient GMM estimator of the auxiliary parameters β does not provide** in general an optimal way to perform indirect inference about θ.

• Such an optimal indirect inference, that we have called GII, corresponds to a choice $\Lambda = \Lambda^*$ solution to (3.21). We are now able to prove the existence of such a solution and in turns of a GII estimator.

Indeed:

* On the one hand, the columns of $V^{-1}\frac{\partial z}{\partial \theta'}(\theta^\circ,\beta^\circ)$ span a subspace E_p of $I\!R^r$ of dimension smaller than or equal to p.

* On the other hand, the columns of $\frac{\partial z}{\partial \beta'}(\theta^\circ,\beta^\circ)$ span a subspace F_q of $I\!R^r$ of dimension equal to $q \geq p$.

Therefore, there exists a (generally infinite) set of one-to-one linear operators on $I\!R^r$ which transform E_p in a subset of F_q. **Any matrix Λ representing the inverse of such an operator** is a convenient choice for GII. If we limit ourselves to regular matrices Λ, one way to represent the set of convenient choices of Λ is the set of restrictions:

$$M\Lambda^{-1}V^{-1}\frac{\partial z}{\partial \theta'}(\theta^\circ,\beta^\circ) = 0, \tag{3.22}$$

[15]We use the symbol $>>$ in the following sense. Let A and B be two square matrices of same size, $A >> B \iff A - B$ is a non negative matrix. $>>$ corresponds to the partial order over the symmetric matrices.

where $M = I_r - \frac{\partial z}{\partial \beta'}(\theta^\circ, \beta^\circ) \left[\frac{\partial z'}{\partial \beta}(\theta^\circ, \beta^\circ) \frac{\partial z}{\partial \beta'}(\theta^\circ, \beta^\circ) \right]^{-1} \frac{\partial z'}{\partial \beta}(\theta^\circ, \beta^\circ)$ is the matrix of the orthogonal projection on $\left[\frac{\partial z}{\partial \beta'}(\theta^\circ, \beta^\circ) \right]^\perp$. M is of rank $(r-q)$, which shows that (3.22) defines $(r-q)p$ independent linear restrictions on the coefficients of Λ^{-1}. We have then proved the main result of this section:

Theorem 3.1 : *Under assumptions $(A1) - (A11)$, a GII estimator $\widehat{\theta}^*_{TS}(\Lambda^*)$ is obtained for any choice of a positive matrix Λ^* of size $r \times r$ solution to:*

$$M \Lambda^{-1} V^{-1} \frac{\partial z}{\partial \theta'}(\theta^\circ, \beta^\circ) = 0,$$

$$z(\theta, \beta) = \underset{\theta}{E} \left[g(\widetilde{w}_t(\theta), \beta) \right],$$

$$V = \underset{\circ}{V}ar_{as} \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} g^*_t(w_t, \beta^\circ) \right],$$

and M is the matrix of the orthogonal projection on $\left[Im \underset{\circ}{E} \frac{\partial g}{\partial \beta'}(w_t, \beta^\circ) \right]^\perp$.
The asymptotic covariance matrix of any GII estimator is:

$$W_S(\Lambda^*) = \left(1 + \frac{1}{S} \right) \left[\frac{\partial z'}{\partial \theta}(\theta^\circ, \beta^\circ) V^{-1} \frac{\partial z}{\partial \theta'}(\theta^\circ, \beta^\circ) \right]^{-1}.$$

3.4 Generalized Indirect Inference and just identifying instrumental model

The main goal of this subsection is to lay out the link between the previous GII theory and the actual moments used for the so-called "matching". More precisely, we want to answer the question: To what extent does the GII method "optimally" use the implicit constraint $\underset{\circ}{E}\left(g(w_t, \beta^\circ)\right) = 0$ or equivalently $z(\theta^\circ, \beta^\circ) = 0$ on the structural parameters θ°?

We stress in the sequel that the GII theory can be reinterpreted in terms of just-identifying calibrated moments or in other words as an indirect estimation of the true unknown value of the structural parameters θ° through a just-identifying instrumental moment type criterion. However, even though this reinterpretation of the GII approach in terms of just-identifying instrumental moment type criterion is appealing for understanding which actual moments are matched, how potential efficiency gains concerning the indirect estimator of the structural parameters θ° are allowed through a larger explained variance of the true unknown conditional score by the regression on $g(w_t, \beta^\circ)$, we will argue in this subsection that it is in practice infeasible since one cannot precisely identify which subset of moments provides (just)identification. So that one is, in general, led to use overidentifying restrictions on the instrumental parameters β°. In this respect, the "just-identification trick" is mentioned here just for sake of interpretation and understanding of the GII approach.

We focus here on an overidentifying moment type instrumental criterion (3.1) and have thus at our disposal r moment conditions defining the instrumental parameters β°:

$$\begin{aligned} &\underset{\circ}{E}[g(w_t, \beta^\circ)] = 0, \\ &dim\ \beta^\circ = q \leq r\cdot \end{aligned} \tag{3.23}$$

We assume that it is possible to disentangle the set of moment restrictions $g(\cdot,\cdot)$ into a subset $g_1(\cdot,\cdot)$ (say) of just-identifying moment restrictions from those providing the overidentification of the instrumental criterion (3.23): $g_2(\cdot,\cdot)$ (say), and we write:

$$\begin{aligned} & g\left(w_t,\beta\right)=\left[\begin{array}{c} g_1\left(w_t,\beta\right) \\ g_2\left(w_t,\beta\right) \end{array}\right], \\ & dim\ g_1\left(\cdot,\cdot\right)=q,\ dim\ g_2\left(\cdot,\cdot\right)=r-q, \\ & \underset{\circ}{E}\left[g_1\left(w_t,\beta\right)\right]=0 \Longrightarrow \beta=\beta^{\circ}\cdot \end{aligned} \tag{3.24}$$

We already know that:

• In order to perform a GII of the structural parameters θ° through an overidentifying moment type criterion (3.23), one has to choose a weighting matrix Λ^* according to theorem 3.1. In this case, the optimal asymptotic covariance matrix of the optimal GII estimator $\widehat{\theta}_{TS}^*$ is:

$$\begin{aligned} & \underset{\circ}{V}ar_{as}\left[\sqrt{T}\left(\widehat{\theta}_{TS}^*-\theta^{\circ}\right)\right]=\left(1+\frac{1}{S}\right)\left[\frac{\partial z'}{\partial\theta}(\theta^{\circ},\beta^{\circ})V^{-1}\frac{\partial z}{\partial\theta'}(\theta^{\circ},\beta^{\circ})\right]^{-1}, \\ & z(\theta,\beta)=\underset{\theta}{E}\left[g(\widetilde{w}_t(\theta),\beta)\right], \\ & V=\underset{\circ}{V}ar_{as}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}g_t^*(w_t,\beta^{\circ})\right], \\ & g_t^*(w_t,\beta^{\circ})=g(w_t,\beta^{\circ})-\underset{\circ}{E}\left[g(w_t,\beta^{\circ})/\underline{x}_T\right]\cdot \end{aligned} \tag{3.25}$$

• When the instrumental moment type criterion is just-identified for the parameters β, that is $r=q$, any weighting positive matrix Λ is optimal since the estimation of the instrumental parameters, either performed under the observed paths or performed under the simulated ones, simply corresponds to solving the following system of q nonlinear equations in the q unknowns β:

$$\frac{1}{T}\sum_{t=1}^{T}g\left(\widetilde{w}_t(\theta),\beta\right)=0 \Longrightarrow \beta=\widetilde{\beta}_T(\theta)\cdot \tag{3.26}$$

In this case, we have for any choice of a positive weighting matrix Λ (for instance $\Lambda=I_q$) that the asymptotic covariance matrix of the GII estimator $\widehat{\theta}_{TS}(\Lambda^*)=\widehat{\theta}_{TS}^*$ is:

$$\begin{aligned} & \underset{\circ}{V}ar_{as}\left[\sqrt{T}\left(\widehat{\theta}_{TS}^*-\theta^{\circ}\right)\right]=\left(1+\frac{1}{S}\right)\left[\frac{\partial z'}{\partial\theta}(\theta^{\circ},\beta^{\circ})V^{-1}\frac{\partial z}{\partial\theta'}(\theta^{\circ},\beta^{\circ})\right]^{-1}, \\ & z(\theta,\beta)=\underset{\theta}{E}\left[g(\widetilde{w}_t(\theta),\beta)\right], \end{aligned} \tag{3.27}$$

and is thus minimal.

In light of the two previous remarks, we propose to replace the overidentifying instrumental model (3.23) by the following just-identifying one introducing a new parametrization $\alpha = (\alpha_1', \alpha_2')'$ defined as follows:

- $\alpha = (\alpha_1', \alpha_2')' \in \mathcal{B} \times \mathcal{A}_2$ *where* $\mathcal{A}_2 \subset I\!R^{r-q}$,

- $$Q_T^*\left(\underline{y}_T, \underline{x}_T, \alpha\right) = \frac{1}{2}\left[\frac{1}{T}\sum_{t=1}^{T} h(w_t,\alpha)\right]'\left[\frac{1}{T}\sum_{t=1}^{T} h(w_t,\alpha)\right], \tag{3.28}$$

- $h\left(w_t,\alpha\right) = \begin{bmatrix} h_1\left(w_t,\alpha_1\right) \\ h_2\left(w_t,\alpha\right)\end{bmatrix}$ *with* $\begin{cases} h_1\left(w_t,\alpha_1\right) = g_1\left(w_t,\alpha_1\right), \\ h_2\left(w_t,\alpha\right) = g_2\left(w_t,\alpha_1\right) - \alpha_2. \end{cases}$

Proposition 3.1 : *Under assumptions (A1) – (A4) and (A10), the instrumental criterion* (3.28)*: $Q_T^*\left(\underline{y}_T, \underline{x}_T, \alpha\right)$ is just-identified for the instrumental parameters α and associated with the true unknown value $\alpha^\circ = \left(\beta^{\circ\prime}, 0'\right)'$.*

Proof : *Since dim α = dim $h = r$, the instrumental model is at best just-identifying. We just have then to prove that the instrumental parameters α are actually identified.*

$$\underset{\circ}{E}\left[h(w_t,\alpha)\right] = 0,$$

$$\Longrightarrow \begin{cases} \underset{\circ}{E}\left[h_1(w_t,\alpha_1)\right] = 0, \\ \\ \underset{\circ}{E}\left[h_2(w_t,\alpha)\right] = 0, \end{cases}$$

$$\Longrightarrow \begin{cases} \underset{\circ}{E}\left[g_1(w_t,\alpha_1)\right] = 0, \\ \\ \underset{\circ}{E}\left[g_2(w_t,\alpha_1)\right] - \alpha_2 = 0, \end{cases}$$

$$\Longrightarrow \begin{cases} \alpha_1 = \beta^\circ, \\ \\ \alpha_2 = \underset{\circ}{E}\left[g_2(w_t,\beta^\circ)\right] = 0, \end{cases}$$

since $g_1\left(w_t,\cdot\right)$ identifies β° by assumption (3.24) *and the moment conditions associated with $g_2\left(\cdot,\cdot\right)$ are null at the true value β°.*

$$\Longrightarrow \quad \alpha^\circ = \left(\alpha_1^{\circ\prime}, \alpha_2^{\circ\prime}\right)' = \left(\beta^{\circ\prime}, 0'\right)'.$$

We define the indirect estimator $\widehat{\theta}^h_{TS}$ performed through the modified instrumental criterion (3.28), $Q^*_T\left(\underline{y}_T, \underline{x}_T, \alpha\right)$:

$$\widehat{\theta}^h_{TS} = \underset{\theta\in\Theta}{Argmin}\, [\widetilde{\alpha}_{TS}(\theta) - \widehat{\alpha}_T]'\, \Omega^*_h\, [\widetilde{\alpha}_{TS}(\theta) - \widehat{\alpha}_T]\,,$$

$$\widehat{\alpha}_T = \underset{\alpha\in\mathcal{B}\times\mathcal{A}_2}{Argmin}\, Q^*_T\left(\underline{y}_T, \underline{x}_T, \alpha\right),$$

$$\widetilde{\alpha}^s_T(\theta) = \underset{\alpha\in\mathcal{B}\times\mathcal{A}_2}{Argmin}\, Q^*_T\left(\underline{\widetilde{y}}^s_T(\theta, z^s_\circ), \underline{x}_T, \alpha\right), \tag{3.29}$$

$$\widetilde{\alpha}_{TS}(\theta) = \frac{1}{S}\sum_{s=1}^{S} \widetilde{\alpha}^s_T(\theta),$$

$$\Omega^*_h = J^{h^{-1}}_\circ \Phi_\circ{}^{-1} J^{h^{-1}}_\circ,$$

$J^h_\circ,\ \Phi_\circ$ are deduced from $(3.13) - (3.14)$.

In this context, we are now able to prove the main result of this subsection:

Proposition 3.2 : *Under assumptions $(A1) - (A11)$, the GII estimator $\widehat{\theta}^h_{TS}$ performed through the instrumental just-identifying moment criterion (3.28): $Q^*_T\left(\underline{y}_T, \underline{x}_T, \alpha\right)$ has the same asymptotic covariance matrix as the one produced by the optimal GII estimator $\widehat{\theta}^g_{TS}(\Lambda^*)$ performed through the instrumental overidentifying moment criterion (3.12): $Q_T\left(\underline{y}_T, \underline{x}_T, \beta\right)$. Thus we have:*

$$\underset{\circ\ \ as}{Var}\left[\sqrt{T}\left(\widehat{\theta}^h_{TS} - \theta^\circ\right)\right] = \left(1 + \frac{1}{S}\right)\left[\frac{\partial z'}{\partial\theta}(\theta^\circ, \beta^\circ) V^{-1} \frac{\partial z}{\partial\theta'}(\theta^\circ, \beta^\circ)\right]^{-1},$$

$$z(\theta, \beta) = \underset{\theta}{E}\, [g(\widetilde{w}_t(\theta), \beta)]\,,$$

$$V = \underset{\circ\ \ as}{Var}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g^*_t(w_t, \beta^\circ)\right],$$

$$g^*_t(w_t, \beta^\circ) = g(w_t, \beta^\circ) - \underset{\circ}{E}\, [g(w_t, \beta^\circ)/\underline{x}_T]\cdot$$

Proof : *We first index all the previously defined quantities either by g or by h depending on whether they are performed respectively through the instrumental model (3.23): $Q_T\left(\underline{y}_T, \underline{x}_T, \beta\right)$ or through the modified instrumental model (3.28): $Q^*_T\left(\underline{y}_T, \underline{x}_T, \alpha\right)$. We have then with obvious notations and according to the previous results that:*

$$\underset{\circ\ \ as}{Var}\left[\sqrt{T}\left(\widehat{\theta}^h_{TS} - \theta^\circ\right)\right] = \left(1 + \frac{1}{S}\right)\left[\frac{\partial z^{h'}}{\partial\theta}(\theta^\circ, \beta^\circ) V^{h^{-1}} \frac{\partial z^h}{\partial\theta'}(\theta^\circ, \beta^\circ)\right]^{-1},$$

$$\underset{\circ\ \ as}{Var}\left[\sqrt{T}\left(\widehat{\theta}^g_{TS}(\Lambda^*) - \theta^\circ\right)\right] = \left(1 + \frac{1}{S}\right)\left[\frac{\partial z^{g'}}{\partial\theta}(\theta^\circ, \beta^\circ) V^{g^{-1}} \frac{\partial z^g}{\partial\theta'}(\theta^\circ, \beta^\circ)\right]^{-1}.$$

We just now have to show the sufficient following conditions:

- $\bullet$ $V^g = V^h,$
- $\bullet\bullet$ $\dfrac{\partial z^h}{\partial\theta'}(\theta^\circ, \alpha^\circ) = \dfrac{\partial z^g}{\partial\theta'}(\theta^\circ, \beta^\circ)\cdot$

- $V^h = \underset{\circ}{Var}_{as}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} h_t^*(w_t,\alpha^\circ)\right]$,

  $\alpha^\circ = \left(\beta^{\circ\prime}, 0'\right)'$,

  $h_t^*(w_t,\alpha^\circ) = h_t(w_t,\alpha^\circ) - \underset{\circ}{E}\left[h_t(w_t,\alpha^\circ)/\underline{x}_T\right]\cdot$

Since $h(w_t,\alpha^\circ) = \begin{bmatrix} h_1(w_t,\alpha_1^\circ) \\ h_2(w_t,\alpha^\circ)\end{bmatrix} = \begin{bmatrix} g_1(w_t,\beta^\circ) \\ g_2(w_t,\beta^\circ)\end{bmatrix}$*, we have* $h(w_t,\alpha^\circ) = g(w_t,\beta^\circ)$*. Thus we have:* $V^h = V^g$*.*

•• *We start with the definition of* $z^h(\cdot,\cdot)$*:*

$$\forall\theta,\alpha\in\Theta\times(\mathcal{B}\times\mathcal{A}_2),\quad z^h(\theta,\alpha) = \underset{\theta}{E}\left[h(\widetilde{w}_t(\theta),\alpha)\right],$$

$$\Longrightarrow \forall\theta,\alpha\in\Theta\times(\mathcal{B}\times\mathcal{A}_2),\quad z^h(\theta,\alpha) = \underset{\theta}{E}\begin{bmatrix} g_1(\widetilde{w}_t(\theta),\alpha_1) \\ g_2(\widetilde{w}_t(\theta),\alpha_1)-\alpha_2\end{bmatrix},$$

$$\Longrightarrow \forall\theta,\alpha\in\Theta\times(\mathcal{B}\times\mathcal{A}_2),\quad z^h(\theta,\alpha) = \underset{\theta}{E}\begin{bmatrix} g_1(\widetilde{w}_t(\theta),\alpha_1) \\ g_2(\widetilde{w}_t(\theta),\alpha_1)\end{bmatrix} - \begin{bmatrix} 0 \\ \alpha_2\end{bmatrix},$$

$$\Longrightarrow \forall\theta,\alpha\in\Theta\times(\mathcal{B}\times\mathcal{A}_2),\quad z^h(\theta,\alpha) = z^g(\theta,\alpha_1) - (0',\alpha_2')',$$

$$\Longrightarrow \forall\theta,\alpha\in\Theta\times(\mathcal{B}\times\mathcal{A}_2),\quad \frac{\partial z^h}{\partial\theta'}(\theta,\alpha) = \frac{\partial z^g}{\partial\theta'}(\theta,\alpha_1),$$

$$\Longrightarrow \qquad \frac{\partial z^h}{\partial\theta'}(\theta^\circ,\alpha^\circ) = \frac{\partial z^g}{\partial\theta'}(\theta^\circ,\alpha_1^\circ) = \frac{\partial z^g}{\partial\theta'}(\theta^\circ,\beta^\circ)\cdot$$

Note that, although the two indirect estimators $\widehat{\theta}_{TS}^h$ and $\widehat{\theta}_{TS}^g$ have the same asymptotic covariance matrix, they do differ in general.

In light of proposition 3.2 results, one may argue that, on the one hand it seems that the implicit constraint $z(\theta^\circ,\beta^\circ)=0$ on the structural parameters θ° is not optimally used since the so-called "just-identification trick" leads to an indirect estimator $\widehat{\theta}_{TS}^h$ of minimum variance[16]. On the other hand, the use of the modified instrumental criterion $Q_T^*\left(\underline{y}_T,\underline{x}_T,\alpha\right)$ would avoid computing the optimal weighting matrix Λ^*.

However we want to mitigate the two previous criticisms with three respects:

• First, the use of the modified instrumental criterion $Q_T^*\left(\underline{y}_T,\underline{x}_T,\alpha\right)$ means asymptotically that one is, roughly speaking, minimizing the following distance criterion:

$$\underset{\theta\in\Theta}{min}\left\|\widetilde{\alpha}(\theta)-\alpha^\circ\right\|_\Omega^2 = \underset{\theta\in\Theta}{min}\left\|\begin{matrix}\widetilde{\alpha}_1(\theta)-\alpha_1^\circ \\ \\ \widetilde{\alpha}_2(\theta)-\alpha_2^\circ\end{matrix}\right\|_\Omega^2,$$

$$\alpha_1^\circ=\beta^\circ,\ \widetilde{\alpha}_1(\theta) = \underset{\alpha_1\in\mathcal{B}}{Argsol}\left\{\underset{\theta}{E}\left[g_1(\widetilde{w}_t(\theta),\alpha_1)\right]=0\right\},\ \alpha_2^\circ=0,\ \widetilde{\alpha}_2(\theta) = \underset{\theta}{E}\left[g_2(\widetilde{w}_t(\theta),\widetilde{\alpha}_1(\theta))\right]\cdot$$

[16] Of course, this optimality refers to the given class of instrumental criterion deduced from $\underset{\circ}{E}\left[g(w_t,\beta^\circ)\right]=0$.

$$\iff \min_{\theta\in\Theta} \left\| \begin{array}{c} \widetilde{\alpha}_1(\theta) - \beta^\circ \\ \\ \underset{\theta}{E}\left[g_2(\widetilde{w}_t(\theta), \widetilde{\alpha}_1(\theta))\right] \end{array} \right\|_\Omega^2, \quad \iff \min_{\theta\in\Theta} \left\| \begin{array}{c} \widetilde{\alpha}_1(\theta) - \beta^\circ \\ \\ \underset{\theta}{E}\left[g_2(\widetilde{w}_t(\theta), \beta^\circ)\right] \end{array} \right\|_\Omega^2,$$

The last equivalence shed some new lights on the GII approach. Indeed the GII estimator seeks to reproduce the dimensions along the instrumental parameters β° while simultaneously imposing the additional implicit constraint $\underset{\theta}{E}\left[g_2(\widetilde{w}_t(\theta), \beta^\circ)\right] = 0$.

In this respect, the GII approach can be regarded as a constrained estimation.

• Second, as already mentioned, in practice it is very unlikely that one is able to select a set of just-identifying moment restrictions. The procedure entails then overidentifying the instrumental criterion in order to ensure the identification of the instrumental parameters β. In this context and according to our results, one has to use the GII approach.

• Last but not least, we want to stress here that to the best of our knowledge, so far either in the II or in the EMM framework, the so-called implicit constraint $z(\theta^\circ, \beta^\circ) = 0$ on the structural parameters θ is never introduced nor treated in the instrumental criterion. The only treatment that can be found in this literature is the use of "quadrature-based" methods or numerical approximations for which one scarcely knows the statistical properties as well as the numerical performance.

This is why one of the messages of this section is that one should use these additional implicit constraints introducing in turn additional parameters in the instrumental criterion in order to minimize the asymptotic covariance matrix of the II estimator[17]. In order to illustrate this precise point, we show the following results (proposition 3.3). We first introduce a set of moment restrictions defining the parameters β_1° as follows:

$$\begin{aligned} &\underset{\circ}{E}\left[g_1(w_t, \beta_1^\circ)\right] = 0, \\ &dim\,(\beta_1^\circ) = q_1 \leq dim\,(g_1) = r_1\cdot \end{aligned} \tag{3.30}$$

This corresponds for instance to the case of SNP score generator ($q_1 = r_1$). Besides, we have at our disposal additional moment restrictions, introducing additional parameters β_2:

$$\begin{aligned} &\underset{\circ}{E}\left[g_2(w_t, \beta_1^\circ, \beta_2^\circ)\right] = 0, \\ &dim\,(\beta_2^\circ) = q_2,\ dim\,(g_2) = r_2\cdot \end{aligned} \tag{3.31}$$

We define the merged subset of moments associated with $g\,(\cdot,\cdot)$:

$$\begin{aligned} &\underset{\circ}{E}\left[g(w_t, \beta^\circ)\right] = 0, \\ &g(\cdot,\cdot) = (g_1(\cdot,\cdot)', g_2(\cdot,\cdot)')', \quad \beta = (\beta_1', \beta_2')'\cdot \end{aligned} \tag{3.32}$$

With a slight abuse of notations and for sake of presentational convenience, we will refer in the sequel to (3.30), (3.31) and (3.32) as the instrumental criteria.

Assumption (A12):

$$\begin{cases} \bullet\ \underset{\circ}{E}\left(g_1(w_t, \beta_1)\right) = 0 \implies \beta_1 = \beta_1^\circ, \\ \\ \bullet\ \underset{\circ}{E}\left(g(w_t, \beta)\right) = 0 \implies \beta = \beta^\circ\cdot \end{cases}$$

[17] This should be done although when using quadrature-based methods.

This identification assumption of the GMM criterion is however stronger than the usual global one:

$$\underset{\circ}{E}\left(g(w_t,\beta)\right)=0 \Longrightarrow \beta=\beta^{\circ}\cdot$$

We have focused on such an assumption because it does correspond to the set-up of the problem we are examining. Note also that assumption $(A12)$ implies that $r_1 \geq q_1$ and $r_2 \geq q_2$.
We define the GII estimators $\widehat{\theta}_{TS}^{g_1{}^*}$ and $\widehat{\theta}_{TS}^{g^*}$ respectively performed through the instrumental criteria (3.30) and (3.32). We are now able to prove the following results.

Proposition 3.3 : *Under assumptions $(A1)-(A12)$, the GII estimator $\widehat{\theta}_{TS}^{g^*}$ is always of smaller asymptotic covariance matrix than the one obtained for the GII estimator $\widehat{\theta}_{TS}^{g_1{}^*}$ and we have:*

$$\underset{\circ}{V}ar_{as}\left(\sqrt{T}\widehat{\theta}_{TS}^{g^*}\right)=\left\{\left[\underset{\circ}{V}ar_{as}\left(\sqrt{T}\widehat{\theta}_{TS}^{g_1^*}\right)\right]^{-1}+\left(1+\frac{1}{S}\right)^{-1}\left[\frac{\partial z_2}{\partial\theta'}(\theta^{\circ},\beta^{\circ})-V_{21}V_{11}{}^{-1}\frac{\partial z_1}{\partial\theta'}(\theta^{\circ},\beta_1^{\circ})\right]'\times\right.$$

$$\left.\left(V_{22}-V_{21}V_{11}{}^{-1}V_{21}{}'\right)^{-1}\left[\frac{\partial z_2}{\partial\theta'}(\theta^{\circ},\beta^{\circ})-V_{21}V_{11}{}^{-1}\frac{\partial z_1}{\partial\theta'}(\theta^{\circ},\beta_1^{\circ})\right]\right\}^{-1}, \tag{3.33}$$

where:

$$z(\theta,\beta)=\left(z_1{}',z_2{}'\right)'(\theta,\beta)=\underset{\theta}{E}\left[g_1'\left(\widetilde{w}_t(\theta),\beta_1\right),g_2'\left(\widetilde{w}_t(\theta),\beta_1,\beta_2\right)\right]',$$

$$V=\underset{\circ}{V}ar_{as}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}g_t^*(w_t,\beta^{\circ})\right],$$

$$g_t^*(w_t,\beta^{\circ})=g(w_t,\beta^{\circ})-\underset{\circ}{E}\left[g(w_t,\beta^{\circ})/\underline{x}_T\right], \tag{3.34}$$

$$V=\begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22}\end{bmatrix},\ dim\ V_{ij}=r_i\times r_j,\ \ i,j=1,2,\ \ V_{ij}{}'=V_{ji},\ \ i,j=1,2\cdot$$

Moreover the two asymptotic covariance matrices are equal if and only if:

$$\frac{\partial z_2}{\partial\theta'}\left(\theta^{\circ},\beta^{\circ}\right)=V_{21}V_{11}{}^{-1}\frac{\partial z_1}{\partial\theta'}\left(\theta^{\circ},\beta_1^{\circ}\right)\cdot \tag{3.35}$$

Proof : *See appendix A.2.*

In other words, as soon as (3.35) is not fulfilled, one should perform a GII estimation $\widehat{\theta}_{TS}^{g^*}$ on the criterion (3.32), whether using quadrature-based methods for imposing the implicit constraint on the structural parameters or not.

4 Simulated Asymptotic Least Squares

The main goal of this section is to develop a theory referred to as the Simulated Asymptotic Least Squares (SALS hereafter) corresponding both to a generalization and to a simulated version of the earlier Asymptotic Least Squares as proposed by Gouriéroux, Monfort and Trognon (1985).
We first lay out the links between the GII approach and the EMM methodology and show how this can be interpreted in terms of Minimum Distance estimation using a set of particular "estimating equations". This does correspond to the spirit of the ALS methodology and allows an interpretation of the GII approach. We then recall the main available results from the ALS procedure and extend it to the Simulated Asymptotic Least Squares.
Both Indirect Inference and Efficient Method of Moments as respectively proposed by Gouriéroux, Monfort and Renault (1993) and Gallant and Tauchen (1996) are nested within the SALS which can be regarded as the natural generalization of the EMM. Moreover the SALS encompasses the SMM approach as developed by McFadden (1989), Ingram and Lee (1991) and Duffie and Singleton (1993), but also the SMM as proposed by Pakes and Pollard (1989) in the semiparametric setting, the Simulated Pseudo Maximum Likelihood as developed by Laroque and Salanié (1989) and the SNLS as introduced by Laffont, Ossard and Vuong (1995) and the GII.

4.1 Generalized Indirect Inference and Efficient Method of Moments

In this subsection we reinterpret the GII approach in terms of the EMM methodology and thus characterize the exact "moment matching" which is performed through GII. We have at our disposal a set of a priori overidentifying moment restrictions:

$$\underset{\circ}{E}\left[g(w_t,\beta^\circ)\right]=0, \tag{4.1}$$

and we associate to (4.1) the natural extremum estimator $Q_T\left(\underline{y}_T,\underline{x}_T,\beta\right)$:

$$\begin{aligned}
&Q_T\left(\underline{y}_T,\underline{x}_T,\beta\right)=\frac{1}{2}\left[\frac{1}{T}\sum_{t=1}^{T}g(w_t,\beta)\right]'\widehat{\Lambda}_T\left[\frac{1}{T}\sum_{t=1}^{T}g(w_t,\beta)\right],\\
&\text{with } \underset{T\to+\infty}{\pi_\circ \lim}\,\widehat{\Lambda}_T=\Lambda, \qquad (4.2)\\
&\text{and } \widehat{\beta}_T=\underset{\beta\in\mathcal{B}}{Argmin}\,Q_T\left(\underline{y}_T,\underline{x}_T,\beta\right)\cdot
\end{aligned}$$

The GII approach leads to define the II estimator $\widehat{\theta}^*_{TS}=\widehat{\theta}_{TS}(\Lambda^*)$ (see theorem 3.1):

$$\begin{aligned}
&\underset{\circ}{V}ar_{as}\left[\sqrt{T}\left(\widehat{\theta}^*_{TS}-\theta^\circ\right)\right]=\left(1+\frac{1}{S}\right)\left[\frac{\partial z'}{\partial\theta}(\theta^\circ,\beta^\circ)V^{-1}\frac{\partial z}{\partial\theta'}(\theta^\circ,\beta^\circ)\right]^{-1},\\
&z(\theta,\beta)=\underset{\theta}{E}\left[g(\widetilde{w}_t(\theta),\beta)\right], \qquad (4.3)\\
&V=\underset{\circ}{V}ar_{as}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}g^*_t(w_t,\beta^\circ)\right]\cdot
\end{aligned}$$

A natural question is then: Can we build an equivalent indirect estimator in terms of the asymptotic covariance matrix through an EMM-type estimator?
In order to do so, we introduce the score associated with the instrumental criterion (4.2):

$$\frac{\partial Q_T}{\partial \beta}\left(\underline{y}_T, \underline{x}_T, \beta\right) = \left[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial g'}{\partial \beta}(w_t,\beta)\right]\widehat{\Lambda}_T\left[\frac{1}{T}\sum_{t=1}^{T} g(w_t,\beta)\right], \tag{4.4}$$

and we define the EMM-type estimator $\widehat{\widehat{\theta}}_{TS}(\Sigma,\Lambda)$:

$$\widehat{\widehat{\theta}}_{TS}(\Sigma,\Lambda) = \underset{\theta\in\Theta}{Argmin}\left[\frac{1}{S}\sum_{t=1}^{T}\frac{\partial Q_T}{\partial \beta'}(\widetilde{w}_t^s(\theta),\widehat{\beta}_T)\right]\Sigma^{18}\left[\frac{1}{S}\sum_{t=1}^{T}\frac{\partial Q_T}{\partial \beta}(\widetilde{w}_t^s(\theta),\widehat{\beta}_T)\right], \tag{4.5}$$

and where $\widehat{\beta}_T$ is defined by (4.2).

Proposition 4.1 : *Under assumptions $(A1)-(A11)$, the GII estimator $\widehat{\theta}_{TS}(\Lambda^*)$ and the EMM-type one $\widehat{\widehat{\theta}}_{TS}(\Sigma^\circ,\Lambda^*)$ where $\Sigma^\circ = \Phi_\circ(\Lambda^*)^{-1}$ are asymptotically equivalent.*

Proof : *The proof is obtained by applying Gouriéroux and Monfort (1995b) proposition 4.3 to the extremum criterion* (4.2).

Proposition 4.1 enables us to exactly characterize the moment matching. Indeed in light of this result, the GII estimator corresponds asymptotically to a Minimum Distance estimator minimizing the following asymptotic criterion:

$$\underset{\theta\in\Theta}{min}\left\|\frac{\partial z'}{\partial \beta}(\theta,\beta^\circ)\Lambda^* z(\theta,\beta^\circ)\right\|_{\Sigma}^{2}, \tag{4.6}$$

or equivalently $\widehat{\theta}_{TS}(\Lambda^*)$ should solve:

$$\frac{\partial z'}{\partial \beta}\left(\widehat{\theta}_{TS}(\Lambda^*),\beta^\circ\right)\Lambda^* z\left(\widehat{\theta}_{TS}(\Lambda^*),\beta^\circ\right) = 0\cdot \tag{4.7}$$

If we introduce the function $H^{\Lambda^*}(\theta,\beta)$ as follows:

$$H^{\Lambda^*}(\theta,\beta) = \frac{\partial z'}{\partial \beta}(\theta,\beta)\Lambda^* z(\theta,\beta), \tag{4.8}$$

we can define the estimating equations:

$$\begin{aligned} & H^{\Lambda^*}(\theta,\beta^\circ) = 0, \\ \Longleftrightarrow\quad & \frac{\partial z'}{\partial \beta}(\theta,\beta^\circ)\Lambda^* z(\theta,\beta^\circ) = 0\cdot \end{aligned} \quad [19] \tag{4.9}$$

In other words, the exact matching corresponds to (4.9) rather than to (4.1). This is the reason why one may argue that the GII estimator is not fully exploiting the constraint (4.1).[20]
However, the use of a Minimum Distance estimator $\widehat{\theta}$ deduced from estimating equations $H^{\Lambda^*}(\theta,\beta^\circ) = 0$ corresponds precisely to the estimation principle developed within the ALS theory. In this respect, the main purpose of the following subsection is to recall the results available for the latter methodology and to extend them to the simulation-based case.

[18] Note that the asymptotic properties of $\widehat{\widehat{\theta}}_{TS}(\Sigma,\Lambda)$ are not modified if we replace Σ by a consistent estimator.
[19] Note that the estimating equations do depend on Λ and this explains why our results differ from the ones obtained by Kodde, Palm and Pfann (1990) in the common ALS literature.
[20] See next subsection for alternative procedures.

4.2 Simulated Asymptotic Least Squares

We first start with the ALS as proposed by Gouriéroux, Monfort and Trognon (1985) and extend them to the SALS approach.

4.2.1 Asymptotic Least Squares theory

We first consider T vectors $y_1, \ldots, y_T$ whose size is n. These vectors can be seen as the first T terms of an infinite sequence $\{y_t, t \in \mathbb{Z}\}$ whose probability distribution $P_\circ$ belongs to some given family $\mathcal{P}$.
We assume that there exists a vector of parameters $\beta(P)$, called the auxiliary (or the instrumental) parameters, defined for any $P \in \mathcal{P}$ whose values belong to $\mathcal{B} \subset \mathbb{R}^q$ a compact set, and for which a consistent asymptotically normal estimator $\widehat{\beta}_T$ is available; therefore we have:
Assumption (A13):

$$\sqrt{T}\left(\widehat{\beta}_T - \beta^\circ\right) \xrightarrow[T\to+\infty]{D} \mathcal{N}\left(0, \Omega^\circ\right), \tag{4.10}$$

where $\beta^\circ = \beta(P_\circ)$ is the true unknown value of β, and Ω° is some symmetric positive matrix. Besides we have at our disposal a set of r "estimating equations" also referred to as the null hypothesis $H_\circ$:
Assumption (A14):

$$\begin{aligned}
&H\left(\theta^\circ, \beta^\circ\right) = 0, \\
&\theta^\circ \in \overset{\circ}{\Theta} \subset \mathbb{R}^p \text{ a compact subset, } \beta^\circ \in \overset{\circ}{\mathcal{B}}, \\
&\text{and we assume: } H\left(\theta, \beta^\circ\right) = 0 \Longrightarrow \theta = \theta^\circ, \\
&\text{and } H(\cdot,\cdot) \text{ is continuously differentiable on } \overset{\circ}{\Theta} \times \overset{\circ}{\mathcal{B}}\cdot
\end{aligned} \tag{4.11}$$

θ is referred to as the structural parameters or the parameters of interest. θ° is the true unknown value of these parameters of interest.
In this framework, one may consider two statistical problems:
- How to test the null hypothesis $H_\circ$.
- How to estimate θ under $H_\circ$[21].

The problem of the estimation of θ under $H_\circ$ has been treated by Gouriéroux, Monfort and Trognon (1985):

$$\widehat{\theta}_T = \underset{\theta\in\Theta}{Argmin}\, H\left(\theta, \widehat{\beta}_T\right)' \widehat{\Sigma}_T H\left(\theta, \widehat{\beta}_T\right), \tag{4.12}$$

where $\pi_\circ \lim\limits_{T\to+\infty} \widehat{\Sigma}_T = \Sigma$ is an $r \times r$ positive matrix[22].
Under assumptions $(A13) - (A15)$, it is shown that $\pi_\circ \lim\limits_{T\to+\infty} \widehat{\theta}_T = \theta^\circ$, $\widehat{\theta}_T$ is asymptotically normal and the

[21] In fact the third problem is the estimation of β° under $H_\circ$.
[22] For sake of computational consistency we have decided to refer to the $\pi_\circ$-probability without any loss of generality. Moreover the assumption of definiteness can be relaxed and in that case the identification assumption is : $\Sigma^{\frac{1}{2}} H\left(\theta, \beta^\circ\right) = 0 \Longrightarrow \theta = \theta^\circ$.

optimal choice Σ^* of Σ is:

$$\begin{aligned} \Sigma^* &= \left[\frac{\partial H}{\partial \beta'}(\theta^\circ, \beta^\circ)\,\Omega^\circ \frac{\partial H'}{\partial \beta}(\theta^\circ, \beta^\circ)\right]^{-1}, \\ \underset{\circ}{V}ar_{as}\left[\sqrt{T}\left(\widehat{\theta}_T^* - \theta^\circ\right)\right] &= \left[\frac{\partial H'}{\partial \theta}(\theta^\circ, \beta^\circ)\,\Sigma^* \frac{\partial H}{\partial \theta'}(\theta^\circ, \beta^\circ)\right]^{-1}. \end{aligned} \tag{4.13}$$

$\widehat{\theta}_T^* = \widehat{\theta}_T(\Sigma^*)$ corresponds to the optimal ALS estimators and the regularity conditions that are required are:

Assumptions (A15):

- $\frac{\partial H'}{\partial \theta}(\theta^\circ, \beta^\circ)$ is of full rank p,[23]
- $\frac{\partial H'}{\partial \beta}(\theta^\circ, \beta^\circ)$ is of rank r.

Note that since the sizes of $\frac{\partial H'}{\partial \theta}(\theta^\circ, \beta^\circ)$ and $\frac{\partial H'}{\partial \beta}(\theta^\circ, \beta^\circ)$ are respectively $p \times r$ and $q \times r$, assumption $(A15)$ implies that $p \leq r \leq q$.

In order to test the null hypothesis $H_\circ$, in Monfort and Rabemananjara (1990) it is shown that under $H_\circ$ and assumptions $(A13) - (A15)$ and additional regularity conditions the statistic ξ_T:

$$\xi_T = T \min_{\theta \in \Theta} H\left(\theta, \widehat{\beta}_T\right)' \widehat{\Sigma}_T H\left(\theta, \widehat{\beta}_T\right), \tag{4.14}$$

where $\pi_\circ \lim\limits_{T \to +\infty} \widehat{\Sigma}_T = \Sigma^*$ is asymptotically distributed as a chi-square with $(r-p)$ degrees of freedom and the critical region associated with the asymptotic level α is defined by:

$$W_\alpha = \left\{\xi_T > \chi^2_{1-\alpha}(r-p)\right\}. \tag{4.15}$$

4.2.2 Simulated Asymptotic Least Squares

The main goal of this subsubsection is to extend the previous ALS principles to a simulation-based inference referred to as the Simulated Asymptotic Least Squares (SALS). We want to stress here that the ALS approach or more precisely the extended SALS one should enjoy some renewal especially in light of the now increasing literature in macroeconometrics and more generally in econometrics often leading to restrictions or "estimating equations" of the type (4.11) which are however intractable or for which it is cumbersome to derive an analytical expression but it is still possible and easy to "approximate" them in some sense through the use of simulations (including Marcet parametrized type expectations procedures) or quadrature-based methods.

We first maintain assumptions $(A13) - (A15)$[24], however since in now most cases $H(\theta, \beta)$ is not known in a closed form (see subsection 3.2 for examples), we will assume that there exists a simulator $\widetilde{H}_{TS}(\theta, \beta)$ such that:

[23] In the case where Σ is not definite, we assume that $\Sigma^{\frac{1}{2}} \frac{\partial H}{\partial \theta'}(\theta^\circ, \beta^\circ)$ is of column rank p.

[24] See however the forthcoming remark concerning assumption $(A15)$.

Assumption $(\mathbf{A16})$:

$(A16)$ [25]

- $\forall \theta, \beta \in \Theta \times \mathcal{B},\ \left\| \widetilde{H}_{TS}(\theta,\beta) - H(\theta,\beta) \right\|_r \xrightarrow[\|\mu(T,S)\|_2^2 \to +\infty]{\pi_\circ} 0,$

- $\widetilde{H}_{TS}(\cdot,\cdot)$ is continuously differentiable (a.s.),

- $\forall \varepsilon > 0, \eta > 0,\ \exists\ \widetilde{\Delta}_T^S(\varepsilon,\eta), \widetilde{\tau}^S(\varepsilon,\eta)$ *such that* $\forall T \geq \widetilde{\tau}^S(\varepsilon,\eta)$:

  * $\pi_\circ\left(\widetilde{\Delta}_T^S(\varepsilon,\eta) > \varepsilon\right) < \eta,$

  * $\forall \theta, \beta \in \Theta \times \mathcal{B},\ \exists\ \mathcal{O}^S_{\theta,\beta,\varepsilon,\eta}$ *an open set containing* (θ,β) *with* : $\underset{\overline{\theta},\overline{\beta} \in \mathcal{O}^S_{\theta,\beta,\varepsilon,\eta}}{Sup} \left\| \widetilde{H}_{TS}\left(\overline{\theta},\overline{\beta}\right) - \widetilde{H}_{TS}(\theta,\beta) \right\|_r \leq \widetilde{\Delta}_T^S(\varepsilon,\eta)\cdot$

The subscript T refers to the T length data the econometrician has at her disposal and the subscript S refers indifferently to the number of replications in the case of simulations or to the number of grid points used in the case of quadrature-based methods. $\mu(\cdot,\cdot)$ is a non decreasing function in both arguments on $\mathbb{R}_+^2$. This notation allows us indeed to encompass the SNLS, Indirect Inference, EMM, GII, SMM methods for which S is fixed and T goes to infinity ; but also the SPML where both S and T goes to infinity at some proper rate. The only thing that really matters is the existence of a simulator as defined by $(A16)$ and which, as shown in proposition 4.2, converges uniformly with respect to θ, β in $\pi_\circ$-probability to $H(\theta,\beta)$.

Proposition 4.2 : *Under assumptions* $(A16)$ *and the compactness of* $\Theta \times \mathcal{B}$*, the simulator* $\widetilde{H}_{TS}(\theta,\beta)$ *converges uniformly with respect to* θ, β *in* $\pi_\circ$*-probability to* $H(\theta,\beta)$*, that is:*

$$\underset{\theta,\beta \in \Theta \times \mathcal{B}}{Sup} \left\| \widetilde{H}_{TS}(\theta,\beta) - H(\theta,\beta) \right\|_r \xrightarrow[\|\mu(T,S)\|_2^2 \to +\infty]{\pi_\circ} 0\cdot$$

Proof : *The proof of proposition* 4.2 *is obtained by simply applying Newey (1991) theorem 2.1. to the vectorial criterion* $\widetilde{H}_{TS}(\theta,\beta)$.

For most of the examples[26] given in this paper assumption $(A16)$ can be replaced by the stronger ones $(A1)$ and $(A17)$: there exists a function $\widetilde{H}_T\left(\underline{y}_T, \underline{x}_T, \beta\right)$ such that for $s = 1, \ldots, S$:

[25] As already pointed out by Pakes and Pollard (1989), assumption $(A16b)$ is restrictive especially when it has been reported that for several simulators the assumption of continuity itself is not fulfilled. However the same kind of technics and weaker assumptions on the simulator can be formulated so as to avoid such a drawback. What is then required is that the limit $H(\theta,\beta)$ is continuously differentiable. The extension of the SALS theory relaxing assumption $(A16b)$ is being processed.

[26] See proposition 5.2 for examples corresponding strictly to assumption $(A16)$.

Assumption (A17):

$(A17)$

- $\forall \theta, \beta \in \Theta \times \mathcal{B},\ \left\| \widetilde{H}_T\left(\underline{\widetilde{y}}_T^s(\theta, z_\circ^s), \underline{x}_T, \beta\right) - H(\theta, \beta) \right\|_r \xrightarrow[T \to +\infty]{\pi_\circ} 0,$
- $\widetilde{H}_T\left(\underline{\widetilde{y}}_T^s(\theta, z_\circ^s), \underline{x}_T, \beta\right)$ is continuously differentiable (a.s.),
- $\forall \varepsilon > 0, \eta > 0,\ \exists\ \widetilde{\Delta}_T^s(\varepsilon, \eta), \widetilde{\tau}^s(\varepsilon, \eta)$ *such that* $\forall T \geq \widetilde{\tau}^s(\varepsilon, \eta):$
  - $\pi_\circ\left(\widetilde{\Delta}_T^s(\varepsilon, \eta) > \varepsilon\right) < \eta,$
  - $\forall \theta, \beta \in \Theta \times \mathcal{B},\ \exists\ \mathcal{O}_{\theta,\beta,\varepsilon,\eta}^s$ *an open set containing* (θ, β) *with* :
$\underset{\overline{\theta},\overline{\beta} \in \mathcal{O}_{\theta,\beta,\varepsilon,\eta}^s}{Sup} \left\| \widetilde{H}_T\left(\underline{\widetilde{y}}_T^s(\overline{\theta}, z_\circ^s), \underline{x}_T, \overline{\beta}\right) - \widetilde{H}_T\left(\underline{\widetilde{y}}_T^s(\theta, z_\circ^s), \underline{x}_T, \beta\right) \right\|_r \leq \widetilde{\Delta}_T^s(\varepsilon, \eta) \cdot$

In this case we will have:

$$\forall s = 1, \ldots, S, \quad \underset{\theta,\beta \in \Theta \times \mathcal{B}}{Sup} \left\| \widetilde{H}_T\left(\underline{\widetilde{y}}_T^s(\theta, z_\circ^s), \underline{x}_T, \beta\right) - H(\theta, \beta) \right\|_r \xrightarrow[T \to +\infty]{\pi_\circ} 0,$$

for S fixed and $\underline{\widetilde{y}}_T^s(\theta, z_\circ^s)$ corresponds to a simulated path of the endogenous variables according to $(A1)$ conditionally on the observed path of the exogenous variables $\underline{x}_T$ and $z_\circ^s$ some initial conditions.
Note however that $(A16)$ corresponds to the semiparametric setting whereas $(A17)$ corresponds to the fully parametric one. For sake of simplicity, we will focus on assumptions $(A1)$ and $(A17)$ rather than on $(A16)$, however the results extend straightforwardly. We define the SALS estimator $\widehat{\theta}_{TS}(\Sigma)$ as follows:

$$\widehat{\theta}_{TS}(\Sigma) = \underset{\theta \in \Theta}{Argmin} \left[\frac{1}{S} \sum_{s=1}^{S} \widetilde{H}_T\left(\underline{\widetilde{y}}_T^s(\theta, z_\circ^s), \underline{x}_T, \widehat{\beta}_T\right)\right]' \widehat{\Sigma}_T \left[\frac{1}{S} \sum_{s=1}^{S} \widetilde{H}_T\left(\underline{\widetilde{y}}_T^s(\theta, z_\circ^s), \underline{x}_T, \widehat{\beta}_T\right)\right], \tag{4.16}$$

where $\widehat{\beta}_T$ is defined according to (4.10), $\pi_\circ \lim\limits_{T \to +\infty} \widehat{\Sigma}_T = \Sigma$ an $r \times r$ positive matrix.[27]
Note that the SALS estimator $\widehat{\theta}_{TS}(\Sigma)$ does not in general collapse to a SMM one since $H(\theta, \beta)$ can differ from a set of moment conditions for instance through the use of a general extremum instrumental criterion in the Indirect Inference framework. We are now able to derive the consistency of the SALS estimator.

Proposition 4.3 : *Under assumptions* $(A1)$, $(A13) - (A14)$ *and* $(A17)$, *the SALS estimator* $\widehat{\theta}_{TS}(\Sigma)$ *is consistent to* θ° *when* T *goes to infinity, that is:* $\pi_\circ \lim\limits_{T \to +\infty} \widehat{\theta}_{TS}(\Sigma) = \theta^\circ$.

Proof : *See appendix A.3.*

We assume in addition that:
Assumption (A18):

$$(A18) \qquad \sqrt{T} \widetilde{H}_T\left(\underline{y}_T, \underline{x}_T, \beta^\circ\right) \xrightarrow[T \to +\infty]{D} \mathcal{N}(0, I_\circ),$$

[27] Again, in the case where Σ is singular, we assume rather that $\Sigma^{\frac{1}{2}} H(\theta, \beta^\circ) = 0 \Longrightarrow \theta = \theta^\circ$.

where $I_\circ$ is of full rank r.

Assumption (A19):

$$
(A19) \quad
\begin{array}{l}
\bullet \quad \pi_\circ \lim\limits_{T\to+\infty} \dfrac{\partial}{\partial\theta} \left[\widetilde{H}'_T\left(\underline{\widetilde{y}}_T(\theta, z_\circ), \underline{x}_T, \beta^\circ\right)\right]_{\theta=\theta^\circ} = \dfrac{\partial H'}{\partial\theta}(\theta^\circ, \beta^\circ)\,, \\
\bullet \quad \pi_\circ \lim\limits_{T\to+\infty} \dfrac{\partial \widetilde{H}_T{}'}{\partial\beta}\left(\underline{y}_T, \underline{x}_T, \beta^\circ\right) = \dfrac{\partial H'}{\partial\beta}(\theta^\circ, \beta^\circ)\,.
\end{array}
$$

Assumption (A20):

$$
(A20) \quad
\begin{array}{l}
\bullet \quad \lim\limits_{T\to+\infty} \underset{\circ}{Cov}\left\{\sqrt{T}\widetilde{H}_T\left(\underline{\widetilde{y}}^s_T(\theta^\circ, z^s_\circ), \underline{x}_T, \beta^\circ\right), \sqrt{T}\widetilde{H}_T\left(\underline{\widetilde{y}}^\ell_T(\theta^\circ, z^\ell_\circ), \underline{x}_T, \beta^\circ\right)\right\} = K_\circ, \\
\bullet \quad \lim\limits_{T\to+\infty} \underset{\circ}{Cov}\left\{\sqrt{T}\widetilde{H}_T\left(\underline{\widetilde{y}}^s_T(\theta^\circ, z^s_\circ), \underline{x}_T, \beta^\circ\right), \dfrac{\partial H}{\partial\beta'}(\theta^\circ, \beta^\circ)\sqrt{T}\left(\widehat{\beta}_T - \beta^\circ\right)\right\} = L_\circ,
\end{array}
$$

independent of the initial values $z^s_\circ$ and $z^\ell_\circ$, for $s \neq \ell$ and $\underline{\widetilde{y}}^s_T(\theta^\circ, z^s_\circ)$ corresponds to a simulated path of the endogenous variables conditionally on the observed path of the exogenous variables and for the initial conditions $z^s_\circ$. Note that in the more general case $L_\circ \neq L_\circ{}'$ and $L_\circ \neq 0$ (see for instance the Indirect Inference where $L_\circ = L_\circ{}' \neq 0$).

We are now able to derive the asymptotic distribution of the SALS estimator $\widehat{\theta}_{TS}(\Sigma)$.

Proposition 4.4 : *Under assumptions (A1), (A13) − (A14) and (A17) − (A20), the SALS estimator $\widehat{\theta}_{TS}(\Sigma)$ is asymptotically normal and its asymptotic covariance matrix is given by:*

$$
\begin{aligned}
&\sqrt{T}\left(\widehat{\theta}_{TS}(\Sigma) - \theta^\circ\right) \xrightarrow[T\to+\infty]{D} \mathcal{N}\left(0, W_S(\Sigma)\right), \\
&W_S(\Sigma) = \left[\frac{\partial H'}{\partial\theta}(\theta^\circ, \beta^\circ)\,\Sigma\frac{\partial H}{\partial\theta'}(\theta^\circ, \beta^\circ)\right]^{-1} \frac{\partial H'}{\partial\theta}(\theta^\circ, \beta^\circ)\,\Sigma\widetilde{\Phi}_\circ(S)\Sigma\frac{\partial H}{\partial\theta'}(\theta^\circ, \beta^\circ) \times \\
&\left[\frac{\partial H'}{\partial\theta}(\theta^\circ, \beta^\circ)\,\Sigma\frac{\partial H}{\partial\theta'}(\theta^\circ, \beta^\circ)\right]^{-1}, \\
&\widetilde{\Phi}_\circ(S) = \frac{1}{S}\left(I_\circ - K_\circ\right) + K_\circ + L_\circ + L_\circ{}' + \frac{\partial H}{\partial\beta'}(\theta^\circ, \beta^\circ)\,\Omega^\circ\frac{\partial H'}{\partial\beta}(\theta^\circ, \beta^\circ)\,.
\end{aligned}
\tag{4.17}
$$

Proof : *See appendix A.4.1.*

As usual there is an optimal choice Σ^* of the weighting matrix Σ in order to derive the more accurate SALS estimator $\widehat{\theta}^*_{TS} = \widehat{\theta}_{TS}(\Sigma^*)$.

Proposition 4.5 : *Under assumptions (A1), (A13) − (A15) and (A17) − (A20), the optimal SALS estimator $\widehat{\theta}^*_{TS} = \widehat{\theta}_{TS}(\Sigma^*)$ is obtained when $\Sigma = \Sigma^* = \widetilde{\Phi}_\circ(S)^{-1}$ and its asymptotic covariance matrix W^*_S is given by:*

$$
W^*_S = W_S(\Sigma^*) = \left[\frac{\partial H'}{\partial\theta}(\theta^\circ, \beta^\circ)\,\widetilde{\Phi}_\circ(S)^{-1}\frac{\partial H}{\partial\theta'}(\theta^\circ, \beta^\circ)\right]^{-1}. \tag{4.18}
$$

Proof : *See appendix A.4.2.*

It is worth noticing that it is implicitly assumed that $\widetilde{\Phi}_\circ(S)$ is invertible rather than $\frac{\partial H}{\partial \beta'}(\theta^\circ,\beta^\circ)\,\Omega^\circ \frac{\partial H'}{\partial \beta}(\theta^\circ,\beta^\circ)$. In other words, we no longer require assumption $(A15b)$ namely $\frac{\partial H'}{\partial \beta}(\theta^\circ,\beta^\circ)$ is of full rank r. As a consequence, it is no longer required that $r \leq q$ and this corresponds in our opinion to a new useful result extending Gouriéroux, Monfort and Trognon (1985) for which $r \leq q$[28]. This latter requirement is in some sense very ad hoc since one cannot really understand (besides the mathematical restrictions) why the implicit constraint on the structural parameters θ°: $H(\theta^\circ,\beta^\circ) = 0$ should be such that the auxiliary parameter vector β° is of higher dimension than the restrictions $H(\cdot,\cdot)$ of which number is r.

Moreover, the case where $r > q$ is now widespread in the macroeconometric literature for instance it corresponds to the SALS estimator that can be deduced from Euler Equations of the type (3.1).

These are the reasons why we advocate the following strategy in order to circumvent the problem of non-invertibility of $\widetilde{\Phi}_\circ(S)$.

• First, in the **simulated** ALS framework and in the case where $\widetilde{\Phi}_\circ(S)$ is invertible, the above theory can be applied without having to perform any transformation. However in the case where $\widetilde{\Phi}_\circ(S)$ is not invertible, we suggest to modify the simulator $\widetilde{H}_{TS}(\theta,\beta)$ and to use a deduced simulator $\widetilde{\widetilde{H}}_{TS}(\theta,\beta,\varepsilon,\lambda)$:

$$\widetilde{\widetilde{H}}_{TS}(\theta,\beta,\varepsilon,\lambda) = \widetilde{H}_{TS}(\theta,\beta) + \lambda\widetilde{\varepsilon}_{TS}(\theta,\beta),$$

$$\text{such that } \widetilde{\varepsilon}_{TS}(\theta,\beta) \perp \widetilde{H}_{TS}(\theta,\beta), \frac{\partial H}{\partial \beta'}(\theta^\circ,\beta^\circ)\sqrt{T}\left(\widehat{\beta}_T - \beta^\circ\right), \quad ^{29} \tag{4.19}$$

$\widetilde{\widetilde{H}}_{TS}(\theta,\beta,\varepsilon,\lambda)$ satisfies assumption $(A16)$ and λ is any given scalar.

In this context, it is easy to see that the transformed asymptotic covariance $\widetilde{\widetilde{\Phi}}_\circ(S,\varepsilon,\lambda)$ associated with the deduced simulator $\widetilde{\widetilde{H}}_{TS}(\theta,\beta,\varepsilon,\lambda)$ (and thus with the deduced SALS) is given by:

$$\begin{aligned} \widetilde{\widetilde{\Phi}}_\circ(S,\varepsilon,\lambda) &= \widetilde{\Phi}_\circ(S) + \lambda^2\widetilde{\widetilde{\Psi}}_\circ(S,\varepsilon), \\ \widetilde{\widetilde{\Psi}}_\circ(S,\varepsilon) &= \underset{\circ}{V}\underset{as}{ar}\left[\sqrt{T}\widetilde{\varepsilon}_{TS}(\theta^\circ,\beta^\circ)\right], \end{aligned} \tag{4.20}$$

and takes the form $\frac{1}{S}\widetilde{\widetilde{\Psi}}^1_\circ(\varepsilon) + \left(1-\frac{1}{S}\right)\widetilde{\widetilde{\Psi}}^2_\circ(\varepsilon)$ under the more restrictive assumptions $(A1)$ and $(A17)$[30].
We now use the fact that the equation:

$$det\left[\widetilde{\widetilde{\Phi}}_\circ(S,\varepsilon,\lambda)\right] = det\left[\widetilde{\Phi}_\circ(S) + \lambda^2\widetilde{\widetilde{\Psi}}_\circ(S,\varepsilon)\right] = 0,$$

in $\lambda \in I\!R^+$ admits a maximum of r solutions (polynomial equation of maximum order r) or is equal to the polynom 0. In the case where $\widetilde{\Phi}_\circ(S)$ is singular, $\lambda = 0$ is one of those solutions. Moreover the polynom in λ^2 is not reduced to the polynom 0 otherwise it would imply that $Tr\left[\widetilde{\Phi}_\circ(S) + \lambda^2\widetilde{\widetilde{\Psi}}_\circ(S,\varepsilon)\right] = 0$[31], so that since $\widetilde{\Phi}_\circ(S)$ and $\widetilde{\widetilde{\Psi}}_\circ(S,\varepsilon)$ are non negative matrices, we would have $\widetilde{\Phi}_\circ(S) = \widetilde{\widetilde{\Psi}}_\circ(S,\varepsilon) = 0$ which is ruled out here. As a consequence, since there is a maximum of r solutions, it is possible to build a

[28]We are going to make this statement clearer since so far we still have the requirement that $\widetilde{\Phi}_\circ(S)$ is invertible. We show indeed that it is possible to modify the simulator such that this condition is fulfilled without any efficiency loss or more precisely with an efficiency loss that can be made as small as desired.

[29]This assumption is done for sake of simplicity but can be relaxed.

[30]See appendix A.4.1.

[31]Since it corresponds to the factor appearing for the monom of order $r-1$.

sequence $\{\lambda_n, n \in I\!N\}$ such that $\lim_{n\to+\infty} \lambda_n = 0$ and $\tilde{\tilde{\Phi}}_\circ(S,\varepsilon,\lambda_n)$ is non singular.

Thus we are led to the first case where $\tilde{\tilde{\Phi}}_\circ(S,\varepsilon,\lambda_n)$ is non singular and the SALS theory can be applied with a loss of efficiency that can be made as small as desired since $\lim_{n\to+\infty} \lambda_n = 0$. We make this statement clearer in the sequel (see theorems $4.1 - 4.2$).

The principle, which underlies this modified simulator is that we have introduced some extra randomness which can be made as small as desired $\lim_{n\to+\infty} \lambda_n = 0$ and such that the deduced SALS criterion or estimating equations lead to desirable or smooth properties such as the invertibility of $\tilde{\tilde{\Phi}}_\circ(S,\varepsilon,\lambda_n)$.[32]

• Second, we want to stress here that the aforementioned principle still holds when one is performing ALS estimation. Indeed, Gouriéroux, Monfort and Trognon (1985) has proposed an efficient ALS **only** in the case where $\frac{\partial H}{\partial \beta'}(\theta^\circ,\beta^\circ)$ is of full rank r (and thus $r \leq q$), we suggest here and although $H(\cdot,\cdot)$ is known in a closed form to use modified estimating equations $\tilde{\tilde{H}}(\cdot,\cdot)$ (that is an SALS estimator) introducing some extra randomness vanishing at the limit and such that the modified asymptotic covariance matrix $\tilde{\tilde{\Phi}}_\circ(S,\varepsilon,\lambda_n)$ is non singular and close to $\Phi_\circ = \frac{\partial H}{\partial \beta'}(\theta^\circ,\beta^\circ)\,\Omega^\circ \frac{\partial H'}{\partial \beta}(\theta^\circ,\beta^\circ)$.

We now state and prove one of the main mathematical result of this paper enabling a rigorous mathematical statement of the previous intuitive strategy.

Theorem 4.1 : *Let A $(r\times p)$, Σ $(r \times r)$, Φ $(r\times r)$ be three matrices such that $r \geq p$, Φ and Σ are symmetric non negative, we do not impose any invertibility assumption on either Φ or on Σ except that rank $\left(\Sigma^{\frac{1}{2}}A\right) = p$. Then the following program:*

$$\underset{\Sigma\in S_r(I\!R)}{Inf} \left\{ W(A,\Phi,\Sigma) = \left[A'\Sigma A\right]^{-1} A'\Sigma\Phi\Sigma A \left[A'\Sigma A\right]^{-1} \right\}, \tag{4.21}$$

where $S_r(I\!R)$ is the set of symmetric non negative matrices (and we have imposed rank $\left(\Sigma^{\frac{1}{2}}A\right) = p$) is such that:

I. In the case where Φ is singular, 4.21 *does not admit in general a unique minimal element. Thus* 4.21 *is written with an abuse of notations and has no unique solution in the "Zorn" sense.* **However**, *we have the following properties. For all Ψ symmetric, non negative and non null matrix:*

(i) there exists a symmetric non negative matrix $B^*(A,\Phi,\Psi)$ such that for all $\Sigma \in S_r(I\!R)$ and rank $\left(\Sigma^{\frac{1}{2}}A\right) = p$:

$$W(A,\Phi,\Sigma) >> B^*(A,\Phi,\Psi)\cdot$$

(ii) $\forall\, \varepsilon > 0$, $\exists\, \tilde{\Phi}(\Phi,\Psi,\varepsilon)$, $\tilde{\Sigma}(\Phi,\Psi,\varepsilon)$ two positive matrices such that:

$$* \quad \tilde{\Sigma}(\Phi,\Psi,\varepsilon) = \left[\tilde{\Phi}(\Phi,\Psi,\varepsilon)\right]^{-1},$$

$$* \quad \left\|\tilde{\Phi}(\Phi,\Psi,\varepsilon) - \Phi\right\|_r < \varepsilon,$$

$$* \quad \left\|B^*(A,\Phi,\Psi) - W\left(A, \tilde{\Phi}(\Phi,\Psi,\varepsilon), \tilde{\Sigma}(\Phi,\Psi,\varepsilon)\right)\right\|_r < \varepsilon,$$

where $\|\cdot\|_r$ is any norm on $\mathcal{M}_r(I\!R)$ the space of square matrices of size $r \times r$.

We will refer to $B^*(A,\Phi,\Psi)$ as the **lower efficiency bound in the direction of Ψ** *and associated with*

[32] Note that $\tilde{\tilde{\Phi}}_\circ(S,\varepsilon,\lambda_n)$ can be made as close as desired to $\tilde{\Phi}_\circ(S)$ but we have $\tilde{\tilde{\Phi}}_\circ(S,\varepsilon,\lambda_n) >> \tilde{\Phi}_\circ(S)$.

given (estimating equations) A, Φ. The word **direction** *has to be taken here* **stricto sensu**. *Indeed, we focus here on $\widetilde{\Phi}(\Phi,\Psi,\varepsilon)$ which are convex combinations of Φ and Ψ. However extensions of this theorem to the case where the direction is not the straight line (in the $\frac{r(r+1)}{2}$ dimension) are straightforward and the same kind of proofs are available. In those cases, one is implicitly specifying a particular way of reaching the target Φ: any curvilinear direction in the $\frac{r(r+1)}{2}$ dimension is a priori possible. We have decided here to start with the most natural one: the straight line.*
Of course, we have that for any Ψ_1 and Ψ_2 non negative (symmetric) matrices that $B^*(A,\Phi,\Psi_1)$ and $B^*(A,\Phi,\Psi_2)$ are either not comparable or equal. We cannot have indeed $B^*(A,\Phi,\Psi_{-i}) << B^*(A,\Phi,\Psi_i), i = 1$ or 2, and where $<<$ is here taken strictly. Moreover, if there exist $\lambda_1 > 0$ and $\lambda_2 > 0$ such that $\lambda_1\Psi_2 << \Psi_1 << \lambda_2\Psi_2$, then $B^*(A,\Phi,\Psi_1) = B^*(A,\Phi,\Psi_2)$.

II. In the case where Φ is invertible, *the infinimum is unique and reached as follows:*

$$\underset{\Sigma\in S_r(I\!R)}{Inf}\{W(A,\Phi,\Sigma)\} = W\left(A,\Phi,\Phi^{-1}\right) = \left[A'\Phi^{-1}A\right]^{-1}\cdot \tag{4.22}$$

In other words, theorem 4.1 extends the latter results to the case where Φ is non invertible.

The proof proceeds in three steps:
1) We first show that for any symmetric non negative and non null $r\times r$ matrix Ψ, there exists a decreasing sequence $\{\lambda_n, n\in I\!N\}$ with $\lim_{n\to+\infty}\lambda_n = 0$ and such that $\widetilde{\Phi}(\Phi,\Psi,\lambda_n) = \Phi+\lambda_n\Psi$ is invertible, decreasing in n with respect to the order $>>$ and has a limit when n goes to infinity that we denote $B^*(A,\Phi,\Psi) = \lim_{n\to+\infty} W\left(A,\widetilde{\Phi}(\Phi,\Psi,\lambda_n),\left[\widetilde{\Phi}(\Phi,\Psi,\lambda_n)\right]^{-1}\right)$.
2) Second we show that, $\forall\ \Sigma\in S_r(I\!R)$ $\left(rank\left(\Sigma^{\frac{1}{2}}A\right)=p\right)$, $\forall n\in I\!N$:

$$W\left(A,\widetilde{\Phi}(\Phi,\Psi,\lambda_n),\Sigma\right) >> W\left(A,\widetilde{\Phi}(\Phi,\Psi,\lambda_n),\left[\widetilde{\Phi}(\Phi,\Psi,\lambda_n)\right]^{-1}\right)\cdot$$

3) Third, we prove that:

$$\lim_{n\to+\infty} W\left(A,\widetilde{\Phi}(\Phi,\Psi,\lambda_n),\Sigma\right) = W(A,\Phi,\Sigma),$$

and since we know that $B^*(A,\Phi,\Psi) = \lim_{n\to+\infty} W\left(A,\widetilde{\Phi}(\Phi,\Psi,\lambda_n),\left[\widetilde{\Phi}(\Phi,\Psi,\lambda_n)\right]^{-1}\right)$, we conclude by analyzing the spectra that, $\forall\ \Sigma\in S_r(I\!R)$ and $rank\left(\Sigma^{\frac{1}{2}}A\right) = p$, $\forall\ \Psi$ symmetric, non null and non negative $r\times r$ matrix:

$$W(A,\Phi,\Sigma) >> B^*(A,\Phi,\Psi),$$

$$\lim_{n\to+\infty} W\left(A,\widetilde{\Phi}(\Phi,\Psi,\lambda_n),\left[\widetilde{\Phi}(\Phi,\Psi,\lambda_n)\right]^{-1}\right) = B^*(A,\Phi,\Psi)\cdot$$ [33]

Of course, in the case where $r=p$ and thus A is invertible, we have:

$$\forall\Psi\in S_r(I\!R),\quad B^*(A,\Phi,\Psi) = B^*(A,\Phi) = A^{-1}\Phi A'^{-1}\cdot$$

Proof : *1) Let Ψ be any non null non negative matrix of size $r\times r$. We define $\widetilde{\Phi}(\Phi,\Psi,\lambda) = \Phi+\lambda\Psi$ and we restrict $\lambda\geq 0$ without any loss of generality. Taking into consideration that the equation in $\lambda\geq 0$:*

[33] The proof of II is provided in appendix A.4.2.

det $\left(\widetilde{\Phi}(\Phi,\Psi,\lambda)\right) = det\,(\Phi+\lambda\Psi) = 0$ *has a maximum of* r *solutions and that the polynom det* $(\Phi+\lambda\Psi)$ *is not reduced to* 0 *(since* $Tr\left(\widetilde{\Phi}(\Phi,\Psi,\lambda)\right) = Tr\,(\Phi+\lambda\Psi) = Tr\,(\Phi) + \lambda\,Tr\,(\Psi) \geq \lambda\,Tr\,(\Psi) > 0$ *for* $\lambda > 0$*), we have that there exists a sequence* $\{\lambda_n, n \in I\!N\}$ *such that* $\lim_{n\to+\infty} \lambda_n = 0$, $\{\lambda_n, n \in I\!N\}$ *is a decreasing sequence and* $\forall\ n \in I\!N$, $\widetilde{\Phi}(\Phi,\Psi,\lambda_n)$ *is non singular. Indeed take any sequence* $\{\mu_n, n \in I\!N\}$ *such that* $\lim_{n\to+\infty} \mu_n = 0$ *and* $\{\mu_n, n \in I\!N\}$ *is a decreasing sequence, then take a suitable subsequence* $\lambda_n = \mu_{\varphi(n)}$ *such that* $\forall\ n \in I\!N,\ \lambda_n \neq \lambda_i^*\ i = 1,\ldots,r$, *where* λ_i^* *are the solutions to the equation det* $(\Phi+\lambda\Psi) = 0$*. In so doing, you will only have to avoid a maximum of* r *points. Moreover:*

- $\widetilde{\Phi}(\Phi,\Psi,\lambda_{n+1}) - \widetilde{\Phi}(\Phi,\Psi,\lambda_n) = (\lambda_{n+1}-\lambda_n)\,\Psi << 0, n \in I\!N$,
  and where the inequality is a strict one.

- $W\left(A, \widetilde{\Phi}(\Phi,\Psi,\lambda_n), \widetilde{\Phi}(\Phi,\Psi,\lambda_n)^{-1}\right) = \left[A'\widetilde{\Phi}(\Phi,\Psi,\lambda_n)^{-1}A\right]^{-1}\cdot$

Since $\widetilde{\Phi}(\Phi,\Psi,\lambda_{n+1}) << \widetilde{\Phi}(\Phi,\Psi,\lambda_n)$ *(with a strict inequality), we have:*

$$W\left(A, \widetilde{\Phi}(\Phi,\Psi,\lambda_{n+1}), \widetilde{\Phi}(\Phi,\Psi,\lambda_{n+1})^{-1}\right) << W\left(A, \widetilde{\Phi}(\Phi,\Psi,\lambda_n), \widetilde{\Phi}(\Phi,\Psi,\lambda_n)^{-1}\right), n \in I\!N\cdot$$

We thus define, applying lemma 4.1[34] *to the ordered chain* $\left\{W\left(A, \widetilde{\Phi}(\Phi,\Psi,\lambda_n), \widetilde{\Phi}(\Phi,\Psi,\lambda_n)^{-1}\right), n \in I\!N\right\}$*, the minimal element* $B^*(A,\Phi,\Psi)$*:*

$$\begin{aligned} B^*(A,\Phi,\Psi) &= \operatorname*{Inf}_{n\in I\!N}\left\{W\left(A, \widetilde{\Phi}(\Phi,\Psi,\lambda_n), \widetilde{\Phi}(\Phi,\Psi,\lambda_n)^{-1}\right)\right\}, \\ &= \lim_{n\to+\infty} W\left(A, \widetilde{\Phi}(\Phi,\Psi,\lambda_n), \widetilde{\Phi}(\Phi,\Psi,\lambda_n)^{-1}\right)\cdot \end{aligned}$$

2) The proof of $\forall\ \Sigma \in S_r(I\!R)$ $\left(rank\left(\Sigma^{\frac{1}{2}}A\right) = p\right)$, $\forall n \in I\!N$*:*

$$W\left(A, \widetilde{\Phi}(\Phi,\Psi,\lambda_n), \Sigma\right) >> W\left(A, \widetilde{\Phi}(\Phi,\Psi,\lambda_n), \left[\widetilde{\Phi}(\Phi,\Psi,\lambda_n)\right]^{-1}\right),$$

corresponds to the application of the result proved in appendix A.4.2. and where:

$$\begin{array}{lll} \widetilde{\Phi}_\circ(S) & \longrightarrow & \widetilde{\Phi}(\Phi,\Psi,\lambda_n), \quad n \in I\!N, \\ \dfrac{\partial H'}{\partial\theta}(\theta^\circ,\beta^\circ) & \longrightarrow & A', \\ \Sigma & \longrightarrow & \Sigma, \end{array}$$

and the arrow means "is replaced by".

3)

$$\begin{aligned} &\left\|W\left(A, \widetilde{\Phi}(\Phi,\Psi,\lambda_n), \Sigma\right) - W(A,\Phi,\Sigma)\right\|_r = \\ &\left\|[A'\Sigma A]^{-1}A'\Sigma\left[\widetilde{\Phi}(\Phi,\Psi,\lambda_n) - \Phi\right]\Sigma A\,[A'\Sigma A]^{-1}\right\|_r = \left\|[A'\Sigma A]^{-1}A'\Sigma\Psi\Sigma A\,[A'\Sigma A]^{-1}\right\|_r \lambda_n, \end{aligned}$$

[34]This lemma states that any decreasing sequence of non negative symmetric matrices with respect to the partial order $<<$ has a unique limit which is a symmetric non negative matrix. See appendix *A.4.3.* for the proofs.

$$\Longrightarrow \quad \begin{aligned} &\bullet \lim_{n\to+\infty}\left\|W\left(A,\widetilde{\Phi}\left(\Phi,\Psi,\lambda_n\right),\Sigma\right)-W\left(A,\Phi,\Sigma\right)\right\|_r=0,\\ &\bullet \lim_{n\to+\infty}\left\|W\left(A,\widetilde{\Phi}\left(\Phi,\Psi,\lambda_n\right),\widetilde{\Phi}\left(\Phi,\Psi,\lambda_n\right)^{-1}\right)-B^*\left(A,\Phi,\Psi\right)\right\|_r=0,\\ &\bullet \ \forall\Sigma\in S_r\left(I\!R\right),\ rank\left(\Sigma^{\frac{1}{2}}A\right)=p,\\ &\quad W\left(A,\widetilde{\Phi}\left(\Phi,\Psi,\lambda_n\right),\Sigma\right)>>W\left(A,\widetilde{\Phi}\left(\Phi,\Psi,\lambda_n\right),\widetilde{\Phi}\left(\Phi,\Psi,\lambda_n\right)^{-1}\right).\end{aligned}$$

We deduce that $W\left(A,\widetilde{\Phi}\left(\Phi,\Psi,\lambda_n\right),\Sigma\right)-W\left(A,\widetilde{\Phi}\left(\Phi,\Psi,\lambda_n\right),\widetilde{\Phi}\left(\Phi,\Psi,\lambda_n\right)^{-1}\right)$ *is a non negative matrix for all* $n\in I\!N$ *and converging to* $W\left(A,\Phi,\Sigma\right)-B^*\left(A,\Phi,\Psi\right)$ *when* n *goes to infinity. We define* $\tau\left(A,\Phi,\Sigma,\Psi\right)=W\left(A,\Phi,\Sigma\right)-B^*\left(A,\Phi,\Psi\right)$ *and* $\tau_n\left(A,\Phi,\Sigma,\Psi\right)=W\left(A,\widetilde{\Phi}\left(\Phi,\Psi,\lambda_n\right),\Sigma\right)-W\left(A,\widetilde{\Phi}\left(\Phi,\Psi,\lambda_n\right),\widetilde{\Phi}\left(\Phi,\Psi,\lambda_n\right)^{-1}\right)$.
$\tau_n\left(\cdot,\cdot,\cdot,\cdot\right)$ *is a symmetric, non negative matrix for all* $n\in I\!N$, $\tau\left(\cdot,\cdot,\cdot,\cdot\right)$ *is a symmetric matrix and we also have that* $\lim_{n\to+\infty}\tau_n\left(A,\Phi,\Sigma,\Psi\right)=\tau\left(A,\Phi,\Sigma,\Psi\right)$. *Therefore* $\lim_{n\to+\infty}Sp\left[\tau_n\left(A,\Phi,\Sigma,\Psi\right)\right]=Sp\left[\tau\left(A,\Phi,\Sigma,\Psi\right)\right]$. *We know that* $Sp\left[\tau_n\left(A,\Phi,\Sigma,\Psi\right)\right]\in\left(I\!R_+\right)^r/\left\{0,\ldots,0\right\}\Longrightarrow Sp\left[\tau\left(A,\Phi,\Sigma,\Psi\right)\right]\in\left(I\!R_+\right)^r$ *; or in other words that* $\tau\left(A,\Phi,\Sigma,\Psi\right)>>0$.
The proofs of the efficiency bounds properties, namely that they do not depend on the chosen sequence $\{\lambda_n,n\in I\!N\}$ *and that if there exist* $\lambda_1>0$, $\lambda_2>0$ *such that* $\lambda_1\Psi_2<<\Psi_1<<\lambda_2\Psi_2\Longrightarrow B^*\left(A,\Phi,\Psi_1\right)=B^*\left(A,\Phi,\Psi_2\right)$, *are given in appendix A.4.4.*

This corresponds to the results announced in theorem 4.1. The principal difficulties come here from the fact, that we have an order $>>$ over the square symmetric matrices set, which is not total, therefore and as usual the Zorn lemma does not ensure the existence of the infinimum (4.21), that the passage at the infinity limit is thus no longer straightforward and that Φ is singular. We have decided to take the point Φ as the heliocenter and then build ordered chains which here correspond to the radiuses (Φ,Ψ). This circumvents the problem of unordered space with respect to $>>$. Then along those radiuses, we have defined the efficiency bounds in the direction of Ψ: $B^*\left(A,\Phi,\Psi\right)$. It is important to realize that a priori $B^*\left(A,\Phi,\Psi\right)$ **does depend** on the “walk” from Ψ to Φ (linear, curvilinear,...)
Moreover, we also want to stress that:
• On the one hand, the proof sketched above straightforwardly extend when rather than focusing on $\widetilde{\Phi}\left(\Phi,\Psi,\lambda_n\right)$ of the form $\Phi+\lambda_n\Psi$, one uses general forms $\widetilde{\Phi}_n\left(\Phi\right)$ such that $\lim_{n\to+\infty}\widetilde{\Phi}_n\left(\Phi\right)=\Phi$ and $\left\{\widetilde{\Phi}_n\left(\Phi\right),n\in I\!N\right\}$ is non singular. In this case, what is required is that the chain $\left\{\widetilde{\Phi}_n\left(\Phi\right),\ n\in I\!N\right\}$ is ordered with respect to $>>$.
• On the other hand, one may wrongly think that theorem 4.1 is solely due to the particular functional form of $W\left(A,\Phi,\Sigma\right)$. However, it turns out that this functional form plays a subordinate role in the proof and what really matters is that: $W\left(A,\cdot,\Sigma\right)$ is continuous with respect to Φ and $W\left(A,\Phi,\Phi^{-1}\right)$ respects the order $>>$ whenever Φ is non singular: $\Phi_1>>\Phi_2\Longrightarrow W\left(A,\Phi_1,{\Phi_1}^{-1}\right)>>W\left(A,\Phi_2,{\Phi_2}^{-1}\right)$.
In light of the two previous remarks, we are now able to state the following generalization of theorem 4.1.

Theorem 4.2 : *The following program:*

$$\underset{\Sigma\in S_r(I\!R)}{Inf}\left\{W\left(\Phi,\Sigma\right)\right\}, \tag{4.23}$$

where $S_r(\mathbb{R})$ is the set of symmetric non negative matrices and $W(\Phi,\Sigma)$ is a non negative matrix such that:

- $\forall\Phi\in S_r^*(\mathbb{R})\,,\ \exists\ B(\Phi)\ /\ \forall\Sigma\in S_r(\mathbb{R})\,,\ W(\Phi,\Sigma)>>B(\Phi)=W\left(\Phi,\xi\left(\Phi^{-1}\right)\right),$ *where ξ is some known function, $\xi:\ S_r^*(\mathbb{R})\longrightarrow S_r(\mathbb{R})\,,\ S_r^*(\mathbb{R})$ is the set of symmetric, positive matrices,*

- $W(\cdot,\Sigma)$ *is continuous in Φ for all $\Sigma\in S_r(\mathbb{R})$,*

- $W\left(\Phi,\xi\left(\Phi^{-1}\right)\right)$ *is compatible with the order $>>$, that is: $\Phi_1>>\Phi_2\Longrightarrow$ $W\left(\Phi_1,\xi\left({\Phi_1}^{-1}\right)\right)>>W\left(\Phi_2,\xi\left({\Phi_2}^{-1}\right)\right),$*

is such that:

I. In the case where Φ is singular, 4.23 *does not admit in general a unique minimal element.* **However**, *we have the following properties:*

(i) For all ordered and decreasing chain $\left\{\widetilde{\Phi}_n,n\in\mathbb{N}\right\}$ (with respect to the partial order $>>$) and such that $\lim\limits_{n\to+\infty}\widetilde{\Phi}_n=\Phi$, there exists a symmetric non negative matrix $\widetilde{B}^*(\Phi)$[35] such that for all $\Sigma\in S_r(\mathbb{R})$:

$$W(\Phi,\Sigma)>>\widetilde{B}^*(\Phi)\cdot$$

(ii) $\forall\ \varepsilon>0,\ \exists\ n_\circ,\widetilde{\Sigma}_n$ a positive matrix such that for all $n\geq n_\circ$:

$$*\quad \widetilde{\Sigma}_n=\widetilde{\Phi}_n^{-1},$$

$$*\quad \left\|\widetilde{\Phi}_n-\Phi\right\|_r<\varepsilon,$$

$$*\quad \left\|\widetilde{B}^*(\Phi)-W\left(\widetilde{\Phi}_n,\xi\left(\widetilde{\Phi}_n^{-1}\right)\right)\right\|_r<\varepsilon,$$

where $\|\cdot\|_r$ is any norm on $\mathcal{M}_r(\mathbb{R})$ the space of square matrices of size $r\times r$.

II. In the case where Φ is invertible, *the infinimum is unique and reached as follows:*

$$\underset{\Sigma\in S_r(\mathbb{R})}{Inf}\{W(\Phi,\Sigma)\}=W\left(\Phi,\xi\left(\Phi^{-1}\right)\right)=B(\Phi)\cdot \tag{4.24}$$

In other words, theorem 4.2 extends the latter results to the case where Φ is non invertible.

Proof :

1) Since $\left\{\widetilde{\Phi}_n,n\in\mathbb{N}\right\}$ is a decreasing ordered chain, we have $\widetilde{\Phi}_{n+1}<<\widetilde{\Phi}_n,n\in\mathbb{N}$. Moreover $W\left(\Phi,\xi\left(\Phi^{-1}\right)\right)$ is compatible with the order. This implies that $\left\{W\left(\widetilde{\Phi}_n,\xi\left(\widetilde{\Phi}_n^{-1}\right)\right),n\in\mathbb{N}\right\}$ is a decreasing ordered chain. By lemma 4.1., there exists therefore a non negative matrix $\widetilde{B}^*(\Phi)$ such that:

$$\begin{aligned}\widetilde{B}^*(\Phi)&=\underset{n\in\mathbb{N}}{Inf}\left\{W\left(\widetilde{\Phi}_n,\xi\left(\widetilde{\Phi}_n^{-1}\right)\right)\right\},\\&=\lim_{n\to+\infty}W\left(\widetilde{\Phi}_n,\xi\left(\widetilde{\Phi}_n^{-1}\right)\right)\cdot\end{aligned}$$

2) $\forall\Sigma\in S_r(\mathbb{R})\,,\ W\left(\widetilde{\Phi}_n,\Sigma\right)>>W\left(\widetilde{\Phi}_n,\xi\left(\widetilde{\Phi}_n^{-1}\right)\right)\cdot$

3) $\forall\Sigma\in S_r(\mathbb{R})\,,\ W(\cdot,\Sigma)$ is continuous $\Longrightarrow\lim\limits_{n\to+\infty}W\left(\widetilde{\Phi}_n,\Sigma\right)=W(\Phi,\Sigma)$.

[35] Therefore a priori depending on the sequence itself or in other words on the way one reaches the target Φ.

To summarize, we have:

- $\lim_{n\to+\infty}\left\|W\left(\widetilde{\Phi}_n,\Sigma\right)-W\left(\Phi,\Sigma\right)\right\|_r=0,$

- $\lim_{n\to+\infty}\left\|W\left(\widetilde{\Phi}_n,\xi\left(\widetilde{\Phi}_n^{-1}\right)\right)-\widetilde{B}^*\left(\Phi\right)\right\|_r=0,$

- $\forall\Sigma\in S_r(I\!R)\,,\ W\left(\widetilde{\Phi}_n,\Sigma\right)>>W\left(\widetilde{\Phi}_n,\xi\left(\widetilde{\Phi}_n^{-1}\right)\right)\cdot$

We define $\tau(\Phi,\Sigma)=W(\Phi,\Sigma)-\widetilde{B}^*(\Phi)$ *and* $\tau_n(\Phi,\Sigma)=W\left(\widetilde{\Phi}_n,\Sigma\right)-W\left(\widetilde{\Phi}_n,\xi\left(\widetilde{\Phi}_n^{-1}\right)\right)$.
$\tau_n(\cdot,\cdot,\cdot,\cdot)$ *is a symmetric, non negative matrix for all* $n\in I\!N$, $\tau(\cdot,\cdot,\cdot,\cdot)$ *is a symmetric matrix and we also have that* $\lim_{n\to+\infty}\|\tau_n(\Phi,\Sigma)-\tau(\Phi,\Sigma)\|_r=0$. *Therefore* $\lim_{n\to+\infty}Sp\left[\tau_n(\Phi,\Sigma)\right]=Sp\left[\tau(\Phi,\Sigma)\right]$. *We know that* $Sp\left[\tau_n(\Phi,\Sigma)\right]\in(I\!R_+)^r/\{0,\ldots,0\}\Longrightarrow Sp\left[\tau(\Phi,\Sigma)\right]\in(I\!R_+)^r$ *; or in other words that* $\tau(\Phi,\Sigma)>>0$.

Endowed with theorems $4.1-4.2$, we are now able to rigorously state the remarks stressed below proposition 4.5.

Theorem 4.3 : *We consider the SALS estimator* $\widetilde{\widetilde{\theta}}^*_{TS}(\varepsilon,\lambda_n)$ *deduced from the modified simulator* (4.19)*:*

$$\widetilde{\widetilde{\theta}}^*_{TS}(\varepsilon,\lambda_n)=\underset{\theta\in\Theta}{Argmin}\left[\widetilde{\widetilde{H}}_{TS}\left(\theta,\widehat{\beta}_T,\varepsilon,\lambda_n\right)\right]'\widehat{\Sigma}^*_T\left[\widetilde{\widetilde{H}}_{TS}\left(\theta,\widehat{\beta}_T,\varepsilon,\lambda_n\right)\right],\tag{4.25}$$

where $\pi_\circ\lim_{T\to+\infty}\widehat{\Sigma}^*_T=\widetilde{\widetilde{\Sigma}}^*=\widetilde{\widetilde{\Phi}}_\circ(S,\varepsilon,\lambda_n)^{-1}=\left[\widetilde{\Phi}_\circ(S)+\lambda_n^2\widetilde{\widetilde{\Psi}}_\circ(S,\varepsilon)\right]^{-1}$ *and* $\{\lambda_n,n\in I\!N\}$ *is a decreasing sequence converging to* 0.
Then under assumptions $(A1)$, $(A13)$, $(A14)$, $(A15a)$ *and that the modified simulator* $\widetilde{\widetilde{H}}_{TS}(\theta,\beta,\varepsilon,\lambda)$ *is such that for all* $\{\lambda_n,n\in I\!N\}$, $(A17)$, $(A19a)$ *are fulfilled and the earlier assumptions* $(A18)$, $(A19b)$, $(A20)$ *on the initial simulator* $\widetilde{H}_{TS}(\theta,\beta)$ *are fulfilled, the SALS estimator* $\widetilde{\widetilde{\theta}}^*_{TS}(\varepsilon,\lambda_n)$ *is consistent, and asymptotically normal. Moreover its asymptotic covariance matrix* $\underset{\circ}{V}ar_{as}\left[\sqrt{T}\left(\widetilde{\widetilde{\theta}}^*_{TS}(\varepsilon,\lambda_n)-\theta^\circ\right)\right]$ *is decreasing with* n *and we have:*

$$\lim_{n\to+\infty}\underset{\circ}{V}ar_{as}\left[\sqrt{T}\left(\widetilde{\widetilde{\theta}}^*_{TS}(\varepsilon,\lambda_n)-\theta^\circ\right)\right]=B^*\left(\frac{\partial H}{\partial\theta'}(\theta^\circ,\beta^\circ),\widetilde{\Phi}_\circ(S),\widetilde{\widetilde{\Psi}}_\circ(S,\varepsilon)\right)\cdot\tag{4.26}$$

In other words, the SALS estimator $\widetilde{\widetilde{\theta}}^*_{TS}(\varepsilon,\lambda_n)$ can be made as close as desired to the efficiency bound in the direction of $\widetilde{\widetilde{\Psi}}_\circ(S,\varepsilon)$.

Proof :
1) The consistency and the $\sqrt{T}$*-asymptotic normality follow from the application of proposition 4.4 to the modified simulator* $\widetilde{\widetilde{H}}_{TS}(\theta,\beta,\varepsilon,\lambda_n)$ *and thus to the modified SALS* $\widetilde{\widetilde{\theta}}^*_{TS}(\varepsilon,\lambda_n)$.

2) The efficiency property follows from the application of theorem 4.1 and where $A=\frac{\partial H}{\partial\theta'}(\theta^\circ,\beta^\circ)\,,\ \Phi=\widetilde{\Phi}_\circ(S)\,,\ \Psi=\widetilde{\widetilde{\Psi}}_\circ(S,\varepsilon)$.

It is clear that the choice of a particular $\widetilde{\widetilde{\Psi}}_\circ(S,\varepsilon)$ leads to a particular efficiency bound $B^*\left(\frac{\partial H}{\partial\theta'}(\theta^\circ,\beta^\circ),\widetilde{\Phi}_\circ(S),\widetilde{\widetilde{\Psi}}_\circ(S,\varepsilon)\right)$, the one corresponding to the direction $\widetilde{\widetilde{\Psi}}_\circ(S,\varepsilon)$. In the case where the efficiency bound varies with the direction or with the type of direction itself (straight line versus curvilinear), we already know that the different efficiency bounds are not comparable. Therefore one is implicitly specifying a particular loss function.

As already announced the SALS estimation nests several other estimation methods and this result is stated in proposition 4.6.

Proposition 4.6 : *The II, the EMM, the GII, the SMM, the SNLS and the SPML estimators are particular SALS estimators respectively associated with the following estimating equations:*

- $H^{II}(\theta,\beta^{\circ}) = \widetilde{\beta}(\theta) - \beta^{\circ}$, *and where* $\widetilde{\beta}(\theta)$ *and* β° *are defined by* (A3),
- $H^{EMM}(\theta,\beta^{\circ}) = \underset{\theta}{E}\left[s_N(\beta^{\circ})\right]$, $s_N(\beta^{\circ})$ *is the seminonparametric score generator,*
- $H^{GII}(\theta,\beta^{\circ}) = \widetilde{\beta}(\theta,\Lambda^*) - \beta^{\circ}$, *where* $\widetilde{\beta}(\theta,\Lambda^*)$ *and* β° *are defined by* (3.16) *and* (3.1),
- $H^{SMM}(\theta,\beta^{\circ}) = \underset{\circ}{E}\left[g(w_t,\theta)\right]$, *where* $\underset{\circ}{E}\left[g(w_t,\theta)\right]$ *is computed through a simulator* $\widetilde{H}_{TS}(\cdot,\cdot)$ *obeying to assumption* (A16),
- $H^{SNLS}(\theta) = \underset{\circ}{E}\left[b_\ell^w - m_\ell(\theta)\right]$, *where* b_ℓ^w *is the winning bid and* $m_\ell(\theta)$ *the associated moments for the given distribution* F_θ *of the private values,*
- $H^{SPML}(\theta,\beta^{\circ}) = \underset{\theta}{E}\left[\widetilde{s}(\beta^{\circ})\right]$, $\widetilde{s}(\beta^{\circ})$ *is the score associated with the exponential p.d.f. family.* β° *correspond to nuisance parameters and again the simulator* $\widetilde{H}_{TS}(\cdot,\cdot)$ *corresponds to assumption* (A16)$\cdot$

Moreover the indirect and the GII estimators are respectively asymptotically equivalent to the following SALS estimators associated with the following estimating equations:

- $H^{II}(\theta,\beta^{\circ}) = \dfrac{\partial q}{\partial \beta}(\theta,\beta^{\circ})$, *and where* $q(\cdot,\cdot)$ *is defined by* (A2),
- $H^{GII}(\theta,\beta^{\circ}) = H^{\Lambda^*}(\theta,\beta^{\circ}) = \dfrac{\partial z'}{\partial \beta}(\theta,\beta^{\circ})\,\Lambda^* z(\theta,\beta^{\circ})$, $z(\theta,\beta)$ *is defined by* (3.16),

and where the **natural** *estimator for the auxiliary parameters* $\widehat{\beta}_T$ *is the one provided by the direct estimation of the instrumental criterion defined respectively for the Indirect Inference, the EMM, the GII and the SPML approaches.*

Proof : *The proofs of these results are straightforward and omitted here, it is indeed just a question of replacement of each expression by each particular form and of usual asymptotic expansions (see however the proof of proposition* 5.2*).*

In other words, either SPML, SMM, SNLS, Indirect Inference, EMM or GII estimation just correspond to a particular selection of "estimating equations" which are not tractable and thus replaced by simulation-based approximation.

5 Specification Tests

As for the Indirect Inference and the EMM, we propose a generalized global specification test. We define the statistic ξ_{TS}^{SALS} as follows:

$$\xi_{TS}^{SALS} = T \min_{\theta \in \Theta} \left[\frac{1}{S} \sum_{s=1}^{S} \widetilde{H}_T \left(\underline{\widetilde{y}}_T^s(\theta, z_\circ^s), \underline{x}_T, \widehat{\beta}_T \right) \right]' \widehat{\Sigma}_T^* \left[\frac{1}{S} \sum_{s=1}^{S} \widetilde{H}_T \left(\underline{\widetilde{y}}_T^s(\theta, z_\circ^s), \underline{x}_T, \widehat{\beta}_T \right) \right], \tag{5.1}$$

where $\widehat{\beta}_T$ is defined according to (4.10), $\pi_\circ \lim_{T \to +\infty} \widehat{\Sigma}_T^* = \Sigma^* = \widetilde{\Phi}_\circ(S)^{-1}$.

Proposition 5.1 : *Under the assumption that the structural model* (2.1) − (2.2) *is well-specified and assumptions* (A1), (A13) − (A15) *and* (A17) − (A20), *the statistic ξ_{TS}^{SALS} is asymptotically distributed as a chi-square with $(r-p)$ degrees of freedom $\chi^2(r-p)$.*
Therefore the test of asymptotic level α is associated with the critical region:

$$\mathcal{W}_\alpha^{SALS} = \{\xi_{TS}^{SALS} > \chi_{1-\alpha}^2(r-p)\}. \tag{5.2}$$

Proof : *See appendix A.5.*

In the case of the GII approach, we advocate the following testing strategy. A specification test for the structural model (2.1) − (2.2) when the instrumental model is of a GMM type is based on a two-steps procedure.
First, in our general setting, we have to test the overidentifying moment restrictions:

$$\underset{\circ}{E}[g(w_t, \beta^\circ)] = 0, \tag{5.3}$$

that is the existence of a unique β° such that these restrictions are fulfilled. This implicitly assumes that one is performing an overidentified estimation for the parameters β° that is $r > q$. Following Hansen (1982) overidentifying test we define the statistic ξ_T^1 by:

$$\xi_T^1 = T \min_{\beta \in \mathcal{B}} \left[\frac{1}{T} \sum_{t=1}^{T} g(w_t, \beta) \right]' \overline{\Lambda}_T \left[\frac{1}{T} \sum_{t=1}^{T} g(w_t, \beta) \right], \tag{5.4}$$

where $\overline{\Lambda}_T$ is a consistent estimator of $\overline{\Lambda} = \left\{ \underset{\circ}{V}ar_{as} \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} g(w_t, \beta^\circ) \right] \right\}^{-1}$. Indeed, the overidentifying test has to be performed with an optimal estimator $\widehat{\beta}_T^*$ of β°. Under assumptions 3.1 − 3.6 of Hansen (1982), the statistic ξ_T^1 converges in distribution to a chi-square distributed random variable with $(r-q)$ degrees of freedom: $\chi^2(r-q)$. The test of asymptotic level α is associated with the critical region:

$$\mathcal{W}_\alpha^1 = \{\xi_T^1 > \chi_{1-\alpha}^2(r-q)\}. \tag{5.5}$$

Second, after having tested the overidentifying restrictions, a specification test for the structural model (2.1) − (2.2) may be based on the optimal value of the objective function used in the second step of the indirect estimation method. Indeed we have:

Proposition 5.2 : *Under the null hypothesis that the structural model* $(2.1)-(2.2)$ *is well specified and assumptions* $(A1)-(A11)$, *the statistics:*

$$\xi_{TS}^{2}=\frac{S}{1+S}\min_{\theta\in\Theta}\left\{\begin{array}{c}\left[\frac{1}{TS}\sum_{s=1}^{S}\sum_{t=1}^{T}\frac{\partial g'}{\partial\beta}\left(\widetilde{w}_t^s(\theta),\widehat{\beta}_T\right)\widehat{\Lambda}_T^*\frac{1}{\sqrt{TS}}\sum_{s=1}^{S}\sum_{t=1}^{T}g\left(\widetilde{w}_t^s(\theta),\widehat{\beta}_T\right)\right]'\\ \widehat{\Sigma}_T\\ \left[\frac{1}{TS}\sum_{s=1}^{S}\sum_{t=1}^{T}\frac{\partial g'}{\partial\beta}\left(\widetilde{w}_t^s(\theta),\widehat{\beta}_T\right)\widehat{\Lambda}_T^*\frac{1}{\sqrt{TS}}\sum_{s=1}^{S}\sum_{t=1}^{T}g\left(\widetilde{w}_t^s(\theta),\widehat{\beta}_T\right)\right]\end{array}\right\}, \tag{5.6}$$

$$\xi_{TS}^{3}=\frac{S}{1+S}\min_{\theta\in\Theta}\left\{\begin{array}{c}\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial g'}{\partial\beta}\left(w_t,\widehat{\beta}_T\right)\widehat{\Lambda}_T^*\frac{1}{\sqrt{TS}}\sum_{s=1}^{S}\sum_{t=1}^{T}g\left(\widetilde{w}_t^s(\theta),\widehat{\beta}_T\right)\right]'\\ \widehat{\Sigma}_T\\ \left[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial g'}{\partial\beta}\left(w_t,\widehat{\beta}_T\right)\widehat{\Lambda}_T^*\frac{1}{\sqrt{TS}}\sum_{s=1}^{S}\sum_{t=1}^{T}g\left(\widetilde{w}_t^s(\theta),\widehat{\beta}_T\right)\right]\end{array}\right\}, \tag{5.7}$$

$$\xi_{TS}^{4}=\frac{S}{1+S}\min_{\theta\in\Theta}\left\{\begin{array}{c}\left[\frac{1}{S}\sum_{s=1}^{S}\left[\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial g'}{\partial\beta}\left(\widetilde{w}_t^s(\theta),\widehat{\beta}_T\right)\right)\widehat{\Lambda}_T^*\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}g\left(\widetilde{w}_t^s(\theta),\widehat{\beta}_T\right)\right)\right]\right]'\\ \widehat{\Sigma}_T\\ \left[\frac{1}{S}\sum_{s=1}^{S}\left[\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial g'}{\partial\beta}\left(\widetilde{w}_t^s(\theta),\widehat{\beta}_T\right)\right)\widehat{\Lambda}_T^*\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}g\left(\widetilde{w}_t^s(\theta),\widehat{\beta}_T\right)\right)\right]\right]\end{array}\right\}, \tag{5.8}$$

are asymptotically equivalent and distributed as chi-squares with $(q-p)$ *degrees of freedom* $\chi^2(q-p)$, *where* $\pi_\circ\lim_{T\to+\infty}\widehat{\Sigma}_T=\Sigma_\circ=\Phi_\circ\left(\Lambda^*\right)^{-1}$ *as defined by* (3.13) *and* $\pi_\circ\lim_{T\to+\infty}\widehat{\Lambda}_T^*=\Lambda^*$.
Therefore the tests of asymptotic level α *are associated with the critical regions* $\mathcal{W}_\alpha^i$, $i=2,3,4$*:*

$$\mathcal{W}_\alpha^i=\{\xi_{TS}^i>\chi_{1-\alpha}^2(q-p)\},\quad i=2,3,4\cdot \tag{5.9}$$

Proof : *These results correspond just to an application of proposition* 5.1, *see however appendix A.6.*

Note however, that under the joint null hypothesis that $(2.1)-(2.2)$ is well specified and (3.1) holds, the result of proposition 5.2 is available whatever choice of Λ. But the power of the test against local alternatives does depend on this choice and is still an issue. In this respect, the test statistic ξ_{TS}^i, $i=2,3,4$ should be jointly performed with the test statistic ξ_T^1. This test can be seen as a generalized global specification test as proposed by Gallant and Tauchen (1996) to the case of overidentifying moment type instrumental criterion. This test can also be regarded as a particular generalized global specification test as developed for the SALS estimator.

6 Concluding Remarks

The main messages of this paper are twofold:

• The SMM, SPML, SNLS, Indirect Inference and EMM are particular cases of the SALS approach and therefore to each simulation-based estimation correspond particular estimating equations.

• The SALS approach enables the treatment of new problems arising from the macroeconometric and econometric literature which cannot be properly handled within the common simulation-based methods. For instance, substantial efficiency gains are achieved when one introduces constraints (overidentifying moment conditions) on the instrumental model. The weighting matrix does no longer correspond to the classical GMM one.

References

Andrews, D. W. K. (1987), "Asymptotic Results for Generalized Wald Tests", *Econometric Theory* , 3, 348-358.

Andrews, D. W. K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation", *Econometrica*, 59, 307-346.

Andrews, D. W. K. (1992), "Generic Uniform Convergence", *Econometric Theory*, 8, 241-257.

Andrews, D. W. K. (1994), "Asymptotics for Semiparametric Econometric Models Via Stochastic Equicontinuity", *Econometrica*, 62, 43-72.

Andrews, D. W. K. (1994), "Empirical Process Methods in Econometrics", in R. F. Engle and D. L. McFadden (eds.), *Handbook of Econometrics*, Vol IV, Amsterdam: North Holland, 2247-2294.

Bansal, R., R. A. Gallant, R. Hussey and G. Tauchen (1995), "Nonparametric Estimation of Structural Models for High-Frequency Currency Market Data", *Journal of Econometrics*, 66, 251-287.

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics*, 31, 307-327.

Bourbaki, N. (1968), "Theory of Sets", *Addison-Wesley, Reading, Massachusetts*, Elements of Mathematics, volume 3.

Broze, L., R. Dridi and E. Renault (1998), "Calibration of Structural Models by Semiparametric Indirect Inference", Mimeo Université de Toulouse.

Calzolari, G., G. Fiorentini and E. Sentana (1999), "Constrained EMM Estimation", Mimeo.

Chamberlain, G. (1982), "Multivariate Regression Models for Panel Data", *Journal of Econometrics*, 18, 5-46.

Chamberlain, G. (1984), "Panel Data", in Z. Griliches and M. D. Intriligator (eds.), *Handbook of Econometrics*, Vol. II, Amsterdam: North Holland, 1247-1318.

Chamberlain, G. (1987), "Asymptotic Efficiency in Estimation with Conditional Moments Restrictions", *Journal of Econometrics*, 34, 305-334.

Chamberlain, G. (1992), "Efficiency Bounds for Semiparametric Regression", *Econometrica*, 60, 567-597.

Clark, P. K. (1973), "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices", *Econometrica*, 41, 135-156.

Den Haan, W. and A. Marcet (1990), "Solving the Stochastic Growth Model by Parametrizing Expectations", *Journal of Business Economics and Statistics*, 8, 31-34.

Dridi, R. (1999), "Essays on Nonlinear Econometric Methodologies: Simulation, Calibration and Instrumental Models", PhD Toulouse.

Dridi, R. and E. Renault (2000), "Semiparametric Indirect Inference", Sticerd Discussion Papers EM 00/392.

Drost, F. and N. Meddahi (1998), "Instrumental Variables Estimation of Stochastic Volatility Models", Mimeo.

Duffie, D. and K. J. Singleton (1993), "Simulated Moments Estimation of Markov Models of Asset Prices", *Econometrica*, 61, 929-952.

Eisenberg, M. (1971), "Axiomatic Theory of Sets and Classes", New York ; London: Holt, Rinehart and Winston, 1971.

Engle, R. (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of the U.K. Inflation", *Econometrica*, 50, 987-1008.

Epstein, L. and S. Zin (1991), "Substitution, Risk Aversion, and Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis", *Journal of Political Economy*, 199, 263-286.

Gallant, R. A. (1987), "Nonlinear Statistical Models", John Wiley, New York.

Gallant, R. A. and G. Tauchen (1989), "Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications", *Econometrica*, 57, 1091-1120.

Gallant, R. A. and G. Tauchen (1996), "Which Moments to Match", *Econometric Theory*, 12, 657-681.

Gallant, R. A. and H. White (1988), "A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models", Basil Blackwell, Oxford.

Gouriéroux, C. and A. Monfort (1989), "A General Framework for Testing a Null Hypothesis in a "Mixed" Form", *Econometric Theory*, 5, 63-82.

Gouriéroux, C. and A. Monfort (1995a), "Statistics and Econometric Models", volumes I and II, Cambridge: Cambridge University Press.

Gouriéroux, C. and A. Monfort (1995b), "Simulation-Based Econometrics", Core Lecture, Oxford: Oxford University Press.

Gouriéroux, C. and A. Monfort (1995c), "Testing, Encompassing and Simulating Dynamic Econometric Models", *Econometric Theory*, 11, 195-228.

Gouriéroux, C., A. Monfort and E. Renault (1993), "Indirect Inference", *Journal of Applied Econometrics*, 8, 85-118.

Gouriéroux, C., A. Monfort and A. Trognon (1985), "Moindres Carrés Asymptotiques", *Annales de l'INSEE*, 58, 91-121.

Hajivassiliou, V. and P. Rudd (1994), "Estimation by Simulation", in R. F. Engle and D. L. McFadden (eds.), *Handbook of Econometrics*, Vol IV, Amsterdam: North Holland, 2383-2441.

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

Hansen, L. P. (1985), "A Method for Calculating Bounds on the Asymptotics Covariance Matrices of Generalized Method of Moments Estimators", *Journal of Econometrics*, 30, 203-238.

Hansen, L. P., J. Heaton and A. Yaron (1996), "Finite Sample Properties of Some Alternative GMM Estimators", *Journal of Business Economics and Statistics*, 14, 262-280.

Hansen, L. P. and K. J. Singleton (1982), "Generalized Instrumental Variables Estimators of Nonlinear Rational Expectations Models", *Econometrica*, 50, 1269-1286.

Hansen, L. P. and K. J. Singleton (1996), "Efficient Estimation of Linear Asset Pricing Models with Moving Average", *Journal of Business Economics and Statistics*, 14, 53-68.

Hausman, J. A., W. K. Newey and W. E. Taylor (1987), "Efficient Estimation and Identification of Simultaneous Equations Models with Covariance Restrictions", *Econometrica*, 55, 849-874.

Ingram, B. F. and B. S. Lee (1991), "Estimation by Simulation of Time Series Models", *Journal of Econometrics*, 47, 197-207.

(1990), *Journal of Business Economics and Statistics*, 8.

Kodde, D. A., F. C. Palm and G. A. Pfann (1990), "Asymptotic Least-Squares Estimation Efficiency Considerations and Applications", *Journal of Applied Econometrics*, 5, 229-243.

Kocherlakota, N. R. (1996), "The Equity Premium: It's Still a Puzzle", *Journal of Economic Literature*, XXXIV, 42-71.

Laffont, J. J., H. Ossard and Q. Vuong (1995), "Econometrics of First-Price Auctions", *Econometrica*, 63, 953-980.

Laroque, G. and B. Salanié (1989), "Estimation of Multi-Market Fix-Price Models: an Application of Pseudo Maximum Likelihood Methods", *Econometrica*, 57, 831-860.

Lucas, R. E. (1978), "Asset Prices in an Exchange Economy", *Econometrica*, 46, 1429-1445.

Lütekepohl, H. and M. M. Burda (1997), "Modified Wald Tests under Nonregular Condition", *Journal of Econometrics*, 78, 315-322.

McFadden, D. L. (1989), "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration", *Econometrica*, 57, 995-1026.

Meddahi, N. and E. Renault (1997), "Aggregations and Marginalization of GARCH and Stochastic Volatility Model", GREMAQ DP 96.30.433.

Monfort, A. and R. Rabemananjara (1990), "From a VAR to a Structural Model, with an Application to the Wage-Price Spiral", *Journal of Applied Econometrics*, 5, 203-227.

Newey, W. K. (1991), "Uniform Convergence in Probability and Stochastic Equicontinuity", *Econometrica*, 59, 1161-1167.

Newey, W. K. (1994), "The Asymptotic Variance of Semiparametric Estimators", *Econometrica*, 62, 1349-1382.

Newey, W. K. and D. L. McFadden (1994), "Large Sample Estimation and Hypothesis Testing", in R. F. Engle and D. L. McFadden (eds.), *Handbook of Econometrics*, Vol IV, Amsterdam: North Holland, 2111-2245.

Newey, W. K. and K. D. West (1987), "A Simple, Positive Definite Heteroscedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica*, 55, 703-708.

Ogaki, M. (1993), "Generalized Method of Moments: Econometric Applications", in G. S. Maddala, C. R. Rao and H. D. Vinod (eds.), *Handbook of Statistics*, Vol XI, Amsterdam: North Holland, 455-488.

Pakes, A. and D. Pollard (1989), "Simulation and the Asymptotics of Optimization Estimators", *Econometrica*, 57, 1027-1058.

Robinson, P. M. (1988), "Root-N-consistent Semiparametric Regression", *Econometrica*, 56, 931-954.

Smith, A. (1993), "Estimating Nonlinear Time Series Models using Simulated Vector Autoregressions", *Journal of Applied Econometrics*, 8, 63-84.

Szroeter, A. (1983), "Generalized Wald Methods for Testing Nonlinear Implicit and Overidentifying Restrictions", *Econometrica*, 51, 335-353.

Tauchen, G. (1997), "New Minimum Chi-Square Methods in Empirical Finance", *Econometric Society Monographs*, n°28, 279-317.

Tauchen, G. and M. Pitts (1983), "The Price Variability Volume Relationship on Speculative Markets", *Econometrica*, 51, 485-505.

Taylor, S. J. (1986), "Modeling Financial Time Series", John Wiley, Chichester, UK.

Taylor, S. J. (1994), "Modeling Stochastic Volatility: A Review and Comparative Study", *Mathematical Finance*, 4, 183-204.

White, H. (1980), "Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct test for Heteroskedasticity", *Econometrica*, 48, 817-836.

Wooldridge, J. M. (1994), "Estimation and Inference for Dependent Processes", in R. F. Engle and D. L. McFadden (eds.), *Handbook of Econometrics*, Vol IV, Amsterdam: North Holland, 2639-2738.

Appendices

A.1. Proof of proposition 2.2:

Since the observed endogenous variables $\underline{y}_T$ can always be regarded as a simulated path of the endogenous variables at the value $\theta = \theta^\circ$ and because of the exogeneity assumption on $\{x_t, t \in \mathbb{Z}\}$, it suffices to prove that $\widetilde{\beta}^s_T(\theta)$ converges to $\widetilde{\beta}(\theta)$.
We have thanks to proposition 2.1:

$$\forall \theta \in \Theta,\ \forall s = 1, \ldots, S,$$

$$\pi_\circ \lim_{T \to +\infty} \mathop{Sup}_{\beta \in \mathcal{B}} \left| Q_T\left(\underline{\widetilde{y}}^s_T(\theta, z^s_\circ), \underline{x}_T, \beta\right) - q\left(\theta, \beta\right) \right| = 0,$$

$$\Longleftrightarrow \quad \forall \theta \in \Theta,\ \forall s = 1, \ldots, S,\ \forall \varepsilon > 0,\ \forall \eta > 0,\ \exists T_{\theta,\varepsilon,\eta}\ /\ \forall T \geq T_{\theta,\varepsilon,\eta},\ \forall \beta \in \mathcal{B}:$$

$$\pi_\circ \left[\left| Q_T\left(\underline{\widetilde{y}}^s_T(\theta, z^s_\circ), \underline{x}_T, \beta\right) - q\left(\theta, \beta\right) \right| < \frac{\eta}{3} \right] > 1 - \frac{\varepsilon}{2},$$

$$\Longrightarrow \quad \forall \theta \in \Theta,\ \forall s = 1, \ldots, S,\ \forall \varepsilon > 0,\ \forall \eta > 0,\ \exists T_{\theta,\varepsilon,\eta}\ /\ \forall T \geq T_{\theta,\varepsilon,\eta}:$$

$$(i)\ \pi_\circ \left[q\left(\theta, \widetilde{\beta}^s_T(\theta)\right) < Q_T\left(\underline{\widetilde{y}}^s_T(\theta, z^s_\circ), \underline{x}_T, \widetilde{\beta}^s_T(\theta)\right) + \frac{\eta}{3} \right] > 1 - \frac{\varepsilon}{2},$$

$$(ii)\ \pi_\circ \left[Q_T\left(\underline{\widetilde{y}}^s_T(\theta, z^s_\circ), \underline{x}_T, \widetilde{\beta}(\theta)\right) < q\left(\theta, \widetilde{\beta}(\theta)\right) + \frac{\eta}{3} \right] > 1 - \frac{\varepsilon}{2},$$

We also have:
$\forall \eta > 0,\ Q_T\left(\underline{\widetilde{y}}^s_T(\theta, z^s_\circ), \underline{x}_T, \widetilde{\beta}^s_T(\theta)\right) < Q_T\left(\underline{\widetilde{y}}^s_T(\theta, z^s_\circ), \underline{x}_T, \widetilde{\beta}(\theta)\right) + \frac{\eta}{3},$
since $\widetilde{\beta}^s_T(\theta)$ corresponds to the $Argmin$ of (2.5). We now define the probability Ψ:

$$\Psi = \pi_\circ \left\{ \left[q\left(\theta, \widetilde{\beta}^s_T(\theta)\right) < Q_T\left(\underline{\widetilde{y}}^s_T(\theta, z^s_\circ), \underline{x}_T, \widetilde{\beta}^s_T(\theta)\right) + \frac{\eta}{3} \right] \ and \right.$$

$$\left[Q_T\left(\underline{\widetilde{y}}^s_T(\theta, z^s_\circ), \underline{x}_T, \widetilde{\beta}(\theta)\right) < q\left(\theta, \widetilde{\beta}(\theta)\right) + \frac{\eta}{3} \right] \ and$$

$$\left. \left[Q_T\left(\underline{\widetilde{y}}^s_T(\theta, z^s_\circ), \underline{x}_T, \widetilde{\beta}^s_T(\theta)\right) < Q_T\left(\underline{\widetilde{y}}^s_T(\theta, z^s_\circ), \underline{x}_T, \widetilde{\beta}(\theta)\right) + \frac{\eta}{3} \right] \right\},$$

$$\Psi = \pi_\circ \left\{ \left[q\left(\theta, \widetilde{\beta}^s_T(\theta)\right) < Q_T\left(\underline{\widetilde{y}}^s_T(\theta, z^s_\circ), \underline{x}_T, \widetilde{\beta}^s_T(\theta)\right) + \frac{\eta}{3} \right] \right\} +$$

$$\pi_\circ \left\{ \left[Q_T\left(\underline{\widetilde{y}}^s_T(\theta, z^s_\circ), \underline{x}_T, \widetilde{\beta}(\theta)\right) < q\left(\theta, \widetilde{\beta}(\theta)\right) + \frac{\eta}{3} \right] \right\} -$$

$$\pi_\circ \left\{ \left[q\left(\theta, \widetilde{\beta}^s_T(\theta)\right) < Q_T\left(\underline{\widetilde{y}}^s_T(\theta, z^s_\circ), \underline{x}_T, \widetilde{\beta}^s_T(\theta)\right) + \frac{\eta}{3} \right] \ or \right.$$

$$\left. \left[Q_T\left(\underline{\widetilde{y}}^s_T(\theta, z^s_\circ), \underline{x}_T, \widetilde{\beta}(\theta)\right) < q\left(\theta, \widetilde{\beta}(\theta)\right) + \frac{\eta}{3} \right] \right\},$$

$$\Psi \geq \pi_\circ \left\{ \left[q\left(\theta, \widetilde{\beta}_T^s(\theta)\right) < Q_T\left(\underline{\widetilde{y}}_T^s(\theta, z_\circ^s), \underline{x}_T, \widetilde{\beta}_T^s(\theta)\right) + \frac{\eta}{3} \right] \right\} +$$

$$\pi_\circ \left\{ \left[Q_T\left(\underline{\widetilde{y}}_T^s(\theta, z_\circ^s), \underline{x}_T, \widetilde{\beta}(\theta)\right) < q\left(\theta, \widetilde{\beta}(\theta)\right) + \frac{\eta}{3} \right] \right\} - 1\cdot$$

Using (i) and (ii) we deduce that:

$$\forall \theta \in \Theta,\ \forall s = 1, \ldots, S,\ \forall \varepsilon > 0,\ \forall \eta > 0,\ \exists T_{\theta,\varepsilon,\eta}\ /\ \forall T \geq T_{\theta,\varepsilon,\eta}:$$

$$\Psi > \left(1 - \frac{\varepsilon}{2}\right) + \left(1 - \frac{\varepsilon}{2}\right) - 1 = 1 - \varepsilon\cdot$$

$$\left\{ \left[q\left(\theta, \widetilde{\beta}_T^s(\theta)\right) < Q_T\left(\underline{\widetilde{y}}_T^s(\theta, z_\circ^s), \underline{x}_T, \widetilde{\beta}_T^s(\theta)\right) + \frac{\eta}{3} \right] \ \textit{and} \right.$$

$$\left[Q_T\left(\underline{\widetilde{y}}_T^s(\theta, z_\circ^s), \underline{x}_T, \widetilde{\beta}(\theta)\right) < q\left(\theta, \widetilde{\beta}(\theta)\right) + \frac{\eta}{3} \right] \ \textit{and}$$

$$\left. \left[Q_T\left(\underline{\widetilde{y}}_T^s(\theta, z_\circ^s), \underline{x}_T, \widetilde{\beta}_T^s(\theta)\right) < Q_T\left(\underline{\widetilde{y}}_T^s(\theta, z_\circ^s), \underline{x}_T, \widetilde{\beta}(\theta)\right) + \frac{\eta}{3} \right] \right\},$$

$$\Longrightarrow \quad \left\{ \left[q\left(\theta, \widetilde{\beta}_T^s(\theta)\right) < q\left(\theta, \widetilde{\beta}(\theta)\right) + \eta \right] \right\},$$

so that:

$$\pi_\circ \left\{ \left[q\left(\theta, \widetilde{\beta}_T^s(\theta)\right) < q\left(\theta, \widetilde{\beta}(\theta)\right) + \eta \right] \right\} \geq \psi > 1 - \varepsilon,$$

In other words we have:

$$\forall \theta \in \Theta,\ \forall s = 1, \ldots, S,\ \forall \varepsilon > 0,\ \forall \eta > 0,\ \exists T_{\theta,\varepsilon,\eta}\ /\ \forall T \geq T_{\theta,\varepsilon,\eta}:$$

$$\pi_\circ \left\{ \left[q\left(\theta, \widetilde{\beta}_T^s(\theta)\right) < q\left(\theta, \widetilde{\beta}(\theta)\right) + \eta \right] \right\} \geq \psi > 1 - \varepsilon\cdot$$

Let now $\mathcal{N}_\theta$ be any open subset of $\mathcal{B}$ containing $\widetilde{\beta}(\theta)$ (we have assumed that $\widetilde{\beta}(\theta) \in \mathring{\mathcal{B}}$). $\mathcal{B} \cap \mathcal{N}_\theta^c$ is a compact set. We define $\widetilde{\beta}^*(\theta)$ by:

$$\widetilde{\beta}^*(\theta) = \underset{\beta \in \mathcal{B} \cap \mathcal{N}_\theta^c}{Argmin}\, q(\theta, \beta) \cdot$$

We have: $\quad \forall \theta \in \Theta, \quad q\left(\theta, \widetilde{\beta}^*(\theta)\right) > q\left(\theta, \widetilde{\beta}(\theta)\right),$

since $\widetilde{\beta}(\theta)$ is the **unique** minimizer of $q(\theta, \cdot)$ and $\widetilde{\beta}^*(\theta) \in \mathcal{B} \cap \mathcal{N}_\theta^c$. ($\widetilde{\beta}(\theta) \in \mathcal{N}_\theta$). Let us define $\eta^*(\theta) = q\left(\theta, \widetilde{\beta}^*(\theta)\right) - q\left(\theta, \widetilde{\beta}(\theta)\right) > 0$. For this particular value of η we have:

$$\forall\theta\in\Theta,\ \forall s=1,\ldots,S,\ \forall\varepsilon>0,\ \exists T_{\theta,\varepsilon}\ /\ \forall T\geq T_{\theta,\varepsilon}:$$

$$\pi_\circ\left\{\left[q\left(\theta,\widetilde{\beta}_T^s(\theta)\right)<q\left(\theta,\widetilde{\beta}(\theta)\right)+\eta^*(\theta)\right]\right\}>1-\varepsilon,$$

$$\Longrightarrow\quad\forall\theta\in\Theta,\ \forall s=1,\ldots,S,\ \forall\varepsilon>0,\ \exists T_{\theta,\varepsilon}\ /\ \forall T\geq T_{\theta,\varepsilon}:$$

$$\pi_\circ\left\{\left[q\left(\theta,\widetilde{\beta}_T^s(\theta)\right)<q\left(\theta,\widetilde{\beta}^*(\theta)\right)\right]\right\}>1-\varepsilon,$$

since $\eta^*(\theta)=q\left(\theta,\widetilde{\beta}^*(\theta)\right)-q\left(\theta,\widetilde{\beta}(\theta)\right)>0,$

$$\Longrightarrow\quad\forall\theta\in\Theta,\ \forall s=1,\ldots,S,\ \forall\varepsilon>0,\ \exists T_{\theta,\varepsilon}\ /\ \forall T\geq T_{\theta,\varepsilon}:$$

$$\pi_\circ\left\{\widetilde{\beta}_T^s(\theta)\notin\mathcal{B}\cap\mathcal{N}_\theta^c\right\}>1-\varepsilon,$$

since $\widetilde{\beta}^*(\theta)=\underset{\beta\in\mathcal{B}\cap\mathcal{N}_\theta^c}{Argmin}\,q\left(\theta,\beta\right),$

and $\eta^*(\theta)=q\left(\theta,\widetilde{\beta}^*(\theta)\right)-q\left(\theta,\widetilde{\beta}(\theta)\right)>0\cdot$

Or in other words, with a probability approaching 1, $\widetilde{\beta}_T^s(\theta)\notin\mathcal{B}\cap\mathcal{N}_\theta^c\Longleftrightarrow\widetilde{\beta}_T^s(\theta)\in\mathcal{N}_\theta$ for each $\mathcal{N}_\theta$ open subset of $\mathcal{B}$ containing $\widetilde{\beta}(\theta)$. We have thus with probability aproaching 1 $\widetilde{\beta}_T^s(\theta)\in\mathcal{N}_\theta$ i.e.:

$$\forall\theta\in\Theta,\ \forall s=1,\ldots,S,:$$

$$\widetilde{\beta}_T^s(\theta)\underset{T\to+\infty}{\xrightarrow{\pi_\circ}}\widetilde{\beta}(\theta)\cdot$$

A.2. Proof of proposition 3.3:

We have:

$$\left[\underset{\circ}{V}ar_{as}\left(\sqrt{T}\widehat{\theta}_{TS}^{g^*}\right)\right]^{-1}\left(1+\frac{1}{S}\right)=\frac{\partial z'}{\partial\theta}\left(\theta^\circ,\beta^\circ\right)V^{-1}\frac{\partial z}{\partial\theta'}\left(\theta^\circ,\beta^\circ\right),$$

$$\left[\underset{\circ}{V}ar_{as}\left(\sqrt{T}\widehat{\theta}_{TS}^{g_1^*}\right)\right]^{-1}\left(1+\frac{1}{S}\right)=\frac{\partial {z_1}'}{\partial\theta}\left(\theta^\circ,\beta_1^\circ\right){V_{11}}^{-1}\frac{\partial z_1}{\partial\theta'}\left(\theta^\circ,\beta_1^\circ\right)\cdot$$

We first use the following matrix lemma:

Lemma: *Let V_{11} $(r_1\times r_1)$, V_{12} $(r_1\times r_2)$, V_{21} $(r_2\times r_1)$, V_{22} $(r_2\times r_2)$, X_1 $(p_1\times r_1)$, X_2 $(p_1\times r_2)$, Y_1 $(r_1\times p_2)$, and Y_2 $(r_2\times p_2)$, then:*

$$\begin{bmatrix}X_1,\\X_2\end{bmatrix}'\begin{bmatrix}V_{11}&V_{12}\\V_{21}&V_{22}\end{bmatrix}^{-1}\begin{bmatrix}Y_1\\Y_2\end{bmatrix}=X_1{V_{11}^{-1}}Y_1+\left[X_2-X_1{V_{11}^{-1}}V_{12}\right]\left[V_{22}-V_{21}{V_{11}^{-1}}V_{12}\right]^{-1}\left[Y_2-V_{21}{V_{11}^{-1}}Y_1\right]\cdot$$

Proof:

Indeed the block-inverse formula gives:

$$\begin{bmatrix}V_{11}&V_{12}\\V_{21}&V_{22}\end{bmatrix}^{-1}=\begin{bmatrix}(V^{-1})_{11}&(V^{-1})_{12}\\(V^{-1})_{21}&(V^{-1})_{22}\end{bmatrix},$$

with:

$$(V^{-1})_{11} = {V_{11}}^{-1} + {V_{11}}^{-1}V_{12}(V^{-1})_{22}V_{21}{V_{11}}^{-1},$$

$$(V^{-1})_{12} = -{V_{11}}^{-1}V_{12}(V^{-1})_{22},$$

$$(V^{-1})_{21} = -(V^{-1})_{22}V_{21}{V_{11}}^{-1},$$

$$(V^{-1})_{22} = \left(V_{22} - V_{21}V_{11}^{-1}V_{12}\right)^{-1}.$$

So that:

$$(X_1, X_2)\begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}^{-1}\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} =$$

$$(X_1, X_2)\begin{bmatrix} {V_{11}}^{-1}Y_1 - {V_{11}}^{-1}V_{12}(V^{-1})_{22}\left(Y_2 - V_{21}{V_{11}}^{-1}Y_1\right) \\ \\ (V^{-1})_{22}\left(Y_2 - V_{21}{V_{11}}^{-1}Y_1\right) \end{bmatrix},$$

$$= X_1{V_{11}}^{-1}Y_1 + \left(X_2 - X_1{V_{11}}^{-1}V_{12}\right)\left(V^{-1}\right)_{22}\left(Y_2 - V_{21}{V_{11}}^{-1}Y_1\right),$$

$$= X_1{V_{11}}^{-1}Y_1 + \left(X_2 - X_1{V_{11}}^{-1}V_{12}\right)\left(V_{22} - V_{21}{V_{11}}^{-1}V_{12}\right)^{-1}\left(Y_2 - V_{21}{V_{11}}^{-1}Y_1\right).$$

We apply the latter lemma to:

$$X_1 = \frac{\partial {z_1}'}{\partial \theta}\left(\theta^\circ, \beta_1^\circ\right),\ X_2 = \frac{\partial {z_2}'}{\partial \theta}\left(\theta^\circ, \beta^\circ\right),\ V = \underset{\circ}{V}ar_{as}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t^*\left(w_t, \beta^\circ\right)\right],\ Y_1 = X_1',\ Y_2 = X_2'.$$

Noticing also that $V_{12} = V_{21}'$, we thus obtain the result. Since $V_{11}' = V_{11}$, $V_{12}' = V_{21}$ and $\left(V_{22} - V_{21}{V_{11}}^{-1}V_{12}\right)^{-1} >> 0$ (inverse of a positive matrix), we have:

$$\left[\underset{\circ}{V}ar_{as}\left(\sqrt{T}\widehat{\theta}_{TS}^{g^*}\right)\right]^{-1}\left(1 + \frac{1}{S}\right) - \left[\underset{\circ}{V}ar_{as}\left(\sqrt{T}\widehat{\theta}_{TS}^{g_1^*}\right)\right]^{-1}\left(1 + \frac{1}{S}\right) =$$

$$\left[\frac{\partial z_2}{\partial \theta'}\left(\theta^\circ, \beta^\circ\right) - V_{21}{V_{11}}^{-1}\frac{\partial z_1}{\partial \theta'}\left(\theta^\circ, \beta_1^\circ\right)\right]'\left(V_{22} - V_{21}{V_{11}}^{-1}V_{21}'\right)^{-1}\left[\frac{\partial z_2}{\partial \theta'}\left(\theta^\circ, \beta^\circ\right) - V_{21}{V_{11}}^{-1}\frac{\partial z_1}{\partial \theta'}\left(\theta^\circ, \beta_1^\circ\right)\right] = \Psi'\Psi,$$

where: $\Psi = A^{\frac{1}{2}}\left[\frac{\partial z_2}{\partial \theta'}\left(\theta^\circ, \beta^\circ\right) - V_{21}V_{11}^{-1}\frac{\partial z_1}{\partial \theta'}\left(\theta^\circ, \beta_1^\circ\right)\right],\ A = \left(V_{22} - V_{21}{V_{11}}^{-1}V_{21}'\right)^{-1}.$

$\Psi'\Psi$ is a non negative matrix and null if and only if $\frac{\partial {z_2}'}{\partial \theta}\left(\theta^\circ, \beta^\circ\right) = \frac{\partial {z_1}'}{\partial \theta}\left(\theta^\circ, \beta_1^\circ\right){V_{11}}^{-1}V_{12}$.

A.3. Proof of proposition 4.3:

We have thanks to assumption $(A17)$:

$$\forall s=1,\ldots,S,\ \underset{T\to+\infty}{\pi_\circ \lim}\ \underset{\theta,\beta\in\Theta\times\mathcal{B}}{Sup}\left\|\widetilde{H}_T^s(\theta,\beta)-H(\theta,\beta)\right\|_r=0,$$

where: $\widetilde{H}_T^s(\theta,\beta)=\widetilde{H}_T\left(\underline{\widetilde{y}}_T^s(\theta,z_\circ^s),\underline{x}_T,\beta\right)$,

$$\Longrightarrow\quad \forall\ S \text{ fixed},\quad \underset{T\to+\infty}{\pi_\circ \lim}\ \underset{\theta,\beta\in\Theta\times\mathcal{B}}{Sup}\left\|\frac{1}{S}\sum_{s=1}^{S}\widetilde{H}_T^s(\theta,\beta)-H(\theta,\beta)\right\|_r=0,$$

We define:

$$\widetilde{Q}_{TS}(\theta,\beta)=\left[\frac{1}{S}\sum_{s=1}^{S}\widetilde{H}_T^s(\theta,\beta)\right]'\widehat{\Sigma}_T\left[\frac{1}{S}\sum_{s=1}^{S}\widetilde{H}_T^s(\theta,\beta)\right],$$

$$q(\theta,\beta)=H(\theta,\beta)'\Sigma H(\theta,\beta),$$

We have (obvious): $\underset{T\to+\infty}{\pi_\circ \lim}\ \underset{\theta,\beta\in\Theta\times\mathcal{B}}{Sup}\left|\widetilde{Q}_{TS}(\theta,\beta)-q(\theta,\beta)\right|=0,$

$$\Longleftrightarrow\quad \forall\varepsilon>0,\ \forall\eta>0,\ \exists T^1_{\varepsilon,\eta}\ /\ \forall T\geq T^1_{\varepsilon,\eta},\ \forall\theta,\beta\in\Theta\times\mathcal{B}:$$

$$\pi_\circ\left[\left|\widetilde{Q}_{TS}\left(\theta,\beta\right)-q\left(\theta,\beta\right)\right|<\frac{\eta}{6}\right]>1-\frac{\varepsilon}{5},$$

$$\Longrightarrow\quad \forall\varepsilon>0,\ \forall\eta>0,\ \exists T^1_{\varepsilon,\eta}\ /\ \forall T\geq T^1_{\varepsilon,\eta}:$$

$$(i)\ \pi_\circ\left[q\left(\widehat{\theta}_{TS}(\Sigma),\widehat{\beta}_T\right)<\widetilde{Q}_{TS}\left(\widehat{\theta}_{TS}(\Sigma),\widehat{\beta}_T\right)+\frac{\eta}{6}\right]>1-\frac{\varepsilon}{5},$$

$$(ii)\ \pi_\circ\left[\widetilde{Q}_{TS}\left(\theta^\circ,\widehat{\beta}_T\right)<q\left(\theta^\circ,\widehat{\beta}_T\right)+\frac{\eta}{6}\right]>1-\frac{\varepsilon}{5},$$

We also have: $\forall\eta>0,\ \widetilde{Q}_{TS}\left(\widehat{\theta}_{TS}(\Sigma),\widehat{\beta}_T\right)<\widetilde{Q}_{TS}\left(\theta^\circ,\widehat{\beta}_T\right)+\frac{\eta}{6}$, since $\widehat{\theta}_{TS}(\Sigma)$ corresponds to the unique minimum of $\widetilde{Q}_{TS}\left(\cdot,\widehat{\beta}_T\right)$.

We now define the probability Ψ:

$$\Psi=\pi_\circ\left\{\left[q\left(\widehat{\theta}_{TS}(\Sigma),\widehat{\beta}_T\right)<\widetilde{Q}_{TS}\left(\widehat{\theta}_{TS}(\Sigma),\widehat{\beta}_T\right)+\frac{\eta}{6}\right]\ and\right.$$

$$\left.\left[\widetilde{Q}_{TS}\left(\theta^\circ,\widehat{\beta}_T\right)<q\left(\theta^\circ,\widehat{\beta}_T\right)+\frac{\eta}{6}\right]\ and\ \left[\widetilde{Q}_{TS}\left(\widehat{\theta}_{TS}(\Sigma),\widehat{\beta}_T\right)<\widetilde{Q}_{TS}\left(\theta^\circ,\widehat{\beta}_T\right)+\frac{\eta}{6}\right]\right\},$$

$$\Psi=\pi_\circ\left\{\left[q\left(\widehat{\theta}_{TS}(\Sigma),\widehat{\beta}_T\right)<\widetilde{Q}_{TS}\left(\widehat{\theta}_{TS}(\Sigma),\widehat{\beta}_T\right)+\frac{\eta}{6}\right]\right\}+$$

$$\pi_\circ\left\{\left[\widetilde{Q}_{TS}\left(\theta^\circ,\widehat{\beta}_T\right)<q\left(\theta^\circ,\widehat{\beta}_T\right)+\frac{\eta}{6}\right]\right\}-$$

$$\pi_\circ\left\{\left[q\left(\widehat{\theta}_{TS}(\Sigma),\widehat{\beta}_T\right)<\widetilde{Q}_{TS}\left(\widehat{\theta}_{TS}(\Sigma),\widehat{\beta}_T\right)+\frac{\eta}{6}\right]\ or\ \left[\widetilde{Q}_{TS}\left(\theta^\circ,\widehat{\beta}_T\right)<q\left(\theta^\circ,\widehat{\beta}_T\right)+\frac{\eta}{6}\right]\right\},$$

$$\Psi\geq\pi_\circ\left\{\left[q\left(\widehat{\theta}_{TS}(\Sigma),\widehat{\beta}_T\right)<\widetilde{Q}_{TS}\left(\widehat{\theta}_{TS}(\Sigma),\widehat{\beta}_T\right)+\frac{\eta}{6}\right]\right\}+$$

$$\pi_\circ\left\{\left[\widetilde{Q}_{TS}\left(\widehat{\theta}_{TS}(\Sigma),\widehat{\beta}_T\right)<q\left(\theta^\circ,\widehat{\beta}_T\right)+\frac{\eta}{6}\right]\right\}-1\cdot$$

Using (i) and (ii) we deduce that:

$$\forall \varepsilon > 0,\ \forall \eta > 0,\ \exists T^1_{\varepsilon,\eta}\ /\ \forall T \geq T^1_{\varepsilon,\eta} : \Psi > \left(1 - \frac{\varepsilon}{5}\right) + \left(1 - \frac{\varepsilon}{5}\right) - 1 = 1 - \frac{2\varepsilon}{5}.$$

$$\left\{\left[q\left(\widehat{\theta}_{TS}(\Sigma), \widehat{\beta}_T\right) < \widetilde{Q}_{TS}\left(\widehat{\theta}_{TS}(\Sigma), \widehat{\beta}_T\right) + \frac{\eta}{6}\right] \ \textit{and}\ \left[\widetilde{Q}_{TS}\left(\theta^\circ, \widehat{\beta}_T\right) < q\left(\theta^\circ, \widehat{\beta}_T\right) + \frac{\eta}{6}\right]\right.$$

$$\left.\textit{and}\ \left[\widetilde{Q}_{TS}\left(\widehat{\theta}_{TS}(\Sigma), \widehat{\beta}_T\right) < \widetilde{Q}_{TS}\left(\theta^\circ, \widehat{\beta}_T\right) + \frac{\eta}{6}\right]\right\},$$

$$\Longrightarrow \quad \left\{\left[q\left(\widehat{\theta}_{TS}(\Sigma), \widehat{\beta}_T\right) < q\left(\theta^\circ, \widehat{\beta}_T\right) + \frac{\eta}{2}\right]\right\},$$

so that:

$$\pi_\circ\left\{\left[q\left(\widehat{\theta}_{TS}(\Sigma), \widehat{\beta}_T\right) < q\left(\theta^\circ, \widehat{\beta}_T\right) + \frac{\eta}{2}\right]\right\} \geq \psi > 1 - \frac{2\varepsilon}{5},$$

In other words we have:

$$\forall \varepsilon > 0,\ \forall \eta > 0,\ \exists T^1_{\varepsilon,\eta}\ /\ \forall T \geq T^1_{\varepsilon,\eta} :\ \pi_\circ\left\{\left[q\left(\widehat{\theta}_{TS}(\Sigma), \widehat{\beta}_T\right) < q\left(\theta^\circ, \widehat{\beta}_T\right) + \frac{\eta}{2}\right]\right\} \geq \psi > 1 - \frac{2\varepsilon}{5}.$$

Let $\widehat{\theta}_{\varphi(T)S}(\Sigma)$ be any subsequence of $\widehat{\theta}_{TS}(\Sigma)$ converging in $\pi_\circ$-probability to $\theta^*_\varphi(\Sigma)$ when T goes to infinity ($\varphi(\cdot)$ is an increasing function from $\mathbb{N}$ onto $\mathbb{N}$). We are going to show that $\theta^*_\varphi(\Sigma) = \theta^\circ$ and since Θ is compact this ends the proof of proposition 4.3.
Indeed since $\varphi(\cdot)$ is increasing and T and $\varphi(T)$ belong to $\mathbb{N}$, we have $\forall T \in \mathbb{N},\ \varphi(T) \geq T$ (proved by induction) and:

$$\forall \varepsilon > 0,\ \forall \eta > 0,\ \exists T^1_{\varepsilon,\eta}\ /\ \forall T \geq T^1_{\varepsilon,\eta} :$$

$$\pi_\circ\left\{\left[q\left(\widehat{\theta}_{\varphi(T)S}(\Sigma), \widehat{\beta}_{\varphi(T)}\right) < q\left(\theta^\circ, \widehat{\beta}_{\varphi(T)}\right) + \frac{\eta}{2}\right]\right\} > 1 - \frac{2\varepsilon}{5}.$$

Since $H(\cdot,\cdot)$ is assumed to be continuous, $q(\cdot,\cdot)$ is also continuous. Moreover since $\left(\widehat{\theta}_{\varphi(T)S}(\Sigma), \widehat{\beta}_{\varphi(T)}\right) \xrightarrow[T\to+\infty]{\pi_\circ} \left(\theta^*_\varphi(\Sigma), \beta^\circ\right)$ and $\left(\widehat{\theta}_{\varphi(T)S}(\Sigma), \beta^\circ\right) \xrightarrow[T\to+\infty]{\pi_\circ} \left(\theta^*_\varphi(\Sigma), \beta^\circ\right)$, we have:

$$\forall \varepsilon > 0,\ \forall \eta > 0,\ \exists T^2_{\varepsilon,\eta}\ /\ \forall T \geq T^2_{\varepsilon,\eta} :$$

- $\pi_\circ\left\{\left[\left|q\left(\widehat{\theta}_{\varphi(T)S}(\Sigma), \widehat{\beta}_{\varphi(T)}\right) - q\left(\theta^*_\varphi(\Sigma), \beta^\circ\right)\right| < \frac{\eta}{6}\right]\right\} > 1 - \frac{\varepsilon}{5}.$
- $\pi_\circ\left\{\left[\left|q\left(\widehat{\theta}_{\varphi(T)S}(\Sigma), \beta^\circ\right) - q\left(\theta^*_\varphi(\Sigma), \beta^\circ\right)\right| < \frac{\eta}{6}\right]\right\} > 1 - \frac{\varepsilon}{5}.$

We define the probability Φ as follows:

$$\Phi = \pi_\circ \left\{ \left[\left| q\left(\widehat{\theta}_{\varphi(T)S}(\Sigma), \widehat{\beta}_{\varphi(T)}\right) - q\left(\theta^*_\varphi(\Sigma), \beta^\circ\right) \right| < \frac{\eta}{6} \right] \ and \ \left[\left| q\left(\widehat{\theta}_{\varphi(T)S}(\Sigma), \beta^\circ\right) - q\left(\theta^*_\varphi(\Sigma), \beta^\circ\right) \right| < \frac{\eta}{6} \right] \right\},$$

$$\Phi = \pi_\circ \left\{ \left[\left| q\left(\widehat{\theta}_{\varphi(T)S}(\Sigma), \widehat{\beta}_{\varphi(T)}\right) - q\left(\theta^*_\varphi(\Sigma), \beta^\circ\right) \right| < \frac{\eta}{6} \right] \right\} + \pi_\circ \left\{ \left[\left| q\left(\widehat{\theta}_{\varphi(T)S}(\Sigma), \beta^\circ\right) - q\left(\theta^*_\varphi(\Sigma), \beta^\circ\right) \right| < \frac{\eta}{6} \right] \right\} -$$

$$\pi_\circ \left\{ \left[\left| q\left(\widehat{\theta}_{\varphi(T)S}(\Sigma), \widehat{\beta}_{\varphi(T)}\right) - q\left(\theta^*_\varphi(\Sigma), \beta^\circ\right) \right| < \frac{\eta}{6} \right] \ or \ \left[\left| q\left(\widehat{\theta}_{\varphi(T)S}(\Sigma), \beta^\circ\right) - q\left(\theta^*_\varphi(\Sigma), \beta^\circ\right) \right| < \frac{\eta}{6} \right] \right\},$$

$$\forall \varepsilon > 0, \ \forall \eta > 0, \ \exists T^2_{\varepsilon,\eta} \ / \ \forall T \geq T^2_{\varepsilon,\eta} : \Phi > \left(1 - \frac{\varepsilon}{5}\right) + \left(1 - \frac{\varepsilon}{5}\right) - 1 = 1 - \frac{2\varepsilon}{5} \cdot$$

Moreover:

$$\left\{ \left[\left| q\left(\widehat{\theta}_{\varphi(T)S}(\Sigma), \widehat{\beta}_{\varphi(T)}\right) - q\left(\theta^*_\varphi(\Sigma), \beta^\circ\right) \right| < \frac{\eta}{6} \right] \ and \ \left[\left| q\left(\widehat{\theta}_{\varphi(T)S}(\Sigma), \beta^\circ\right) - q\left(\theta^*_\varphi(\Sigma), \beta^\circ\right) \right| < \frac{\eta}{6} \right] \right\},$$

$$\Longrightarrow \quad \left\{ \left[\left| q\left(\widehat{\theta}_{\varphi(T)S}(\Sigma), \widehat{\beta}_{\varphi(T)}\right) - q\left(\widehat{\theta}_{\varphi(T)S}(\Sigma), \beta^\circ\right) \right| < \frac{\eta}{3} \right] \right\} \cdot$$

So that we have:

$$\forall \varepsilon > 0, \ \forall \eta > 0, \ \exists T^2_{\varepsilon,\eta} \ / \ \forall T \geq T^2_{\varepsilon,\eta} : \ \pi_\circ \left\{ \left[\left| q\left(\widehat{\theta}_{\varphi(T)S}(\Sigma), \widehat{\beta}_{\varphi(T)}\right) - q\left(\widehat{\theta}_{\varphi(T)S}(\Sigma), \beta^\circ\right) \right| < \frac{\eta}{3} \right] \right\} > 1 - \frac{2\varepsilon}{5},$$

$$\Longrightarrow \ \forall \varepsilon > 0, \ \forall \eta > 0, \ \exists T^2_{\varepsilon,\eta} \ / \ \forall T \geq T^2_{\varepsilon,\eta} : \ \pi_\circ \left\{ \left[q\left(\widehat{\theta}_{\varphi(T)S}(\Sigma), \beta^\circ\right) < q\left(\widehat{\theta}_{\varphi(T)S}(\Sigma), \widehat{\beta}_{\varphi(T)}\right) + \frac{\eta}{3} \right] \right\} > 1 - \frac{2\varepsilon}{5} \cdot$$

By using exactly the same kind of arguments, we also have:

$$\forall \varepsilon > 0, \ \forall \eta > 0, \ \exists T^3_{\varepsilon,\eta} \ / \ \forall T \geq T^3_{\varepsilon,\eta} : \pi_\circ \left\{ \left[q\left(\theta^\circ, \widehat{\beta}_{\varphi(T)}\right) < q\left(\theta^\circ, \beta^\circ\right) + \frac{\eta}{6} \right] \right\} > 1 - \frac{\varepsilon}{5} \cdot$$

We now define $T_{\varepsilon,\eta} = \underset{i=1,2,3.}{max} \left(T^i_{\varepsilon,\eta}\right)$ and we have: $\forall \varepsilon > 0, \ \forall \eta > 0, \ \exists T_{\varepsilon,\eta} \ / \ \forall T \geq T_{\varepsilon,\eta}$:

- $\pi_\circ \left\{ \left[q\left(\widehat{\theta}_{\varphi(T)S}(\Sigma), \widehat{\beta}_{\varphi(T)}\right) < q\left(\theta^\circ, \widehat{\beta}_{\varphi(T)}\right) + \frac{\eta}{2} \right] \right\} > 1 - \frac{2\varepsilon}{5},$
- $\pi_\circ \left\{ \left[q\left(\widehat{\theta}_{\varphi(T)S}(\Sigma), \beta^\circ\right) < q\left(\widehat{\theta}_{\varphi(T)S}(\Sigma), \widehat{\beta}_{\varphi(T)}\right) + \frac{\eta}{3} \right] \right\} > 1 - \frac{2\varepsilon}{5},$
- $\pi_\circ \left\{ \left[q\left(\theta^\circ, \widehat{\beta}_{\varphi(T)}\right) < q\left(\theta^\circ, \beta^\circ\right) + \frac{\eta}{6} \right] \right\} > 1 - \frac{\varepsilon}{5} \cdot$

$$\Longrightarrow \ \forall \varepsilon > 0, \ \forall \eta > 0, \ \exists T_{\varepsilon,\eta} \ / \ \forall T \geq T_{\varepsilon,\eta} : \ \pi_\circ \left\{ \left[q\left(\widehat{\theta}_{\varphi(T)S}(\Sigma), \beta^\circ\right) < q\left(\theta^\circ, \beta^\circ\right) + \eta \right] \right\} > 1 - \varepsilon \cdot$$

Let now $\mathcal{N}_{\theta^\circ}$ be any open subset of Θ containing θ° (we have assumed that $\theta^\circ \in \overset{\circ}{\Theta}$). $\Theta \cap \mathcal{N}^c_{\theta^\circ}$ is a compact set and we have $q\left(\theta^\circ, \beta^\circ\right) = 0$. We define θ^* by: $\theta^* = \underset{\theta \in \Theta \cap \mathcal{N}^c_{\theta^\circ}}{Argmin}\, q\left(\theta, \beta^\circ\right) \cdot$

We have: $q\left(\theta^*, \beta^\circ\right) > 0$ since θ° is the **unique** minimizer of $q\left(\cdot, \beta^\circ\right)$ and $\theta^* \in \Theta \cap \mathcal{N}^c_{\theta^\circ}$. ($\theta^\circ \in \mathcal{N}_{\theta^\circ}$). Let us define $\eta^*(\theta) = q\left(\theta^*, \beta^\circ\right) > 0$. For this particular value of η we have:

$$\forall \varepsilon > 0,\ \exists T_\varepsilon\ /\ \forall T \geq T_\varepsilon : \pi_\circ \left\{ \left[q\left(\widehat{\theta}_{\varphi(T)S}(\Sigma), \beta^\circ \right) < q\left(\theta^*, \beta^\circ\right) \right] \right\} > 1 - \varepsilon,$$

$$\Longrightarrow \quad \forall \varepsilon > 0,\ \exists T_\varepsilon\ /\ \forall T \geq T_\varepsilon : \pi_\circ \left\{ \widehat{\theta}_{\varphi(T)S}(\Sigma) \notin \Theta \cap \mathcal{N}^c_{\theta^\circ} \right\} > 1 - \varepsilon,$$

since $\theta^* = \underset{\theta \in \Theta \cap \mathcal{N}^c_{\theta^\circ}}{Argmin}\, q\left(\theta, \beta^\circ\right),$

Or in other words, with a probability approaching 1, $\widehat{\theta}_{\varphi(T)S}(\Sigma) \notin \Theta \cap \mathcal{N}^c_{\theta^\circ} \Longleftrightarrow \widehat{\theta}_{\varphi(T)S}(\Sigma) \in \mathcal{N}_{\theta^\circ}$ for each $\mathcal{N}_{\theta^\circ}$ open subset of Θ containing θ°. We have thus with probability approaching 1, $\widehat{\theta}_{\varphi(T)S}(\Sigma) \in \mathcal{N}_{\theta^\circ}$ i.e.:

$$\widehat{\theta}_{\varphi(T)S}(\Sigma) \underset{T \to +\infty}{\overset{\pi_\circ}{\longrightarrow}} \theta^\circ, \text{ that is: } \theta^*_\varphi(\Sigma) = \theta^\circ \cdot$$

A.4.1. Proof of proposition 4.4:

We start with the first order conditions associated with the minimization program (4.16):

$$\frac{\partial}{\partial \theta} \left[\widetilde{H}_{TS}\left(\theta, \widehat{\beta}_T\right) \right]'_{\theta = \widehat{\theta}_{TS}(\Sigma)} \widehat{\Sigma}_T \sqrt{T} \widetilde{H}_{TS}\left(\widehat{\theta}_{TS}(\Sigma), \widehat{\beta}_T \right) = 0 \cdot$$

Expanding the latter expression around the point $(\theta^\circ, \beta^\circ)$ we have:

$$\frac{\partial H'}{\partial \theta}\left(\theta^\circ, \beta^\circ\right) \Sigma \left[\sqrt{T} \widetilde{H}_{TS}\left(\theta^\circ, \beta^\circ\right) + \frac{\partial H}{\partial \theta'}\left(\theta^\circ, \beta^\circ\right) \sqrt{T}\left(\widehat{\theta}_{TS}(\Sigma) - \theta^\circ \right) + \frac{\partial H}{\partial \beta'}\left(\theta^\circ, \beta^\circ\right) \sqrt{T}\left(\widehat{\beta}_T - \beta^\circ \right) \right] + o_{\pi_\circ}(1) = 0 \cdot$$

$$\sqrt{T}\left[\widehat{\theta}_{TS}(\Sigma) - \theta^\circ \right] = \left[\frac{\partial H'}{\partial \theta}\left(\theta^\circ, \beta^\circ\right) \Sigma \frac{\partial H}{\partial \theta'}\left(\theta^\circ, \beta^\circ\right) \right]^{-1} \frac{\partial H'}{\partial \theta}\left(\theta^\circ, \beta^\circ\right) \Sigma \times$$
$$\left[\sqrt{T} \widetilde{H}_{TS}\left(\theta^\circ, \beta^\circ\right) + \frac{\partial H}{\partial \beta'}\left(\theta^\circ, \beta^\circ\right) \sqrt{T}\left(\widehat{\beta}_T - \beta^\circ \right) \right] + o_{\pi_\circ}(1) \cdot$$

$\left[\sqrt{T} \widetilde{H}_{TS}\left(\theta^\circ, \beta^\circ\right) + \frac{\partial H}{\partial \beta'}\left(\theta^\circ, \beta^\circ\right) \sqrt{T}\left(\widehat{\beta}_T - \beta^\circ \right) \right]$ is under assumptions $(A13) - (A16)$, $(A18) - (A20)$ asymptotically normal with an asymptotic zero mean and an asymptotic covariance matrix given by $\widetilde{\Phi}_\circ(S)$ and:

$$W_S(\Sigma) = \left[\frac{\partial H'}{\partial \theta}\left(\theta^\circ, \beta^\circ\right) \Sigma \frac{\partial H}{\partial \theta'}\left(\theta^\circ, \beta^\circ\right) \right]^{-1} \frac{\partial H'}{\partial \theta}\left(\theta^\circ, \beta^\circ\right) \Sigma \widetilde{\Phi}_\circ(S) \Sigma \frac{\partial H}{\partial \theta'}\left(\theta^\circ, \beta^\circ\right) \left[\frac{\partial H'}{\partial \theta}\left(\theta^\circ, \beta^\circ\right) \Sigma \frac{\partial H}{\partial \theta'}\left(\theta^\circ, \beta^\circ\right) \right]^{-1} \cdot$$

In the semiparametric setting $(A16)$, we have:

$$\begin{aligned}
\widetilde{\Phi}_\circ(S) &= \underset{\circ}{V}ar_{as} \left[\sqrt{T} \widetilde{H}_{TS}\left(\theta^\circ, \beta^\circ\right) + \frac{\partial H}{\partial \beta'}\left(\theta^\circ, \beta^\circ\right) \sqrt{T}\left(\widehat{\beta}_T - \beta^\circ \right) \right], \\
&= \underset{\circ}{V}ar_{as} \left[\sqrt{T} \widetilde{H}_{TS}\left(\theta^\circ, \beta^\circ\right) \right] + \frac{\partial H}{\partial \beta'}\left(\theta^\circ, \beta^\circ\right) \Omega^\circ \frac{\partial H'}{\partial \beta}\left(\theta^\circ, \beta^\circ\right) + \mathcal{L}_\circ + \mathcal{L}'_\circ,
\end{aligned}$$

where $\mathcal{L}_\circ = \lim\limits_{T \to +\infty} \underset{\circ}{C}ov \left\{ \sqrt{T} \widetilde{H}_{TS}\left(\theta^\circ, \beta^\circ\right), \frac{\partial H}{\partial \beta'}\left(\theta^\circ, \beta^\circ\right) \sqrt{T}\left(\widehat{\beta}_T - \beta^\circ \right) \right\}$. This general formula collapses in the fully parametric case $(A1)$, $(A17)$ to:

$$\widetilde{H}_{TS}\left(\theta^\circ, \beta^\circ\right) = \frac{1}{S} \sum_{s=1}^{S} \widetilde{H}^s_T\left(\underline{\widetilde{y}}^s_T\left(\theta^\circ, z^s_\circ\right), \underline{x}_T, \beta^\circ \right),$$

$$\underset{\circ}{V}ar_{as} \left[\sqrt{T} \widetilde{H}_{TS}\left(\theta^\circ, \beta^\circ\right) \right] = \frac{1}{S^2} \underset{\circ}{V}ar_{as} \left[\sum_{s=1}^{S} \sqrt{T} \widetilde{H}^s_T\left(\underline{\widetilde{y}}^s_T\left(\theta^\circ, z^s_\circ\right), \underline{x}_T, \beta^\circ \right) \right],$$

$$= \frac{1}{S^2} \left[S \underset{\circ}{V}ar_{as} \left[\sqrt{T} \widetilde{H}^s_T\left(\underline{\widetilde{y}}^s_T\left(\theta^\circ, z^s_\circ\right), \underline{x}_T, \beta^\circ \right) \right] + S(S-1) \underset{\circ}{C}ov_{as} \left[\sqrt{T} \widetilde{H}^s_T\left(\underline{\widetilde{y}}^s_T\left(\theta^\circ, z^s_\circ\right), \underline{x}_T, \beta^\circ \right), \right.\right.$$
$$\left.\left. \sqrt{T} \widetilde{H}^\ell_T\left(\underline{\widetilde{y}}^\ell_T\left(\theta^\circ, z^\ell_\circ\right), \underline{x}_T, \beta^\circ \right) \right] \right] = \frac{1}{S} I_\circ + \left(1 - \frac{1}{S}\right) K_\circ, \quad s \neq \ell \cdot$$

$$\mathcal{L}_\circ = \lim_{T\to+\infty} \underset{\circ}{Cov}\left\{\frac{1}{S}\sum_{s=1}^{S}\sqrt{T}\widetilde{H}_T^s\left(\underline{\widetilde{y}}_T^s(\theta^\circ,z_\circ^s),\underline{x}_T,\beta^\circ\right),\frac{\partial H}{\partial\beta'}(\theta^\circ,\beta^\circ)\sqrt{T}\left(\widehat{\beta}_T-\beta^\circ\right)\right\},$$

$$= \lim_{T\to+\infty} \underset{\circ}{Cov}\left\{\sqrt{T}\widetilde{H}_T^s\left(\underline{\widetilde{y}}_T^s(\theta^\circ,z_\circ^s),\underline{x}_T,\beta^\circ\right),\frac{\partial H}{\partial\beta'}(\theta^\circ,\beta^\circ)\sqrt{T}\left(\widehat{\beta}_T-\beta^\circ\right)\right\} = L_\circ\cdot$$

Therefore $\widetilde{\Phi}_\circ(S) = \frac{1}{S}I_\circ + \left(1-\frac{1}{S}\right)K_\circ + L_\circ + L_\circ' + \frac{\partial H}{\partial\beta'}(\theta^\circ,\beta^\circ)\,\Omega^\circ\frac{\partial H'}{\partial\beta}(\theta^\circ,\beta^\circ)$.

$$\widetilde{\widetilde{\Phi}}_\circ(S) = \underset{\circ}{V}ar_{as}\left[\sqrt{T}\widetilde{H}_{TS}(\theta^\circ,\beta^\circ)+\sqrt{T}\lambda\widetilde{\varepsilon}_{TS}(\theta^\circ,\beta^\circ)+\frac{\partial H}{\partial\beta'}(\theta^\circ,\beta^\circ)\sqrt{T}\left(\widehat{\beta}_T-\beta^\circ\right)\right],$$

$$= \underset{\circ}{V}ar_{as}\left[\sqrt{T}\widetilde{H}_{TS}(\theta^\circ,\beta^\circ)\right]+\lambda^2\underset{\circ}{V}ar_{as}\left[\sqrt{T}\widetilde{\varepsilon}_{TS}(\theta^\circ,\beta^\circ)\right]+\frac{\partial H}{\partial\beta'}(\theta^\circ,\beta^\circ)\,\Omega^\circ\frac{\partial H'}{\partial\beta}(\theta^\circ,\beta^\circ)+\mathcal{L}_\circ+\mathcal{L}_\circ',$$

since $\widetilde{\varepsilon}_{TS}(\theta,\beta) \perp \widetilde{H}_{TS}(\theta,\beta), \frac{\partial H}{\partial\beta'}(\theta^\circ,\beta^\circ)\sqrt{T}\left(\widehat{\beta}_T-\beta^\circ\right)$, therefore: $\widetilde{\widetilde{\Phi}}_\circ(S) = \widetilde{\Phi}_\circ(S)+\lambda^2\widetilde{\widetilde{\Psi}}_\circ(S,\varepsilon)$, with $\widetilde{\widetilde{\Psi}}_\circ(S,\varepsilon) = \underset{\circ}{V}ar_{as}\left[\sqrt{T}\widetilde{\varepsilon}_{TS}(\theta^\circ,\beta^\circ)\right]$.

In the parametric case, we have: $\widetilde{\widetilde{\Psi}}_\circ(S,\varepsilon) = \frac{1}{S}\widetilde{\widetilde{\Psi}}_\circ^1(\varepsilon)+\left(1-\frac{1}{S}\right)\widetilde{\widetilde{\Psi}}_\circ^2(\varepsilon)$, with $\widetilde{\widetilde{\Psi}}_\circ^1(\varepsilon) = \underset{\circ}{V}ar_{as}\left[\sqrt{T}\widetilde{\varepsilon}_T^s(\theta^\circ,\beta^\circ)\right]$ and $\widetilde{\widetilde{\Psi}}_\circ^2(\varepsilon) = \underset{\circ}{C}ov_{as}\left[\sqrt{T}\widetilde{\varepsilon}_T^s(\theta^\circ,\beta^\circ),\sqrt{T}\widetilde{\varepsilon}_T^\ell(\theta^\circ,\beta^\circ)\right]\ s\neq\ell$.

A.4.2. Proof of proposition 4.5:

We first prove that:

$$\left[\frac{\partial H'}{\partial\theta}(\theta^\circ,\beta^\circ)\,\Sigma\frac{\partial H}{\partial\theta'}(\theta^\circ,\beta^\circ)\right]^{-1}\frac{\partial H'}{\partial\theta}(\theta^\circ,\beta^\circ)\,\Sigma\widetilde{\Phi}_\circ(S)\Sigma\frac{\partial H}{\partial\theta'}(\theta^\circ,\beta^\circ)\times$$
$$\left[\frac{\partial H'}{\partial\theta}(\theta^\circ,\beta^\circ)\,\Sigma\frac{\partial H}{\partial\theta'}(\theta^\circ,\beta^\circ)\right]^{-1} >> \left[\frac{\partial H'}{\partial\theta}(\theta^\circ,\beta^\circ)\,\widetilde{\Phi}_\circ(S)^{-1}\frac{\partial H}{\partial\theta'}(\theta^\circ,\beta^\circ)\right]^{-1},$$

and then since $W_S\left(\widetilde{\Phi}_\circ(S)^{-1}\right) = \left[\frac{\partial H'}{\partial\theta}(\theta^\circ,\beta^\circ)\,\widetilde{\Phi}_\circ(S)^{-1}\frac{\partial H}{\partial\theta'}(\theta^\circ,\beta^\circ)\right]^{-1}$, we thus prove the result of proposition 4.5.

We first define:

$$A = \Sigma\widetilde{\Phi}_\circ(S)\Sigma,$$

$$B = A^{\frac{1}{2}}\frac{\partial H}{\partial\theta'}(\theta^\circ,\beta^\circ),$$

$$P = \frac{\partial H'}{\partial\theta}(\theta^\circ,\beta^\circ)\,\Sigma\frac{\partial H}{\partial\theta'}(\theta^\circ,\beta^\circ)\cdot$$

We know that:

$$I_r - B\left(B'B\right)^{-1}B' >> 0,$$

$$\Longrightarrow \quad A^{-1} >> A^{-\frac{1}{2}}B\left(B'B\right)^{-1}B'A^{-\frac{1}{2}},$$

$$\Longrightarrow \quad A^{-1} >> \frac{\partial H}{\partial \theta'}\left(\theta^\circ,\beta^\circ\right)\left[\frac{\partial H'}{\partial \theta}\left(\theta^\circ,\beta^\circ\right)A\frac{\partial H}{\partial \theta'}\left(\theta^\circ,\beta^\circ\right)\right]^{-1}\frac{\partial H'}{\partial \theta}\left(\theta^\circ,\beta^\circ\right),$$

$$\Longrightarrow \quad \Sigma^{-1}\widetilde{\Phi}_\circ(S)^{-1}\Sigma^{-1} - \frac{\partial H}{\partial \theta'}\left(\theta^\circ,\beta^\circ\right)\left[\frac{\partial H'}{\partial \theta}\left(\theta^\circ,\beta^\circ\right)\Sigma\widetilde{\Phi}_\circ(S)\Sigma\frac{\partial H}{\partial \theta'}\left(\theta^\circ,\beta^\circ\right)\right]^{-1}\frac{\partial H'}{\partial \theta}\left(\theta^\circ,\beta^\circ\right) >> 0,$$

$$\Longrightarrow \quad \frac{\partial H'}{\partial \theta}\left(\theta^\circ,\beta^\circ\right)\Sigma\left\{\Sigma^{-1}\widetilde{\Phi}_\circ(S)^{-1}\Sigma^{-1} - \frac{\partial H}{\partial \theta'}\left(\theta^\circ,\beta^\circ\right)\left[\frac{\partial H'}{\partial \theta}\left(\theta^\circ,\beta^\circ\right)\Sigma\widetilde{\Phi}_\circ(S)\Sigma\frac{\partial H}{\partial \theta'}\left(\theta^\circ,\beta^\circ\right)\right]^{-1}\right.$$

$$\left.\frac{\partial H'}{\partial \theta}\left(\theta^\circ,\beta^\circ\right)\right\}\Sigma\frac{\partial H}{\partial \theta'}\left(\theta^\circ,\beta^\circ\right) >> 0,$$

$$\Longrightarrow \quad \frac{\partial H'}{\partial \theta}\left(\theta^\circ,\beta^\circ\right)\widetilde{\Phi}_\circ(S)^{-1}\frac{\partial H}{\partial \theta'}\left(\theta^\circ,\beta^\circ\right) >> P\left[\frac{\partial H'}{\partial \theta}\left(\theta^\circ,\beta^\circ\right)\Sigma\widetilde{\Phi}_\circ(S)\Sigma\frac{\partial H}{\partial \theta'}\left(\theta^\circ,\beta^\circ\right)\right]^{-1}P,$$

$$\Longrightarrow \quad \left[\frac{\partial H'}{\partial \theta}\left(\theta^\circ,\beta^\circ\right)\widetilde{\Phi}_\circ(S)^{-1}\frac{\partial H}{\partial \theta'}\left(\theta^\circ,\beta^\circ\right)\right]^{-1} << P^{-1}\frac{\partial H'}{\partial \theta}\left(\theta^\circ,\beta^\circ\right)\Sigma\widetilde{\Phi}_\circ(S)\Sigma\frac{\partial H}{\partial \theta'}\left(\theta^\circ,\beta^\circ\right)P^{-1}.$$

A.4.3. Proof of Lemma 4.1:

Lemma 4.1:
Let $\{W_n, n \in I\!N\}$ a decreasing sequence of non negative (symmetric) matrices with respect to the partial order $<<$ over the square symmetric matrices set of order r, then there exists a unique symmetric non negative matrix W^* such that $\lim\limits_{n\to+\infty} W_n = W^*$.[36]

Proof : *Indeed, since $\{W_n, n \in I\!N\}$ is decreasing with respect to $<<$, we have:*

$$\forall n \in I\!N, \quad W_n << W_\circ,$$

$$\Longleftrightarrow \quad \forall n \in I\!N, \quad Sp\left(W_n\right) \leq Sp\left(W_\circ\right),$$

$$\Longrightarrow \quad \forall n \in I\!N, \quad \left\|W_n\right\|_1 \leq \left\|W_\circ\right\|_1,$$

where $\|W\|_1 = \underset{i=1,\ldots,r}{Sup}\ |\lambda_i|$.
We now consider $M = \{W \in \mathcal{M}_r\left(I\!R\right) / \left\|W_n\right\|_1 \leq \left\|W_\circ\right\|_1\}$ and where $\mathcal{M}_r\left(I\!R\right)$ is the vectorial space of symmetric matrices. M is a compact subspace of $\mathcal{M}_r\left(I\!R\right)$ and $\{W_n, n \in I\!N\} \in M$. Then, we just need to show that $card\{\mathcal{V}\left(W_n\right)\} \leq 1$.
Let φ_1 and φ_2 two increasing functions from $I\!N$ onto $I\!N$, $\lim\limits_{n\to+\infty}\varphi_1(n) = \lim\limits_{n\to+\infty}\varphi_2(n) = +\infty$ and such

[36] The norm is any norm on $I\!R^{\frac{r(r+1)}{2}}$ since in finite dimension, all norms are equivalent. We will focus here on $\|W\|_1 = \underset{i=1,\ldots,r}{Sup}\ |\lambda_i|$, where $\{\lambda_i, i = 1,\ldots,r\}$ correspond to the spectra of W.

that there exist $W^*_{\varphi_1}$ and $W^*_{\varphi_2}$ such that:

$$\lim_{n\to+\infty} W_{\varphi_1(n)} = W^*_{\varphi_1},$$

$$\lim_{n\to+\infty} W_{\varphi_2(n)} = W^*_{\varphi_2}.$$

We show that $W^*_{\varphi_1} = W^*_{\varphi_2}$ and this ends the proof of lemma 4.1.

$$\forall\varepsilon > 0,\ \exists\ n_\circ/\ \forall\ n \geq n_\circ,\quad \left\|W_{\varphi_1(n)} - W^*_{\varphi_1}\right\|_1 < \varepsilon,$$

$$\left\|W_{\varphi_2(n)} - W^*_{\varphi_2}\right\|_1 < \varepsilon.$$

We have $W^*_{\varphi_i} << W_{\varphi_i(n)},\ i = 1,2$ for $n \in I\!N$, since $Sp\left\{W_{\varphi_i(n)}\right\},\ i = 1,2$ is a decreasing subsequence of $I\!R_+{}^r$. Let now for each $n \in I\!N$ there exists $n' \in I\!N/\ \varphi_1(n') > \varphi_2(n)$ (since $\varphi_1(n)$ is increasing and $\lim_{n\to+\infty} \varphi_1(n) = +\infty$). Since $\{W_n, n \in I\!N\}$ is decreasing with respect to $<<$, we have $W_{\varphi_1(n')} << W_{\varphi_2(n)}$. We also know that $W^*_{\varphi_1} << W_{\varphi_1(n')}$, therefore:

$$\Longrightarrow\quad W^*_{\varphi_1} << W_{\varphi_2(n)},\ for\ n \in I\!N,$$

$$\Longrightarrow\quad W^*_{\varphi_1} << W^*_{\varphi_2},$$

again this is just a consequence of the limit of the spectra of $W_{\varphi_2(n)} - W^*_{\varphi_1}$.
Using exactly the same tricks and because of the symmetry in the proofs, we also have:

$$W^*_{\varphi_2} << W^*_{\varphi_1},$$

$$\Longrightarrow\quad W^*_{\varphi_1} = W^*_{\varphi_2}.$$

A.4.4. Proof of the efficiency bounds properties:

• We now show that the efficiency bound in the direction of Ψ does not depend on the sequence $\{\lambda_n, n \in I\!N\}$. Let $\{\lambda_n, n \in I\!N\}$ and $\{\mu_n, n \in I\!N\}$ two decreasing sequences such that $\lim_{n\to+\infty} \lambda_n = \lim_{n\to+\infty} \mu_n = 0$. We denote $B^*(A,\Phi,\Psi,\lambda)$ and $B^*(A,\Phi,\Psi,\mu)$ the two associated efficiency bounds and we show that $B^*(A,\Phi,\Psi,\lambda) = B^*(A,\Phi,\Psi,\mu)$.
Indeed:

$$B^*(A,\Phi,\Psi,\lambda) = \lim_{n\to+\infty} W\left(A, \widetilde{\Phi}(\Phi,\Psi,\lambda_n), \widetilde{\Phi}(\Phi,\Psi,\lambda_n)^{-1}\right),$$

$$B^*(A,\Phi,\Psi,\mu) = \lim_{n\to+\infty} W\left(A, \widetilde{\Phi}(\Phi,\Psi,\mu_n), \widetilde{\Phi}(\Phi,\Psi,\mu_n)^{-1}\right),$$

$$W\left(A, \widetilde{\Phi}(\Phi,\Psi,\lambda_n), \widetilde{\Phi}(\Phi,\Psi,\lambda_n)^{-1}\right) = \left[A'\widetilde{\Phi}(\Phi,\Psi,\lambda_n)^{-1}A\right]^{-1},$$

$$W\left(A, \widetilde{\Phi}(\Phi,\Psi,\mu_n), \widetilde{\Phi}(\Phi,\Psi,\mu_n)^{-1}\right) = \left[A'\widetilde{\Phi}(\Phi,\Psi,\mu_n)^{-1}A\right]^{-1}.$$

$\forall\ n \in I\!N\ \exists\ \varphi(n) \in I\!N/\ \mu_{\varphi(n)} < \lambda_n$. Therefore $\forall\ n \in I\!N,\ \widetilde{\Phi}\left(\Phi,\Psi,\mu_{\varphi(n)}\right) << \widetilde{\Phi}(\Phi,\Psi,\lambda_n)$.

$$\Longrightarrow \forall\ n \in I\!N,\ W\left(A, \widetilde{\Phi}\left(\Phi,\Psi,\mu_{\varphi(n)}\right), \widetilde{\Phi}\left(\Phi,\Psi,\mu_{\varphi(n)}\right)^{-1}\right) << W\left(A, \widetilde{\Phi}(\Phi,\Psi,\lambda_n), \widetilde{\Phi}(\Phi,\Psi,\lambda_n)^{-1}\right).$$

We thus obtain at the limit or equivalently by analyzing the spectra that:

$$\lim_{n\to+\infty} W\left(A,\widetilde{\Phi}\left(\Phi,\Psi,\mu_{\varphi(n)}\right),\widetilde{\Phi}\left(\Phi,\Psi,\mu_{\varphi(n)}\right)^{-1}\right) << \lim_{n\to+\infty} W\left(A,\widetilde{\Phi}\left(\Phi,\Psi,\lambda_n\right),\widetilde{\Phi}\left(\Phi,\Psi,\lambda_n\right)^{-1}\right),$$

$$\Longrightarrow B^*\left(A,\Phi,\Psi,\mu\right) << B^*\left(A,\Phi,\Psi,\lambda\right)\cdot$$

Symmetrically it is easy to prove that: $B^*\left(A,\Phi,\Psi,\lambda\right) << B^*\left(A,\Phi,\Psi,\mu\right)$ and therefore $B^*\left(A,\Phi,\Psi,\mu\right) = B^*\left(A,\Phi,\Psi,\lambda\right)$.

• We now show that $B^*\left(A,\Phi,\lambda\Psi\right) = B^*\left(A,\Phi,\Psi\right)$ for all $\lambda > 0$. Indeed endowed with the previous notations we have $B^*\left(A,\Phi,\lambda\Psi\right) = B^*\left(A,\Phi,\Psi,\mu\right)$, with $\mu_n = \lambda\lambda_n$, therefore the result is straightforward.

• Let now Ψ_1 and Ψ_2 two non negative (symmetric) matrices. Suppose that $B^*\left(A,\Phi,\Psi_1\right) << B^*\left(A,\Phi,\Psi_2\right)$ and where the inequality is a strict one. Then $\forall\ \lambda > 0, \mu > 0$ $B^*\left(A,\Phi,\lambda\Psi_1\right) << B^*\left(A,\Phi,\mu\Psi_2\right)$ since $B^*\left(A,\Phi,\lambda\Psi\right) = B^*\left(A,\Phi,\Psi\right)$. This implies that:

$$\begin{aligned}&\exists\ n_\circ/\ \forall\ n \geq n_\circ,\ \forall\ \lambda > 0, \mu > 0,\\&W\left(A,\widetilde{\Phi}\left(\Phi,\lambda\Psi_1,\lambda_n\right),\widetilde{\Phi}\left(\Phi,\lambda\Psi_1,\lambda_n\right)^{-1}\right) << W\left(A,\widetilde{\Phi}\left(\Phi,\mu\Psi_2,\lambda_n\right),\widetilde{\Phi}\left(\Phi,\mu\Psi_2,\lambda_n\right)^{-1}\right)\cdot\end{aligned}$$

It is crucial to note that this is **only** true because $B^*\left(A,\Phi,\Psi_1\right) << B^*\left(A,\Phi,\Psi_2\right)$ with a **strict** inequality. Therefore:

$$\begin{aligned}&\exists\ n_\circ/\ \forall\ n \geq n_\circ,\ \forall\ \lambda > 0, \mu > 0, && \left[A'\widetilde{\Phi}\left(\Phi,\lambda\Psi_1,\lambda_n\right)^{-1}A\right]^{-1} << \left[A'\widetilde{\Phi}\left(\Phi,\mu\Psi_2,\lambda_n\right)^{-1}A\right]^{-1},\\ \Longrightarrow\ &\exists\ n_\circ/\ \forall\ n \geq n_\circ,\ \forall\ \lambda > 0, \mu > 0, && \widetilde{\Phi}\left(\Phi,\lambda\Psi_1,\lambda_n\right) << \widetilde{\Phi}\left(\Phi,\mu\Psi_2,\lambda_n\right),\\ \Longrightarrow\ &\exists\ n_\circ/\ \forall\ n \geq n_\circ,\ \forall\ \lambda > 0, \mu > 0, && \lambda\Psi_1 << \mu\Psi_2,\\ \Longrightarrow\ &\forall\ \lambda > 0, \mu > 0, && \lambda\Psi_1 << \mu\Psi_2,\\ \Longrightarrow\ & && \Psi_1 = 0,\end{aligned}$$

which is ruled out here.

Therefore we cannot have $B^*\left(A,\Phi,\Psi_{-i}\right) << B^*\left(A,\Phi,\Psi_i\right),\ i = 1,\ or\ 2$, and where the inequality is strict. Consequently for all Ψ_1 and Ψ_2 non negative matrices either $B^*\left(A,\Phi,\Psi_1\right) = B^*\left(A,\Phi,\Psi_2\right)$ or $B^*\left(A,\Phi,\Psi_1\right)$ and $B^*\left(A,\Phi,\Psi_2\right)$ are not comparable.

We now assume that there exist $\lambda_1 > 0$ and $\lambda_2 > 0$ such that $\lambda_1\Psi_2 << \Psi_1 << \lambda_2\Psi_2$ then $B^*\left(A,\Phi,\Psi_1\right) = B^*\left(A,\Phi,\Psi_2\right)$. Indeed:

$$\begin{aligned}& \lambda_1\Psi_2 << \Psi_1,\\ \Longrightarrow\ & B^*\left(A,\Phi,\lambda\Psi_2\right) << B^*\left(A,\Phi,\Psi_1\right),\\ \Longrightarrow\ & B^*\left(A,\Phi,\Psi_2\right) << B^*\left(A,\Phi,\Psi_1\right)\cdot\end{aligned}$$

For exactly the same arguments, we have $B^*\left(A,\Phi,\Psi_1\right) << B^*\left(A,\Phi,\Psi_2\right)$ and therefore $B^*\left(A,\Phi,\Psi_1\right) = B^*\left(A,\Phi,\Psi_2\right)$.

A.5. Proof of proposition 5.1:

$$\xi_{TS}^{SALS} = \left[\frac{1}{S}\sqrt{T}\sum_{s=1}^{S}\widetilde{H}_T\left(\underline{\widetilde{y}}_T^s(\widehat{\theta}_{TS}^{SALS^*}, z_\circ^s), \underline{x}_T, \widehat{\beta}_T\right)\right]'\widehat{\Sigma}_T^*\left[\frac{1}{S}\sum_{s=1}^{S}\sqrt{T}\widetilde{H}_T\left(\underline{\widetilde{y}}_T^s(\widehat{\theta}_{TS}^{SALS^*}, z_\circ^s), \underline{x}_T, \widehat{\beta}_T\right)\right],$$

where $\pi_\circ \lim_{T\to+\infty} \widehat{\Sigma}_T^* = \widetilde{\Phi}_\circ(S)^{-1}$.

$$\frac{1}{S}\sum_{s=1}^{S}\sqrt{T}\widetilde{H}_T\left(\underline{\widetilde{y}}_T^s(\widehat{\theta}_{TS}^{SALS^*}, z_\circ^s), \underline{x}_T, \widehat{\beta}_T\right) =$$

$$= \frac{1}{S}\sum_{s=1}^{S}\sqrt{T}\widetilde{H}_T\left(\underline{\widetilde{y}}_T^s(\theta^\circ, z_\circ^s), \underline{x}_T, \beta^\circ\right) + \frac{\partial H}{\partial \theta'}(\theta^\circ,\beta^\circ)\sqrt{T}\left(\widehat{\theta}_{TS}^{SALS^*} - \theta^\circ\right) +$$

$$\frac{\partial H}{\partial \beta'}(\theta^\circ,\beta^\circ)\sqrt{T}\left(\widehat{\beta}_T - \beta^\circ\right) + o_{\pi_\circ}(1),$$

$$= \frac{1}{S}\sum_{s=1}^{S}\sqrt{T}\widetilde{H}_T\left(\underline{\widetilde{y}}_T^s(\theta^\circ, z_\circ^s), \underline{x}_T, \beta^\circ\right) - \frac{\partial H}{\partial \theta'}(\theta^\circ,\beta^\circ)\left[\frac{\partial H'}{\partial \theta}(\theta^\circ,\beta^\circ)\widetilde{\Phi}_\circ(S)^{-1}\frac{\partial H}{\partial \theta'}(\theta^\circ,\beta^\circ)\right]^{-1}\times$$

$$\frac{\partial H'}{\partial \theta}(\theta^\circ,\beta^\circ)\widetilde{\Phi}_\circ(S)^{-1}\left[\frac{1}{S}\sum_{s=1}^{S}\sqrt{T}\widetilde{H}_T\left(\underline{\widetilde{y}}_T^s(\theta^\circ, z_\circ^s), \underline{x}_T, \beta^\circ\right) + \frac{\partial H}{\partial \beta'}(\theta^\circ,\beta^\circ)\sqrt{T}\left(\widehat{\beta}_T - \beta^\circ\right)\right] +$$

$$\frac{\partial H}{\partial \beta'}(\theta^\circ,\beta^\circ)\sqrt{T}\left(\widehat{\beta}_T - \beta^\circ\right) + o_{\pi_\circ}(1),$$

$$= \left\{I_r - \frac{\partial H}{\partial \theta'}(\theta^\circ,\beta^\circ)\left[\frac{\partial H'}{\partial \theta}(\theta^\circ,\beta^\circ)\widetilde{\Phi}_\circ(S)^{-1}\frac{\partial H}{\partial \theta'}(\theta^\circ,\beta^\circ)\right]^{-1}\frac{\partial H'}{\partial \theta}(\theta^\circ,\beta^\circ)\widetilde{\Phi}_\circ(S)^{-1}\right\}\times$$

$$\left[\frac{1}{S}\sum_{s=1}^{S}\sqrt{T}\widetilde{H}_T\left(\underline{\widetilde{y}}_T^s(\theta^\circ, z_\circ^s), \underline{x}_T, \beta^\circ\right) + \frac{\partial H}{\partial \beta'}(\theta^\circ,\beta^\circ)\sqrt{T}\left(\widehat{\beta}_T - \beta^\circ\right)\right] + o_{\pi_\circ}(1)\cdot$$

We denote:
$\widehat{A}_{TS} = \widetilde{\Phi}_\circ(S)^{-\frac{1}{2}}\left[\frac{1}{S}\sum_{s=1}^{S}\sqrt{T}\widetilde{H}_T\left(\underline{\widetilde{y}}_T^s(\theta^\circ, z_\circ^s), \underline{x}_T, \beta^\circ\right) + \frac{\partial H}{\partial \beta'}(\theta^\circ,\beta^\circ)\sqrt{T}\left(\widehat{\beta}_T - \beta^\circ\right)\right]$. We have $\widehat{A}_{TS} \xrightarrow[T\to+\infty]{D} \mathcal{N}(0, I_r)$.

$$\frac{1}{S}\sum_{s=1}^{S}\sqrt{T}\widetilde{H}_T\left(\underline{\widetilde{y}}_T^s(\widehat{\theta}_{TS}^{SALS^*}, z_\circ^s), \underline{x}_T, \widehat{\beta}_T\right)$$

$$= \widetilde{\Phi}_\circ(S)^{\frac{1}{2}}\left\{I_r - \widetilde{\Phi}_\circ(S)^{-\frac{1}{2}}\frac{\partial H}{\partial \theta'}(\theta^\circ,\beta^\circ)\left[\frac{\partial H'}{\partial \theta}(\theta^\circ,\beta^\circ)\widetilde{\Phi}_\circ(S)^{-1}\frac{\partial H}{\partial \theta'}(\theta^\circ,\beta^\circ)\right]^{-1}\times\right.$$

$$\left.\frac{\partial H'}{\partial \theta}(\theta^\circ,\beta^\circ)\widetilde{\Phi}_\circ(S)^{-\frac{1}{2}}\right\}\widehat{A}_{TS} + o_{\pi_\circ}(1)\cdot$$

We denote:

$$Q = \widetilde{\Phi}_\circ(S)^{-\frac{1}{2}}\frac{\partial H}{\partial \theta'}(\theta^\circ,\beta^\circ)\left[\frac{\partial H'}{\partial \theta}(\theta^\circ,\beta^\circ)\widetilde{\Phi}_\circ(S)^{-1}\frac{\partial H}{\partial \theta'}(\theta^\circ,\beta^\circ)\right]^{-1}\frac{\partial H'}{\partial \theta}(\theta^\circ,\beta^\circ)\widetilde{\Phi}_\circ(S)^{-\frac{1}{2}}.$$

Q is the orthogonal projector on the orthogonal space spanned by the columns of $\widetilde{\Phi}_\circ(S)^{-\frac{1}{2}}\frac{\partial H}{\partial \theta'}(\theta^\circ,\beta^\circ)$.
Thus we have:

$$\xi_{TS}^{SALS} = \left[\widetilde{\Phi}_\circ(S)^{\frac{1}{2}}\left(I_r - Q\right)\widehat{A}_{TS}\right]'\widetilde{\Phi}_\circ(S)^{-1}\widetilde{\Phi}_\circ(S)^{\frac{1}{2}}\left(I_r - Q\right)\widehat{A}_{TS} + o_{\pi_\circ}(1),$$

$$= \widehat{A}'_{TS}\left(I_r - Q\right)\widehat{A}_{TS} + o_{\pi_\circ}(1),$$

this proves the result of proposition 5.1.

A.6. Proof of proposition 5.2:

We start with the statistics $\xi_{TS}^i,\ i = 2,3,4$:

- $\xi_{TS}^i = \dfrac{1}{1+\dfrac{1}{S}}\widehat{A}_{TS}^{i'}\widehat{\Sigma}_T\widehat{A}_{TS}^i, \quad i = 2,3,4\cdot$

$$\widehat{A}_{TS}^2 = \frac{1}{TS}\sum_{s=1}^{S}\sum_{t=1}^{T}\frac{\partial g'}{\partial \beta}\left(\widetilde{w}_t^s(\widehat{\theta}_{TS}^{2^*}),\widehat{\beta}_T\right)\widehat{\Lambda}_T^*\frac{1}{\sqrt{TS}}\sum_{s=1}^{S}\sum_{t=1}^{T}g\left(\widetilde{w}_t^s(\widehat{\theta}_{TS}^{2^*}),\widehat{\beta}_T\right),$$

$$\widehat{A}_{TS}^3 = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial g'}{\partial \beta}\left(w_t,\widehat{\beta}_T\right)\widehat{\Lambda}_T^*\frac{1}{\sqrt{TS}}\sum_{s=1}^{S}\sum_{t=1}^{T}g\left(\widetilde{w}_t^s(\widehat{\theta}_{TS}^{3^*}),\widehat{\beta}_T\right),$$

$$\widehat{A}_{TS}^4 = \frac{1}{S}\sum_{s=1}^{S}\left[\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial g'}{\partial \beta}\left(\widetilde{w}_t^s(\widehat{\theta}_{TS}^{4^*}),\widehat{\beta}_T\right)\right)\widehat{\Lambda}_T^*\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}g\left(\widetilde{w}_t^s(\widehat{\theta}_{TS}^{4^*}),\widehat{\beta}_T\right)\right)\right]\cdot$$

The expansion of $\widehat{A}_{TS}^i,\ i = 2,3,4$ around the point $(\theta^\circ,\beta^\circ)$ gives:

$$\widehat{A}_{TS}^i = \underset{\circ}{E}\left[\frac{\partial g'}{\partial \beta}\left(w_t,\beta^\circ\right)\right]\Lambda^*\left\{\frac{1}{\sqrt{TS}}\sum_{s=1}^{S}\sum_{t=1}^{T}g\left(\widetilde{w}_t^s(\theta^\circ),\beta^\circ\right) + \right.$$
$$\frac{1}{\sqrt{TS}}\frac{\partial}{\partial \theta'}\left[\sum_{s=1}^{S}\sum_{t=1}^{T}g\left(\widetilde{w}_t^s(\theta),\beta^\circ\right)\right]_{\theta=\theta^\circ}\left(\widehat{\theta}_{TS}^{i^*} - \theta^\circ\right) +$$
$$\left.\frac{1}{\sqrt{TS}}\sum_{s=1}^{S}\sum_{t=1}^{T}\frac{\partial g}{\partial \beta'}\left(\widetilde{w}_t^s(\theta^\circ),\beta^\circ\right)\left(\widehat{\beta}_T - \beta^\circ\right)\right\} + o_{\pi_\circ}^i(1), \quad i = 2,3,4,$$

$$\widehat{A}_{TS}^i = \underset{\circ}{E}\left[\frac{\partial g'}{\partial \beta}\left(w_t,\beta^\circ\right)\right]\Lambda^*\left\{\frac{1}{\sqrt{TS}}\sum_{s=1}^{S}\sum_{t=1}^{T}g\left(\widetilde{w}_t^s(\theta^\circ),\beta^\circ\right) + \right.$$
$$\frac{1}{TS}\frac{\partial}{\partial \theta'}\left[\sum_{s=1}^{S}\sum_{t=1}^{T}g\left(\widetilde{w}_t^s(\theta),\beta^\circ\right)\right]_{\theta=\theta^\circ}\sqrt{T}\left(\widehat{\theta}_{TS}^{i^*} - \theta^\circ\right) +$$
$$\left.\frac{1}{TS}\sum_{s=1}^{S}\sum_{t=1}^{T}\frac{\partial g}{\partial \beta'}\left(\widetilde{w}_t^s(\theta^\circ),\beta^\circ\right)\sqrt{T}\left(\widehat{\beta}_T - \beta^\circ\right)\right\} + o_{\pi_\circ}^i(1), \quad i = 2,3,4,$$

- $\widehat{A}_{TS}^i = \dfrac{\partial z'}{\partial \beta}\left(\theta^\circ,\beta^\circ\right)\Lambda^*\left\{\dfrac{1}{\sqrt{TS}}\sum_{s=1}^{S}\sum_{t=1}^{T}g\left(\widetilde{w}_t^s(\theta^\circ),\beta^\circ\right) + \dfrac{\partial z}{\partial \theta'}\left(\theta^\circ,\beta^\circ\right)\sqrt{T}\left(\widehat{\theta}_{TS}^{i^*} - \theta^\circ\right) + \right.$
$\left.\dfrac{\partial z}{\partial \beta'}\left(\theta^\circ,\beta^\circ\right)\sqrt{T}\left(\widehat{\beta}_T - \beta^\circ\right)\right\} + o_{\pi_\circ}^i(1), \quad i = 2,3,4\cdot$

Besides we have:

$$\widehat{\beta}_T = \underset{\beta\in\mathcal{B}}{Argmin}\left[\frac{1}{T}\sum_{t=1}^{T}g\left(w_t,\beta\right)\right]'\widehat{\Lambda}_T^*\left[\frac{1}{T}\sum_{t=1}^{T}g\left(w_t,\beta\right)\right]\cdot$$

The first order conditions are:

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial g'}{\partial\beta}\left(w_t,\widehat{\beta}_T\right)\widehat{\Lambda}_T^{*}\frac{1}{T}\sum_{t=1}^{T}g\left(w_t,\widehat{\beta}_T\right)=0,$$

or equivalently:

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial g'}{\partial\beta}\left(w_t,\widehat{\beta}_T\right)\widehat{\Lambda}_T^{*}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}g\left(w_t,\widehat{\beta}_T\right)=0\cdot$$

Expanding the latter expression around the point β°, we have:

$$\underset{\circ}{E}\left[\frac{\partial g'}{\partial\beta}\left(w_t,\beta^{\circ}\right)\right]\Lambda^{*}\frac{1}{\sqrt{T}}\left[\sum_{t=1}^{T}g\left(w_t,\beta^{\circ}\right)+\sum_{t=1}^{T}\frac{\partial g}{\partial\beta'}\left(w_t,\beta^{\circ}\right)\left(\widehat{\beta}_T-\beta^{\circ}\right)\right]+o_{\pi_\circ}\left(1\right)=0,$$

$$\underset{\circ}{E}\left[\frac{\partial g'}{\partial\beta}\left(w_t,\beta^{\circ}\right)\right]\Lambda^{*}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}g\left(w_t,\beta^{\circ}\right)+$$
$$\underset{\circ}{E}\left[\frac{\partial g'}{\partial\beta}\left(w_t,\beta^{\circ}\right)\right]\Lambda^{*}\underset{\circ}{E}\left[\frac{\partial g}{\partial\beta'}\left(w_t,\beta^{\circ}\right)\right]\sqrt{T}\left(\widehat{\beta}_T-\beta^{\circ}\right)+o_{\pi_\circ}\left(1\right)=0,$$

$$\sqrt{T}\left(\widehat{\beta}_T-\beta^{\circ}\right)=-\left[\frac{\partial z'}{\partial\beta}\left(\theta^{\circ},\beta^{\circ}\right)\Lambda^{*}\frac{\partial z}{\partial\beta'}\left(\theta^{\circ},\beta^{\circ}\right)\right]^{-1}\frac{\partial z'}{\partial\beta}\left(\theta^{\circ},\beta^{\circ}\right)\Lambda^{*}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}g\left(w_t,\beta^{\circ}\right)+o_{\pi_\circ}\left(1\right),$$

- $\sqrt{T}\left(\widehat{\beta}_T-\beta^{\circ}\right)=-J_\circ\left(\Lambda^{*}\right)^{-1}\frac{\partial z'}{\partial\beta}\left(\theta^{\circ},\beta^{\circ}\right)\Lambda^{*}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}g\left(w_t,\beta^{\circ}\right)+o_{\pi_\circ}\left(1\right)\cdot$

Using the same kind of expansions we have:

- $\sqrt{T}\left(\widetilde{\beta}_T^{s}\left(\theta^{\circ}\right)-\beta^{\circ}\right)=-J_\circ\left(\Lambda^{*}\right)^{-1}\frac{\partial z'}{\partial\beta}\left(\theta^{\circ},\beta^{\circ}\right)\Lambda^{*}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}g\left(\widetilde{w}_t^{s}\left(\theta^{\circ}\right),\beta^{\circ}\right)+o_{\pi_\circ}\left(1\right),$

  and thus:
- $\sqrt{T}\left(\widetilde{\beta}_{TS}\left(\theta^{\circ}\right)-\beta^{\circ}\right)=-J_\circ\left(\Lambda^{*}\right)^{-1}\frac{\partial z'}{\partial\beta}\left(\theta^{\circ},\beta^{\circ}\right)\Lambda^{*}\frac{1}{\sqrt{T}S}\sum_{s=1}^{S}\sum_{t=1}^{T}g\left(\widetilde{w}_t^{s}\left(\theta^{\circ}\right),\beta^{\circ}\right)+o_{\pi_\circ}\left(1\right),$

  this implies that:
- $\sqrt{T}\left(\widetilde{\beta}_{TS}\left(\theta^{\circ}\right)-\widehat{\beta}_T\right)=-J_\circ\left(\Lambda^{*}\right)^{-1}\frac{\partial z'}{\partial\beta}\left(\theta^{\circ},\beta^{\circ}\right)\Lambda^{*}\frac{1}{\sqrt{T}S}\times$
  $\sum_{s=1}^{S}\sum_{t=1}^{T}\left[g\left(\widetilde{w}_t^{s}\left(\theta^{\circ}\right),\beta^{\circ}\right)-g\left(w_t,\beta^{\circ}\right)\right]+o_{\pi_\circ}\left(1\right)\cdot$

Moreover we have that for $i=2,3,4$:

$$\sqrt{T}\left(\widehat{\theta}_{TS}^{i*}-\theta^{\circ}\right)=-\left[\frac{\partial\widetilde{\beta}'}{\partial\theta}\left(\theta^{\circ},\Lambda^{*}\right)\Omega^{*}\frac{\partial\widetilde{\beta}}{\partial\theta'}\left(\theta^{\circ},\Lambda^{*}\right)\right]^{-1}\frac{\partial\widetilde{\beta}'}{\partial\theta}\left(\theta^{\circ},\Lambda^{*}\right)\Omega^{*}\sqrt{T}\left(\widetilde{\beta}_{TS}\left(\theta^{\circ}\right)-\widehat{\beta}_T\right)+o_{\pi_\circ}^{i}\left(1\right),$$

with:

$$\Omega^{*}=J_\circ\left(\Lambda^{*}\right)\Phi_\circ\left(\Lambda^{*}\right)^{-1}J_\circ\left(\Lambda^{*}\right),$$

$$\frac{\partial\widetilde{\beta}}{\partial\theta'}\left(\theta^{\circ},\Lambda^{*}\right)=-J_\circ\left(\Lambda^{*}\right)^{-1}\frac{\partial z'}{\partial\beta}\left(\theta^{\circ},\beta^{\circ}\right)\Lambda^{*}\frac{\partial z}{\partial\theta'}\left(\theta^{\circ},\beta^{\circ}\right)\cdot$$

Indeed it is easy to show that the three estimators are asymptotically equivalent and that their asymptotic expansion corresponds to the latter one. Here we have privileged the GII expression which is actually the common one by virtue of asymptotic equivalence.

Plugging the different relations into the expression of $\sqrt{T}\left(\widehat{\theta}_{TS}^{i*}-\theta^{\circ}\right)$, $i=2,3,4$, we obtain:

$$\sqrt{T}\left(\widehat{\theta}_{TS}^{i*}-\theta^{\circ}\right)=-\left[\frac{\partial z'}{\partial\theta}\left(\theta^{\circ},\beta^{\circ}\right)\Lambda^{*}\frac{\partial z}{\partial\beta'}\left(\theta^{\circ},\beta^{\circ}\right)\Phi_{\circ}\left(\Lambda^{*}\right)^{-1}\frac{\partial z'}{\partial\beta}\left(\theta^{\circ},\beta^{\circ}\right)\Lambda^{*}\frac{\partial z}{\partial\theta'}\left(\theta^{\circ},\beta^{\circ}\right)\right]^{-1}\times$$
$$\frac{\partial z'}{\partial\theta}\left(\theta^{\circ},\beta^{\circ}\right)\Lambda^{*}\frac{\partial z}{\partial\beta'}\left(\theta^{\circ},\beta^{\circ}\right)\Phi_{\circ}\left(\Lambda^{*}\right)^{-1}\frac{\partial z'}{\partial\beta}\left(\theta^{\circ},\beta^{\circ}\right)\Lambda^{*}\frac{1}{\sqrt{TS}}\sum_{s=1}^{S}\sum_{t=1}^{T}\left[g\left(\widetilde{w}_{t}^{s}\left(\theta^{\circ}\right),\beta^{\circ}\right)-g\left(w_{t},\beta^{\circ}\right)\right]$$

$$+o_{\pi_{\circ}}^{i}\left(1\right),\quad i=2,3,4\cdot$$

We can now write the expression of $\widehat{A}_{TS}^{i}$, $i=2,3,4$:

$$\widehat{A}_{TS}^{i}=\left\{I_{q}-\frac{\partial z'}{\partial\beta}\left(\theta^{\circ},\beta^{\circ}\right)\Lambda^{*}\frac{\partial z}{\partial\theta'}\left(\theta^{\circ},\beta^{\circ}\right)\left[\frac{\partial z'}{\partial\theta}\left(\theta^{\circ},\beta^{\circ}\right)\Lambda^{*}\frac{\partial z}{\partial\beta'}\left(\theta^{\circ},\beta^{\circ}\right)\Phi_{\circ}\left(\Lambda^{*}\right)^{-1}\times\right.\right.$$
$$\left.\left.\frac{\partial z'}{\partial\beta}\left(\theta^{\circ},\beta^{\circ}\right)\Lambda^{*}\frac{\partial z}{\partial\theta'}\left(\theta^{\circ},\beta^{\circ}\right)\right]^{-1}\frac{\partial z'}{\partial\theta}\left(\theta^{\circ},\beta^{\circ}\right)\Lambda^{*}\frac{\partial z}{\partial\beta'}\left(\theta^{\circ},\beta^{\circ}\right)\Phi_{\circ}\left(\Lambda^{*}\right)^{-1}\right\}$$
$$\frac{\partial z'}{\partial\beta}\left(\theta^{\circ},\beta^{\circ}\right)\Lambda^{*}\frac{1}{\sqrt{TS}}\sum_{s=1}^{S}\sum_{t=1}^{T}\left[g\left(\widetilde{w}_{t}^{s}\left(\theta^{\circ}\right),\beta^{\circ}\right)-g\left(w_{t},\beta^{\circ}\right)\right]+o_{\pi_{\circ}}^{i}\left(1\right),\quad i=2,3,4,$$

$$\widehat{A}_{TS}^{i}=\Phi_{\circ}\left(\Lambda^{*}\right)^{\frac{1}{2}}\left\{I_{q}-\Phi_{\circ}\left(\Lambda^{*}\right)^{-\frac{1}{2}}\frac{\partial z'}{\partial\beta}\left(\theta^{\circ},\beta^{\circ}\right)\Lambda^{*}\frac{\partial z}{\partial\theta'}\left(\theta^{\circ},\beta^{\circ}\right)\left[\frac{\partial z'}{\partial\theta}\left(\theta^{\circ},\beta^{\circ}\right)\Lambda^{*}\frac{\partial z}{\partial\beta'}\left(\theta^{\circ},\beta^{\circ}\right)\times\right.\right.$$
$$\left.\left.\Phi_{\circ}\left(\Lambda^{*}\right)^{-1}\frac{\partial z'}{\partial\beta}\left(\theta^{\circ},\beta^{\circ}\right)\Lambda^{*}\frac{\partial z}{\partial\theta'}\left(\theta^{\circ},\beta^{\circ}\right)\right]^{-1}\frac{\partial z'}{\partial\theta}\left(\theta^{\circ},\beta^{\circ}\right)\Lambda^{*}\frac{\partial z}{\partial\beta'}\left(\theta^{\circ},\beta^{\circ}\right)\Phi_{\circ}\left(\Lambda^{*}\right)^{-\frac{1}{2}}\right\}\Phi_{\circ}\left(\Lambda^{*}\right)^{-\frac{1}{2}}\times$$
$$\frac{\partial z'}{\partial\beta}\left(\theta^{\circ},\beta^{\circ}\right)\Lambda^{*}\frac{1}{\sqrt{TS}}\sum_{s=1}^{S}\sum_{t=1}^{T}\left[g\left(\widetilde{w}_{t}^{s}\left(\theta^{\circ}\right),\beta^{\circ}\right)-g\left(w_{t},\beta^{\circ}\right)\right]+o_{\pi_{\circ}}^{i}\left(1\right),\quad i=2,3,4\cdot$$

We denote by P the orthogonal projector on the orthogonal space spanned by the columns of $\Phi_{\circ}\left(\Lambda^{*}\right)^{-\frac{1}{2}}\frac{\partial z'}{\partial\beta}\left(\theta^{\circ},\beta^{\circ}\right)\Lambda^{*}\frac{\partial z}{\partial\theta'}\left(\theta^{\circ},\beta^{\circ}\right)$. We can rewrite $\widehat{A}_{TS}^{i}$, $i=2,3,4$:

$$\widehat{A}_{TS}^{i}=\Phi_{\circ}\left(\Lambda^{*}\right)^{\frac{1}{2}}P\Phi_{\circ}\left(\Lambda^{*}\right)^{-\frac{1}{2}}\frac{\partial z'}{\partial\beta}\left(\theta^{\circ},\beta^{\circ}\right)\Lambda^{*}\frac{1}{\sqrt{TS}}\sum_{s=1}^{S}\sum_{t=1}^{T}\left[g\left(\widetilde{w}_{t}^{s}\left(\theta^{\circ}\right),\beta^{\circ}\right)-g\left(w_{t},\beta^{\circ}\right)\right]+o_{\pi_{\circ}}^{i}\left(1\right)\cdot$$

$$\xi_{TS}^{i}=\frac{1}{1+\frac{1}{S}}\widehat{A}_{TS}^{i\prime}\Phi_{\circ}\left(\Lambda^{*}\right)^{-1}\widehat{A}_{TS}^{i}+o_{\pi_{\circ}}^{i}\left(1\right),\quad i=2,3,4,$$

$$\xi_{TS}^{i}=\frac{1}{1+\frac{1}{S}}\left\{\frac{\partial z'}{\partial\beta}\left(\theta^{\circ},\beta^{\circ}\right)\Lambda^{*}\frac{1}{\sqrt{TS}}\sum_{s=1}^{S}\sum_{t=1}^{T}\left[g\left(\widetilde{w}_{t}^{s}\left(\theta^{\circ}\right),\beta^{\circ}\right)-g\left(w_{t},\beta^{\circ}\right)\right]\right\}'\times$$
$$\Phi_{\circ}\left(\Lambda^{*}\right)^{-\frac{1}{2}}P\Phi_{\circ}\left(\Lambda^{*}\right)^{\frac{1}{2}}\Phi_{\circ}\left(\Lambda^{*}\right)^{-1}\Phi_{\circ}\left(\Lambda^{*}\right)^{\frac{1}{2}}P\Phi_{\circ}\left(\Lambda^{*}\right)^{-\frac{1}{2}}\times$$
$$\left\{\frac{\partial z'}{\partial\beta}\left(\theta^{\circ},\beta^{\circ}\right)\Lambda^{*}\frac{1}{\sqrt{TS}}\sum_{s=1}^{S}\sum_{t=1}^{T}\left[g\left(\widetilde{w}_{t}^{s}\left(\theta^{\circ}\right),\beta^{\circ}\right)-g\left(w_{t},\beta^{\circ}\right)\right]\right\}+o_{\pi_{\circ}}^{i}\left(1\right),\quad i=2,3,4,$$

$$\xi^i_{TS} = \widehat{C}'_{TS} P P \widehat{C}_{TS} + o^i_{\pi_\circ}(1), \quad i = 2, 3, 4,$$

with:

$$\widehat{C}_{TS} = \frac{1}{\sqrt{1 + \frac{1}{S}}} \Phi_\circ (\Lambda^*)^{-\frac{1}{2}} \frac{\partial z'}{\partial \beta} (\theta^\circ, \beta^\circ) \Lambda^* \frac{1}{\sqrt{TS}} \sum_{s=1}^{S} \sum_{t=1}^{T} \left[g\left(\widetilde{w}^s_t(\theta^\circ), \beta^\circ\right) - g\left(w_t, \beta^\circ\right) \right],$$

and $\widehat{C}_{TS} \xrightarrow[T \to +\infty]{D} \mathcal{N}(0, I_q)$, this ends the proof of proposition 5.2.