SEMI-PARAMETRIC INDIRECT INFERENCE*

by

Ramdan Dridi
London School of Economics and Political Science

and

Eric Renault
CREST-INSEE and Université de Paris IX Dauphine

Contents:

Discussion Paper
No. EM/00/392
May 2000

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

---

* We thank Jean-Pierre Florens, Lars Peter Hansen, Nour Meddahi, Alain Monfort and seminar audiences at ESRC Bristol, LSE, Montreal and Toulouse for helpful comments. All remaining errors are our own.

Abstract

We develop in this paper a generalization of the Indirect Inference (II) to semi-parametric settings and termed Semi-parametric Indirect Inference (SII). We introduce a new notion of Partial Encompassing which lays the emphasis on Pseudo True Values of Interest. The main difference with the older notion of encompassing is that some components of the pseudo-true value of interest associated with the structural parameters do correspond to true unknown values. This enables us to produce a theory of robust estimation despite mis-specifications in the structural model being used as a simulator. We also provide the asymptotic probability distributions of our SII estimators as well as Wald Encompassing Tests (WET) and advocate the use of Hausman type tests on the required assumptions for the consistency of the SII estimators. We illustrate our theory with examples based on semi-parametric stochastic volatility models.

Keywords: Indirect inference; partial encompassing; pseudo-true value of interest; structural models; instrumental models; Wald encompassing tests.

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

1 Introduction

The so-called Indirect Inference methodology was recently introduced in the literature by Smith (1993), Gouriéroux, Monfort and Renault (1993), Gallant and Tauchen (1996), for a simulation-based inference on generally intractable structural models through an instrumental model, conceived as easier to handle. This methodology allows the use of somewhat mis-specified instrumental models, since the simulation process in the **well-specified** structural model and the calibration of the simulated paths against the observed one through the same instrumental model will provide an automatic mis-specification bias correction.

However this theory crucially depends on the correct specification assumption concerning the structural model. And as stressed by Bergstrom (1985) among others, "one of the main differences between econometrics and the application of statistical methods in the physical sciences is that the functional forms in the structural equations of an econometric model are seldom given by the theory". This fact has been explicitly recognized in several important recent advances in the econometrics of mis-specified models (see Monfort (1996) for a recent appraisal of this issue). In this respect, we consider in this paper a semi-parametric framework which specifies only some parameters of interest θ_1 (say) raised out by the economic theory and corresponding to a true unknown value θ_1°. This may be defined through a set of identifying moment conditions. In such a semi-parametric setting, not only the Maximum Likelihood Estimator is no longer available in general, but even more robust M-estimators or Minimum Distance estimators may be unpalatable due to a complicated dynamic structure of the Data Generating Process (DGP) (unobservable state variables, non markovianity...). Consequently the econometrician is led to perform a semi-parametric indirect inference associated with a given pair of structural model and instrumental model.

In order to get a simulator useful for indirect inference about θ_1, the econometrician has to plug this semi-parametric setting into a structural model that is fully parametric and mis-specified in general since it introduces additional assumptions on the law of motion of the DGP. These additional assumptions may require a vector θ_2 of additional parameters so that the vector θ of "structural parameters" is given by $\theta = ({\theta_1}', {\theta_2}')'$.

In this framework, we are naturally led to define the notion of Pseudo True Value of Interest $\theta^* = \left(\theta_1^{\circ\prime}, \overline{\theta}_2^{\prime}\right)'$, where θ_1° corresponds to the true unknown value of the parameters θ_1 and $\overline{\theta}_2$ belongs to Θ_2 a subset of $I\!R^{p_2}$. In order to answer the issue on consistently estimating the true unknown value θ_1° through a semi-parametric indirect inference, we introduce the notion of Partial Encompassing. The main difference with the older notion of encompassing as proposed by Mizon and Richard (1986), or for a simulated version à la Gouriéroux and Monfort (1995) and Dhaene, Gouriéroux and Scaillet (1998) is that the emphasis is led on a pseudo-true value of interest defined by the true value θ_1° for any given $\overline{\theta}_2$. Moreover, in this framework and when required by the partial encompassing property, some of the nuisance parameters θ_{22} (say) are not estimated in the first step SII but in a second step introducing some general simulation-based loss function. The basic idea of partial encompassing is something like a "ceteris paribus" condition which ensures that consistency is maintained for the estimation of θ_1° while $\overline{\theta}_2$ might have to be fixed from some other extra information. In the same line as Bierens and Swanson (2000) who have recently called upon a ceteris paribus condition to formalize some ideas stemming from the so-called Calibration methodology of the new classical macro-economics, we do think that SII provides a useful framework to understand what calibrato2s exactly do. We develop this thesis in a companion paper (Broze, Dridi and Renault (1999)).

Actually, we consider that our SII methodology starts from the same issues as the Calibration one but complements it by some inference tools that are needed for a comprehensive statistical strategy.

By analogy with the Quasi Maximum Likelihood methodology (White (1982), Gouriéroux, Monfort and Trognon (1984)), we show that standard GMM or Indirect Inference results cannot be directly applied in the calibration context but need a preliminary "robustification" against the likely mis-specification of the structural model. The formalization of this mis-specification, through our new notion of partial encompassing enables us to derive the asymptotic probability distribution of the SII estimators and associated Wald Encompassing Tests (WET). Moreover, we lay out a general specification strategy involving Hausman type tests.

The paper is organized as follows. We first recall in section 2 a brief overview of the available results on Indirect Inference when the structural model is correctly specified. Then we provide an extended semi-parametric framework for indirect inference. We address, in section 3 the issue on robustness of Indirect Inference with respect to mis-specifications in the structural model. We propose a formalization of a general setting, termed Semi-parametric Indirect Inference (SII), where the consistency of the estimators of the structural parameters of interest is maintained. In section 4 we deduce the asymptotic probability distribution of the SII estimators. We also provide a diagnostic procedure of tests about the null hypothesis that ensures the consistency of the SII estimators; this procedure is based on Wald Encompassing Tests. In order to increase the power of the testing procedure against spurious fit, as stressed by Tauchen (1997), we advocate the use of simulated Hausman type tests. We discuss in section 5 the issue on estimating the nuisance parameters θ_{22} in a second step estimation by using a general loss function and the consequences on the asymptotic results. We give in section 6 an example based on semi-parametric stochastic volatility modeling where our SII and procedures of tests ensures the desired consistency property. Finally section 7 states some concluding remarks.

2 An Extended Framework for Semi-parametric Indirect Inference

2.1 Indirect Inference principle

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

$$\begin{aligned} r(y_t, y_{t-1}, x_t, u_t, \theta) &= 0, & (2.1)\\ \varphi(u_t, u_{t-1}, \varepsilon_t, \theta) &= 0, & (2.2) \end{aligned}$$

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

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 strongly exogenous observable variables. The variables $\{u_t, t \in \mathbb{Z}\}$ and $\{\varepsilon_t, t \in \mathbb{Z}\}$ are not observed.

We assume that $\{x_t, t \in \mathbb{Z}\}$ 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 and the process $\{(y_t, x_t), t \in \mathbb{Z}\}$ is stationary[1]. For each given value of the parameters θ, it is possible to simulate values $\{\tilde{y}_1(\theta, z_\circ), \ldots, \tilde{y}_T(\theta, z_\circ)\}$ 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 $\{(y_t, u_t), t \in \mathbb{Z}\}$.

Let θ° be the true unknown value of θ assuming that the structural model (2.1) – (2.2) is well-specified. A direct estimation of θ° is often cumbersome since the conditional probability density function (p.d.f. hereafter) of $\{y_1, \ldots, y_T\}$ given $\{z_\circ, x_1, \ldots, x_T\}$ may be computationally intractable. The idea is then to replace the intractable log-likelihood function of the structural model:

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

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

$$\begin{gathered} Q_T(\underline{y}_T, \underline{x}_T, \beta) = \sum_{t=1}^{T} q_t(y_t / \underline{y}_{t-1}, \underline{x}_t, \beta), \\ \beta \in \mathcal{B} \text{ a compact subset of } I\!R^q. \end{gathered} \tag{2.4}$$

[1]This assumption may be relaxed.

On the one hand, we associate with Q_T the M-estimator $\widehat{\beta}_T$ of β defined by:

$$\widehat{\beta}_T = \underset{\beta\in\mathcal{B}}{Argmin}\, Q_T(\underline{y}_T, \underline{x}_T, \beta), \tag{2.5}$$

on the other hand, from simulated values $\widetilde{y}_t^s(\theta, z_\circ^s),\ t = 1,\ldots,\ T,\ s = 1,\ldots,\ S$ of the endogenous variables, we can compute for $s = 1,\ldots,\ S$:

$$\begin{aligned} \widetilde{\beta}_T^s(\theta) &= \underset{\beta\in\mathcal{B}}{Argmin}\, Q_T\left(\underline{\widetilde{y}}_T^s(\theta, z_\circ^s), \underline{x}_t, \beta\right), \\ \widetilde{\beta}_{TS}(\theta) &= \frac{1}{S}\sum_{s=1}^{S} \widetilde{\beta}_T^s(\theta)\cdot \end{aligned} \tag{2.6}$$

Under usual regularity conditions, this defines the so-called binding function:

$$\begin{aligned} \widetilde{\beta}(\theta) &= \underset{T\to+\infty}{P_\circ \lim}\, \widetilde{\beta}_{TS}(\theta), \\ \widetilde{\beta}(\theta^\circ) &= \beta^\circ = \underset{T\to+\infty}{P_\circ \lim}\, \widehat{\beta}_T\cdot \end{aligned} \tag{2.7}$$ [2]

The class of indirect estimators is indexed by a choice of a positive weighting matrix Ω of size $q\times q$. For a given Ω, the indirect inference (II hereafter) estimator is defined by:

$$\widehat{\theta}_{TS}(\Omega) = \underset{\theta\in\Theta}{Argmin}\left[\widehat{\beta}_T - \widetilde{\beta}_{TS}(\theta)\right]'\Omega\left[\widehat{\beta}_T - \widetilde{\beta}_{TS}(\theta)\right]\cdot \tag{2.8}$$

As usual, the indirect inference estimator $\widehat{\theta}_{TS}(\Omega)$ will be computed in practice by 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. This justifies the notation (2.8). But, in order to minimize the asymptotic covariance matrix of $\widehat{\theta}_{TS}(\Omega)$, an optimal choice of Ω:

$$\Omega^* = J_\circ\left(I_\circ - K_\circ\right)^{-1} J_\circ, \tag{2.9}$$

$$J_\circ = \underset{\circ}{E}\left[\frac{\partial^2 q_t}{\partial\beta\partial\beta'}(y_t/\underline{y}_{t-1}, \underline{x}_t, \beta^\circ)\right],$$

$$I_\circ = \underset{\circ}{Var}_{as}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial q_t}{\partial\beta}(y_t/\underline{y}_{t-1}, \underline{x}_t, \beta^\circ)\right],$$

$$K_\circ = \underset{\circ}{Var}_{as}\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\underset{\circ}{E}\left[\frac{\partial q_t}{\partial\beta}(y_t/\underline{y}_{t-1}, \underline{x}_t, \beta^\circ)/\underline{x}_t\right]\right\}\cdot$$

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

$$W_S = \underset{\circ}{Var}_{as}\left[\sqrt{T}(\widehat{\theta}_{TS}^* - \theta^\circ)\right] = \left(1 + \frac{1}{S}\right)\left[\frac{\partial\widetilde{\beta}'}{\partial\theta}(\theta^\circ) J_\circ\left(I_\circ - K_\circ\right)^{-1} J_\circ \frac{\partial\widetilde{\beta}}{\partial\theta'}(\theta^\circ)\right]^{-1}\cdot \tag{2.10}$$

[2] We denote by $\underset{T\to+\infty}{P_\circ \lim}$ the limit in probability (with respect to $P_\circ$) when T goes to infinity.

Gouriéroux and Monfort (1996) has shown that, under convenient regularity conditions, these results are still valid for a general instrumental criterion $Q_T\left(\underline{y}_T, \underline{x}_T, \beta\right)$. In particular, we have now to consider matrices $I_\circ - K_\circ$ and $J_\circ$ according to the following general definitions:

$$I_\circ - K_\circ = \underset{\circ}{V}\underset{as}{ar}\left[\sqrt{T}\frac{\partial Q_T}{\partial \beta}\left(\underline{y}_T, \underline{x}_T, \beta^\circ\right) - \underset{\circ}{E}\left[\sqrt{T}\,\frac{\partial Q_T}{\partial \beta}\left(\underline{y}_T, \underline{x}_T, \beta^\circ\right)/\underline{x}_T\right]\right],$$
$$J_\circ = P_\circ \lim_{T\to+\infty} \frac{\partial^2 Q_T}{\partial \beta \partial \beta'}\left(\underline{y}_T, \underline{x}_T, \beta^\circ\right)\cdot$$

2.2 Indirect Inference in a semi-parametric setting

As previously announced, the main goal of this subsection is to extend the Indirect Inference principle to semi-parametric settings. The semi-parametric modeling widely adapted in modern econometrics does correspond indeed to an alternative to the "quest for the Holy Grail" (see Monfort (1996)), that is the hopeless search for a well-specified parametric model that is more often than not impossible to deduce from the Economic Theory and specifies only some parameters of interest θ_1 (say) raised out by the underlying Economic Theory. Therefore, we have first in this subsection to extend the semi-parametric point of view to an II framework before revisiting the issues on consistency (section 3) and asymptotic probability distributions of II estimators and tests in this setting (section 4).

As in the previous subsection, 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 $\{(y_t, x_t), t \in \mathbb{Z}\}$[3].

Assumption (A1):

(i) $P_\circ$ belongs to a family $\mathcal{P}$ of probability distributions on $(\mathcal{X} \times \mathcal{Y})^{\mathbb{Z}}$.

(ii) $\widetilde{\theta_1}$ is an application from $\mathcal{P}$ onto a compact set $\Theta_1 = \widetilde{\theta_1}(\mathcal{P})$ of $\mathbb{R}^{p_1}$.

(iii) $\widetilde{\theta_1}(P_\circ) = \theta_1^\circ$, the true unknown value of the parameters of interest, belongs to the interior $\overset{\circ}{\Theta}_1$ of Θ_1.

$\widetilde{\theta_1}(P) = \theta_1$ is the vector of unknown parameters of interest. Several illustrations of the relevance of this framework are indeed envisioned and developed in two companion papers (Dridi (2000) and Broze, Dridi and Renault (1999)). Actually, we stress that there are nowadays various occasions in Economics as well as in Statistics, where a semi-parametric set-up is available for the **definition** of the parameters of interest θ_1 according to $(A1)$; but because of unobservable components (as in state variables models), because of non availability of relevant aggregate economic variables at the proper frequency, the standard semi-parametric methods (QML, GMM, PMLE) do no longer apply.

Typically, in the case of a stationary process $\{(y_t, x_t), t \in \mathbb{Z}\}$, the parameters of interest may be defined through a set g of identifying moment restrictions:

$$\underset{P}{E}\, g\left(y_t, x_t, u_t, y_{t-1}, x_{t-1}, u_{t-1}, \ldots, y_{t-K}, x_{t-K}, u_{t-K}, \theta_1\right) = 0 \quad \Longrightarrow \quad \theta_1 = \widetilde{\theta_1}(P)\cdot \tag{2.11}$$

[3]It is essential to keep in mind that the disentangling (y, x) by no way and in accordance with the forthcoming assumption $(A1)$ means that $\{x_t, t \in \mathbb{Z}\}$ is exogenous.

Conditional moment restrictions may also be considered. The explicit occurrence of latent processes $\{u_t, t \in \mathbb{Z}\}$ in this moment restrictions (think for instance about a stochastic volatility process) prevents one from using GMM. Thus one may always imagine **to perform semi-parametric indirect inference** associated with a given pair of "structural" model (used as simulator) and "auxiliary" (or "instrumental") criterion.

In order to get a simulator useful for indirect inference on θ_1, we have to plug the semi-parametric model defined by $(A1)$ into a structural model that is **fully parametric** (at least with respect to the conditional probability distribution of y given x) and mis-specified in general since it introduces additional assumptions on the law of motion of (y, x) which are not suggested by any Economic Theory. These additional assumptions may require a vector θ_2 of additional parameters in such a way that the vector θ of "structural parameters" is given by $\theta = ({\theta_1}', {\theta_2}')'$. We then formulate a nominal assumption $(B1)$ to specify a structural model conformable to the previous section, even though we know that $(B1)$ is likely to be inconsistent with the true DGP. [4]

Nominal assumptions (B1):

$\{(y_t, x_t), t \in \mathbb{Z}\}$ is a stationary process conformable to the following nonlinear simultaneous equations model:

- $$\begin{cases} r(y_t, y_{t-1}, x_t, u_t, \theta) &= 0, \\ \varphi(u_t, u_{t-1}, \varepsilon_t, \theta) &= 0, \end{cases} \tag{2.12}$$

  $\theta = ({\theta_1}', {\theta_2}')' \in (\Theta_1 \times \Theta_2) = \Theta$ a compact subset of $\mathbb{R}^{p_1+p_2}$,

- the exogenous process $\{x_t, t \in \mathbb{Z}\}$ is independent of $\{\varepsilon_t, t \in \mathbb{Z}\}$,

- $\{\varepsilon_t, t \in \mathbb{Z}\}$ is a white noise with a known distribution G_*.

We denote π_* the p.d.f. of the process $\{x_t,\ \varepsilon_t, t \in \mathbb{Z}\}$ defined as the product of the true unknown p.d.f. of $\{x_t, t \in \mathbb{Z}\}$ (marginalization of $P_\circ$) and G_*. Note that the space of unknown parameters Θ is defined as a product space $\Theta_1 \times \Theta_2$ for sake of notations simplicity.

As a joint hypothesis, the structural model $(B1)$ is mis-specified in general for at least two reasons:

- Economic Theory provides little guidance about the functional forms r and φ including the number of lags, of unobserved state variables u (and ε) and nuisance (or technology) parameters θ_2.

- Even if the structural equations (2.12) are valid, because the underlying Economic Theory is itself correct, the purely statistical assumptions (exogeneity property for $\{x_t, t \in \mathbb{Z}\}$, known distribution G_* for ε_t) may not be fulfilled by the DGP. [5]

We focus here on indirect inference about the true value θ_1° of the parameters of interest θ_1. This indirect inference is termed semi-parametric since we do not trust the nominal assumptions $(B1)$. However the Indirect Inference principle is still defined from the two basic components: a

[4]We denote by B the nominal assumptions, i.e. assumptions that are used for a quasi-indirect inference even though they are known to be mis-specified. This mis-specification will be taken into account in our results that we prove thanks to the maintained assumptions, denoted by A.

[5]For instance, the exogeneity assumption concerning the process $\{x_t, t \in \mathbb{Z}\}$ is maintained within the nominal assumptions $(B1)$ whereas it is generally not within the semi-parametric model $(A1)$. Consequently and as already pointed out, the disentangling (y, x) just corresponds to modeling (and therefore may be mis-specified) and not to a genuine fulfilled property.

"structural" model $(B1)$ and a general instrumental criterion:

$$Q_T\left(\underline{y}_T, \underline{x}_T, \beta\right), \qquad \beta \in \mathcal{B} \text{ a compact subset of } I\!\!R^q \cdot \tag{2.13}$$

Assumption (A2):

$$(A2) \qquad \begin{aligned} &P_\circ \lim_{T\to+\infty} \mathop{Sup}_{\beta\in\mathcal{B}} \left| Q_T\left(\underline{y}_T, \underline{x}_T, \beta\right) - q_\circ(\beta)\right| = 0, \\ \forall\theta\in\Theta, \quad &\pi_* \lim_{T\to+\infty} \mathop{Sup}_{\beta\in\mathcal{B}} \left| Q_T\left(\underline{\widetilde{y}}^s_T(\theta, z^s_\circ), \underline{x}_T, \beta\right) - q_M(\theta,\beta)\right| = 0 \cdot \end{aligned}$$ [6]

$\{\widetilde{y}^s_1(\theta, z^s_\circ), \ldots, \widetilde{y}^s_T(\theta, z^s_\circ)\}$ correspond to simulated paths of the dependent variable according to the model $(B1)$ conditionally on $\{x_1, \ldots, x_T\}$ and $z^s_\circ$ for $s = 1, \ldots, S$.
$q_\circ(\beta)$ and $q_M(\theta,\beta)$ are assumed to be non stochastic twice differentiable functions not depending on the initial conditions $z^s_\circ$ and with a unique minimum with respect to β. Let β° and $\widetilde{\beta}(\theta_1,\theta_2)$ be respectively the minimum of $q_\circ(\beta)$ and $q_M(\theta,\beta)$.
Assumption (A3):

$$(A3) \qquad \begin{aligned} \beta^\circ = \beta(P_\circ) &= \mathop{Argmin}_{\beta\in\mathcal{B}} q_\circ(\beta), \\ \widetilde{\beta}(\theta_1,\theta_2) &= \mathop{Argmin}_{\beta\in\mathcal{B}} q_M(\theta_1,\theta_2,\beta) \cdot \end{aligned}$$

Assumption (A4):
$\widetilde{\beta}(\cdot,\cdot)$ is one-to-one.
According to Gouriéroux and Monfort (1995) definitions, β° is the pseudo true value for the instrumental model (2.13) and $\widetilde{\beta}(\cdot,\cdot)$ is the binding function from the structural model (2.12) to the instrumental one (2.13).[7] In their terminology, assumption $(A4)$ is referred to as the indirect identification of θ from β. This is related to the indirect inference procedure described below. Let us introduce the following estimators:

$$\begin{aligned} &\widehat{\beta}_T = \mathop{Argmin}_{\beta\in\mathcal{B}} Q_T\left(\underline{y}_T, \underline{x}_T, \beta\right), \\ &\widetilde{\beta}^s_T(\theta_1,\theta_2) = \mathop{Argmin}_{\beta\in\mathcal{B}} Q_T\left(\underline{\widetilde{y}}^s_T(\theta_1,\theta_2, z^s_\circ), \underline{x}_T, \beta\right), \\ &\widetilde{\beta}_{TS}(\theta_1,\theta_2) = \frac{1}{S}\sum_{s=1}^S \widetilde{\beta}^s_T(\theta_1,\theta_2) \cdot \end{aligned} \tag{2.14}$$

Under assumptions $(A2) - (A3)$, these estimators converge to:

$$\begin{aligned} &P_\circ \lim_{T\to+\infty} \widehat{\beta}_T = \beta\left(P_\circ\right), \\ &\pi_* \lim_{T\to+\infty} \widetilde{\beta}^s_T(\theta_1,\theta_2) = \pi_* \lim_{T\to+\infty} \widetilde{\beta}_{TS}(\theta_1,\theta_2) = \widetilde{\beta}(\theta_1,\theta_2) \cdot \end{aligned}$$

[6] We denote by $\pi_* \lim_{T\to+\infty}$ the limit in probability (with respect to π_*) when T goes to infinity.

[7] The instrumental criterion is generally speaking suggested by an instrumental model. We refer to (2.13) as an instrumental model to be conformable to Gouriéroux and Monfort (1995) terminology.

Assumption $(A5)$**:**

We assume in addition that the latter convergence is uniform in θ, that is for $s = 1, \ldots, S$:

$$(A5) \qquad \pi_* \lim_{T \to +\infty} \underset{(\theta_1', \theta_2')' \in \Theta}{Sup} \left\| \widetilde{\beta}_T^s(\theta_1, \theta_2) - \widetilde{\beta}(\theta_1, \theta_2) \right\|_q = 0\cdot$$

An indirect inference estimator $\widehat{\theta}_{TS}$ is then defined as follows:

$$\widehat{\theta}_{TS} = \left(\widehat{\theta}_{1,TS}', \widehat{\theta}_{2,TS}'\right)' = \underset{(\theta_1, \theta_2) \in \Theta_1 \times \Theta_2}{Argmin} \left[\widehat{\beta}_T - \widetilde{\beta}_{TS}(\theta_1, \theta_2)\right]' \widehat{\Omega}_{TS} \left[\widehat{\beta}_T - \widetilde{\beta}_{TS}(\theta_1, \theta_2)\right], \tag{2.15}$$

where $\widehat{\Omega}_{TS}$ is a positive matrix which may depend on both simulated $\widetilde{\varepsilon}_t$ and observed (y_t, x_t). Thus, in order to extend the standard minimum distance setting of indirect inference, we have to assume now that: $P_* \lim_{T \to +\infty} \widehat{\Omega}_{TS} = \Omega$ a positive matrix on $I\!R^q$, where P_* denotes the joint p.d.f. of (y, x, ε) defined with obvious notations by:

$$P_* = \pi_* \otimes P_\circ^{Y/X}\cdot \tag{2.16}$$

Endowed with such a notation, all the probability limits considered in this paper can be viewed w.r.t. P_*.

However, several important differences with respect to the standard setting of indirect inference (as reminded in subsection 2.1 above) have to be emphasized:

• First, due to the mis-specification of the structural model $(B1)$, there is in general no reason to hope that the limit problem:

$$\underset{\theta_1, \theta_2 \in \Theta_1 \times \Theta_2}{Min} \left\| \beta\left(P_\circ\right) - \widetilde{\beta}\left(\theta_1, \theta_2\right) \right\|_\Omega, \tag{2.17}$$

(where $\|\beta\|_\Omega^2 = \beta' \Omega \beta$) has a null value. A fortiori, **for a given choice of the weighting matrix $\boldsymbol{\Omega}$**, there will be in general a set of minimizers which, first, is not reduced to a singleton and second does depend on this choice of Ω. This is indeed a standard issue on GMM estimation applied to mis-specified moment conditions. There does not exist in general a unique pseudo-true value which allows one to properly define a consistency concept. We will maintain in our theoretical developments below the assumption that (2.17) has a null value and therefore admits a unique minimizer by virtue of the one-to-one mapping assumption $(A4)$ on $\widetilde{\beta}\left(\cdot, \cdot\right)$[8].

• Second, besides the aforementioned pitfall resulting from some mis-specification in the set of "moments to match", there is a second pitfall, more specific to simulation-based inference, where data simulated from a wrong DGP may give the fallacious feeling of a perfect fit. Namely, it may exist a pseudo-true value θ_1^* different from the true value θ_1° such that $\widetilde{\beta}\left(\theta_1^*, \theta_2^*\right) = \beta\left(P_\circ\right)$ for some $\theta_2^* \in \Theta_2$.

To summarize, a consistent semi-parametric indirect inference estimator $\widehat{\theta}_{1,TS}$ for a family $\mathcal{P}$ of possible DGP has to be defined in light of the two previous pitfalls:

$\star$ First, (2.17) should define a unique pseudo-true value $\theta^* = \left(\theta_1^{*'}, \theta_2^{*'}\right)'$ of the structural parameters

[8]We will stress in subsection 3.2 below about the encompassing that an assumption of uniqueness of the minimizer which would not correspond to a null value of the criterion is not releva.t, since this minimizer would be unique only up to the arbitrary choice of Ω. Moreover, it would be always possible to relax the assumption of uniqueness by introducing correspondences of minimizers but this is beyond the scope of this paper.

for any allowed (w.r.t. $(A1)$ and $(B1)$) probability distribution P_*. Let us denote by $\theta_1^* = \theta_1^*(P_*)$ the corresponding pseudo-true value of the structural parameters of interest.

$\star$ Second, this pseudo-true value $\theta_1^*(P_*)$ should coincide with the true unknown value $\widetilde{\theta}_1(P)$ (where P_* corresponds to the product of π_* and $P^{Y/X}$ as in (2.16)). Of course, this coincidence issue makes sense only if the nominal parametric model $(B1)$ is compatible with the maintained semi-parametric model $(A1)$. In other words, we have:

Assumption (A6):

For any $\theta = (\theta_1', \theta_2')' \in \Theta_1 \times \Theta_2, \quad \widetilde{\theta}_1(P_\theta) = \theta_1$, where P_θ denotes the **actual** probability distribution of $(\widetilde{y}(\theta), x)$ from the simulator[9] $(B1)$.

In the general semi-parametric framework delineated by assumptions $(A1) - (A6)$, we provide in the next section the consistency criteria which are required to deal with the two previous pitfalls.

[9]Of course, as already explained, the path of the process x associated with the simulated path $\widetilde{y}(\theta)$ is the observed one and is not simulated. Indeed, P_θ is fully characterized by the endowment of π_* and θ plugged into (2.12).

3 Semi-parametric Indirect Estimation

We first provide in subsection 3.1 an necessary and sufficient condition for the consistency of the Semi-parametric Indirect Inference (SII) estimator $\widehat{\theta}_{1,TS}$ defined by (2.15), while we focus in subsection 3.2 on sufficient and testable conditions.

3.1 Consistency of the semi-parametric indirect inference estimator

In order to derive a necessary and sufficient condition for the consistency of $\widehat{\theta}_{1,TS}$ to θ_1°, let us define the so-called "generalized inverse" $\widetilde{\beta}^-$ of $\widetilde{\beta}$ by:

$$\widetilde{\beta}^-(\beta) = \underset{(\theta_1,\theta_2)\in\Theta_1\times\Theta_2}{Argmin} \left\|\beta - \widetilde{\beta}(\theta_1,\theta_2)\right\|_\Omega \cdot \tag{3.1}$$

$\widetilde{\beta}^-$ is a correspondence from $I\!R^q$ onto $\Theta_1 \times \Theta_2$ whose restriction to $\widetilde{\beta}\,(\Theta)$ is a genuine inverse of the function $\widetilde{\beta}$, which is one-to-one by virtue of the identification assumption $(A4)$. In the case of standard indirect inference with a well-specified structural model, the consistency of the indirect inference estimator is guaranteed thanks to the fact that $\widetilde{\beta}^-\,[\beta(P)]$ coincides with the parameterizations of the structural model. More precisely, while a function $P \longrightarrow \beta(P)$ is always defined by extension of definition $(A3)$ to any possible DGP (P) in the model $(A1)$, a function $\widetilde{\theta}\,(P) = \left(\widetilde{\theta_1}\,(P)', \widetilde{\theta_2}\,(P)'\right)'$ is only defined in the particular case where the structural model $(A1)$ is fully parametric ; in this case $(A1)$ coincides with $(B1)$ seen as well-specified by a one-to-one parameterization:

$$({\theta_1}', {\theta_2}')' \longrightarrow P_{({\theta_1}',{\theta_2}')'},$$

$$\widetilde{\theta}\left(P_{({\theta_1}',{\theta_2}')'}\right) = ({\theta_1}', {\theta_2}')'\cdot$$

In our semi-parametric setting, we are only interested in the projection of $\widetilde{\beta}^-\,[\beta(P)]$ on the set Θ_1 of the parameters of interest. Let us denote by Q_1 the projection operator:

$$\begin{array}{cccccc} Q_1 & : & I\!R^{p_1} \times I\!R^{p_2} & \longrightarrow & I\!R^{p_1}, \\ & & ({\theta_1}', {\theta_2}')' & \longrightarrow & \theta_1\cdot \end{array}$$

We are then led to the following consistency criterion:

Proposition 3.1 : *Under assumptions* $(A1)-(A6)$, $\widehat{\theta}_{1,TS}$ *is a consistent estimator of the parameters of interest* θ_1° *if and only if, for any* P *in the family* $\mathcal{P}$ *of probability distributions delineated by the semi-parametric model* $(A1)$*:*

$$Q_1\left[\widetilde{\beta}^- \circ \beta\,(P)\right] = \widetilde{\theta_1}\,(P)\cdot$$

Proof : *Under assumptions* $(A1)-(A6)$, $\widehat{\theta}_{1,TS}(P)$ *is consistent to* $\widetilde{\theta}_1(P)$ *if and only if:*

$$\begin{aligned}\forall P\in\mathcal{P},\quad \underset{T\to+\infty}{P_*\lim}\,\widehat{\theta}_{1,TS}(P) &= \theta_1^*(P)=\widetilde{\theta}_1(P),\\ &= Q_1\left[\underset{\theta\in\Theta}{Argmin}\left\|\beta(P)-\widetilde{\beta}\left(\theta_1,\theta_2\right)\right\|_{\Omega}\right],\\ &= Q_1\left[\widetilde{\beta}^{-}\circ\beta(P)\right]\cdot\end{aligned}$$

This ends the proof of proposition 3.1.

In order to illustrate to what extent the criterion of proposition 3.1 imposes constraints on both the semi-parametric model, the nominal structural model and the instrumental model, it may be helpful to have the following setting in mind. Let us imagine that the nominal structural model is mis-specified because it imposes some invalid constraints on some nuisance parameters θ_3. In other words, we start from a "parametric" representation of the set $\mathcal{P}$ of probability distributions of interest:

$$\mathcal{P}=\left\{P_\lambda,\ \lambda=({\theta_1}',{\theta_2}',{\theta_3}')'\in\Theta_1\times\Theta_2\times\Theta_3\right\}\cdot$$

The term "parametric" is used here with a very general meaning: the nuisance "parameters" θ_3 may be functional that is Θ_3 may be of infinite dimension. The only important assumption consists in the correct specification of this "parametric" model:

$$P_\circ=P_{\lambda^\circ}\text{ for }\lambda^\circ=\left(\theta_1^{\circ\prime},\theta_2^{\circ\prime},\theta_3^{\circ\prime}\right)'\in\Theta_1\times\Theta_2\times\Theta_3\cdot \tag{3.2}$$

Therefore, a slight change of notation allows us to rewrite: $\beta\left(P\right)=\beta(\theta_1,\theta_2,\theta_3)$ when $P=P_\lambda$ with $\lambda=({\theta_1}',{\theta_2}',{\theta_3}')'$. Furthermore, the nominal structural model $(B1)$ is mis-specified whenever it imposes some invalid constraint on the nuisance parameters θ_3: $\theta_3=0,$ while $\theta_3^\circ\neq0$ (say). In other words: $\widetilde{\beta}(\theta_1,\theta_2)=\beta(\theta_1,\theta_2,0)\cdot$ In such a setting:

$$\widetilde{\beta}^{-}(\beta)=\underset{(\theta_1,\theta_2)\in\Theta_1\times\Theta_2}{Argmin}\left\|\beta-\widetilde{\beta}(\theta_1,\theta_2,0)\right\|_{\Omega},$$

and therefore:

$$\widetilde{\beta}^{-}\left[\beta\left(P_\lambda\right)\right]=\underset{\left(\overline{\theta}_1,\overline{\theta}_2\right)\in\Theta_1\times\Theta_2}{Argmin}\left\|\beta(\theta_1,\theta_2,\theta_3)-\beta(\overline{\theta}_1,\overline{\theta}_2,0)\right\|_{\Omega}\cdot \tag{3.3}$$

The program (3.3) highlights with a new perspective the two already announced pitfalls, which may prevent from getting semi-parametric consistency:

• First, when the nominal structural model does not contain the DGP ($\theta_3^\circ\neq0$) there does not exist in general $\overline{\theta}_2$ such that:

$$\beta(\theta_1^\circ,\overline{\theta}_2,0)=\beta(\theta_1^\circ,\theta_2^\circ,\theta_3^\circ)\cdot$$

• Second, a perfect fit may occur with wrong values $\left(\theta_1^{*\prime},\theta_2^{*\prime}\right)'$ of the parameters:

$$\beta(\theta_1^*,\theta_2^*,0)=\beta(\theta_1^\circ,\theta_2^\circ,\theta_3^\circ),\text{ while }\theta_1^*\neq\theta_1^\circ\cdot$$

These pitfalls illustrate the well-known "joint hypothesis" issue in Econometrics:

• Either, one wants to perform inference without any a priori restriction, that is estimating

$(\theta_1, \theta_2, \theta_3)$ without constraint. But such a fully unrestricted approach is generally infeasible due to the curse of dimensionality: inference on a reasonable number of instrumental parameters β does not provide a sufficient indirect information to identify all the relevant features of the DGP described by $(\theta_1, \theta_2, \theta_3)$.

• Or, due to this lack of identification, the econometrician is led to add to the knowledge which often comes from the Economic Theory and is never brought into question some arbitrary and a priori restrictions $\theta_3 = 0$ in order to identify the parameters of interest θ_1 and some nuisance parameters θ_2 thanks to an identification assumption extending $(A4)$: $\beta(\cdot, \cdot, \overline{\theta}_3)$ is a one-to-one function of $(\theta_1, \theta_2) \in \Theta_1 \times \Theta_2$ for any fixed value $\overline{\theta}_3 \in \Theta_3$. In such a setting, the lack of consistency stressed above is precisely due the wrong "joint hypothesis" about θ_3: $\beta(\theta_1^*, \theta_2^*, 0) = \beta(\theta_1^\circ, \theta_2^\circ, \theta_3^\circ)$ may occur with $\theta_1^* \neq \theta_1^\circ$ because $\theta_3^\circ \neq 0$.

Last but not least, in case of mis-specification, the minimum (3.3) is generally not reached at a null value. This means that the discrepancy between θ_1° and the limit (in probability) $Q_1 \left[\widetilde{\beta}^- \left(\beta \left(P_{\lambda_\circ}\right)\right)\right]$ of the semi-parametric II estimator $\widehat{\theta}_{1,TS}$ will generally depend on the choice Ω of the weighting matrix. Consequently, the following subsection 3.2 focuses on a case where, whatever the mis-specification, the minimum (3.3) is reached at a null value. This is the only case where **general statements**, that is, statements that are independent of the arbitrary choice of a metric on instrumental parameters may make sense.

3.2 An encompassing sufficient condition for consistency

Following Mizon and Richard (1986), Gouriéroux and Monfort (1995) have used the concept of binding function to set up a formal definition of the encompassing principle. This principle involves notions of pseudo-true values and binding function which are underpinned by the Kullback Leibler Information Criterion (KLIC) as a proximity criterion. But it is clear that other proximity criteria may be used to capture some structural a-statistical ideas, which lead to loss functions different from the log-likelihood ratio as explained by Dhaene, Gouriéroux and Scaillet (1998). Besides this, the specific feature of our setting is that we consider a parametric model $(B1)$ which is mis-specified but introduces a vector of unknown parameters $(\theta_1', \theta_2')'$ whose first p_1 components do correspond to some structural well-specified ideas (according to $(A1)$). As a consequence, we propose here to focus on pseudo-true values of $(B1)$ of the form $\left(\theta_1^{\circ'}, \overline{\theta}_2'\right)'$ where $\theta_1^\circ = \widetilde{\theta_1}\left(P_\circ\right)$ is the true unknown value of the parameters of interest. On the other hand, the instrumental criterion (2.13) defines a pseudo-true value β° of the "instrumental model" (N_β). Typically, the instrumental criterion (2.13) may be the log-likelihood of an instrumental model which is a proxy of some structural model ; in such a case β° is a pseudo-true value conformable to Gouriéroux and Monfort (1995) terminology. By extension to Gouriéroux and Monfort (1995) definition we are allowed to interpret the function $\widetilde{\beta}(\cdot, \cdot)$ defined by $(A3)$ as a link function from $(B1)$ to (N_β). Then, we say that:

Definition 3.1 : *$(B1)$ endowed with the true unknown value θ_1° fully encompasses (N_β) if there exists $\overline{\theta}_2 \in \Theta_2$ such that:*

$$\beta^\circ = \widetilde{\beta}(\theta_1^\circ, \overline{\theta}_2) \cdot \tag{3.4}$$

In this framework, we are able to prove the following sufficient condition for the consistency of the semi-parametric II estimator $\widehat{\theta}_{1,TS}$:

Proposition 3.2 : *Under assumptions $(A1)-(A6)$ and if $(B1)$ endowed with the true value θ_1° fully encompasses (N_β), then $\widehat{\theta}_{1,TS}$ is a consistent estimator of the parameters of interest θ_1°.*

Proof : *Proposition 3.2 is a direct corollary of proposition 3.1 since:*

$$\begin{aligned}
\beta^\circ &= \widetilde{\beta}(\theta_1^\circ, \overline{\theta}_2), \\
\Longrightarrow \quad \left(\theta_1^{\circ'}, \overline{\theta}_2'\right)' &= \underset{(\theta_1,\theta_2)\in\Theta_1\times\Theta_2}{Argmin} \left\| \beta^\circ - \widetilde{\beta}(\theta_1, \theta_2) \right\|_\Omega, \\
&= \widetilde{\beta}^-(\beta^\circ), \\
\Longrightarrow \quad \theta_1^\circ &= Q_1\left[\widetilde{\beta}^-(\beta^\circ)\right],
\end{aligned}$$

which corresponds to the criterion of proposition 3.1.

Moreover, let us notice that the above minimization program is reached at a null value, as already announced.

When the structural mis-specified model $(B1)$ endowed with the true value θ_1° does not fully encompass the instrumental model (N_β), and when the minimum program (2.17) is reached at a null value, we know thanks to the identification assumption $(A4)$, that the semi-parametric indirect inference estimator $\widehat{\theta}_{1,TS}$ is not consistent to the true unknown value θ_1° of the parameters of interest. Facing this inconsistency, one may imagine two alternative strategies:

• First, one may believe that the encompassing property is violated only because some particular moments or more generally a subset β_2 of instrumental parameters cannot be "matched" while some proper subset β_1 $(\beta = (\beta_1', \beta_2')')$ does fulfill the encompassing property:

$$\exists \overline{\theta}_2 / \quad \beta_1^\circ = \widetilde{\beta}_1\left(\theta_1^\circ, \overline{\theta}_2\right) \cdot \tag{3.5}$$

In this case, the required asymptotic theory is almost unchanged when one replaces β by β_1, to the extent that β_1 is also one-to-one. Of course, nothing is changed when the instrumental criterion $Q_T\left(\underline{y}_T, \underline{x}_T, \beta\right)$ is defined by separable just-identifying moment conditions:

$$Q_T\left(\underline{y}_T, \underline{x}_T \beta\right) = \left\| \frac{1}{T}\sum_{t=1}^{T} h\left(\underline{y}_t, \underline{x}_t, \beta_1\right)\right\|_{q_1}^2 + \left\| \frac{1}{T}\sum_{t=1}^{T} g\left(\underline{y}_t, \underline{x}_t, \beta_2\right)\right\|_{q_2}^2,$$

where $dimh = dim\beta_1 = q_1$ and $dimg = dim\beta_2 = q_2$. But in the general case of indirect estimators built on instrumental estimators $\widehat{\beta}_{1,T}$ that cannot be disentangled in the criterion $Q_T\left(\underline{y}_T, \underline{x}_T, \beta\right)$

from $\widehat{\beta}_{2,T}$, there is a need for a well suited asymptotic theory dealing with North-West blocks of $I_\circ, J_\circ$,and $K_\circ$ matrices. This asymptotic theory will be developed in section 4 below in a more general setting[10] where $\widetilde{\beta}_1(\theta)$ depends on θ only through a sub-vector $(\theta_1', \theta_{21}')'$ where $\theta_2 = (\theta_{21}', \theta_{22}')'$.

• Second, it may make sense to think about a reduction of the set θ_2 of nuisance parameters which are really identified from the subset β_1 of selected instrumental parameters. This latter remark should be referred back to a proposal by Gouriéroux and Monfort (1995). They indeed suggest to extend the indirect identification concept to a sub-vector $(\theta_1', \theta_{21}')'$ by relaxing $(A4)$ as follows:

$$\begin{aligned} &\Theta_2 = \Theta_{21} \times \Theta_{22},\ \forall\,(\theta_{22}, \theta_{22}^*) \in \Theta_{22} \times \Theta_{22}, \\ &\widetilde{\beta}\,(\theta_1, \theta_{21}, \theta_{22}) = \widetilde{\beta}\,(\theta_1^*, \theta_{21}^*, \theta_{22}^*) \Longrightarrow (\theta_1, \theta_{21}) = (\theta_1^*, \theta_{21}^*)\cdot \end{aligned} \tag{3.6}$$

In light of this, we propose to revisit Gouriéroux and Monfort (1995) definition in three respects:
⋆ First, since one is only interested in identifying a sub-vector $(\theta_1', \theta_{21}')'$, it may be relevant to select only a subset $\widetilde{\beta}_1(\theta)$ of moments to match: $\widetilde{\beta}(\theta) = \left(\widetilde{\beta}_1'(\theta), \widetilde{\beta}_2'(\theta)\right)'$, as long as it also fulfills the required indirect identification condition (3.6).
⋆ Second, since this identification condition means that the knowledge of $\widetilde{\beta}_1\,(\theta_1, \theta_2)$ provides the knowledge of (θ_1, θ_{21}), it implies that there exists a function g such that:

$$(\theta_1', \theta_{21}')' = g\left(\widetilde{\beta}_1(\theta_1, \theta_2)\right),$$

for any $(\theta_1, \theta_2) \in \Theta_1 \times \Theta_2$. Therefore, up to a reduction by the transformation g of the set of moments to match (or more generally of instrumental parameters), one may consider that $\widetilde{\beta}_1(\theta)$ depends upon θ only through (θ_1, θ_{21}). Therefore, we will often refer in the sequel to the following extension of assumption $(A4)$:

Assumption (A7):

$$\widetilde{\beta}\,(\theta_1, \theta_{21}, \theta_{22}) = \left(\widetilde{\beta}_1'(\theta_1, \theta_{21}), \widetilde{\beta}_2'(\theta_1, \theta_{21}, \theta_{22})\right)',$$

where $\widetilde{\beta}_1(\cdot, \cdot)$ is one-to-one.
⋆ Third, one may imagine to relax the quite restrictive indirect assumption (3.6) by assuming only that:

$$\begin{aligned} &\exists \overline{\theta}_{22} \in \Theta_{22}, \\ &\widetilde{\beta}_1(\theta_1, \theta_{21}, \overline{\theta}_{22}) = \widetilde{\beta}_1(\theta_1^*, \theta_{21}^*, \overline{\theta}_{22}) \Longrightarrow (\theta_1, \theta_{21}) = (\theta_1^*, \theta_{21}^*)\cdot \end{aligned} \tag{3.7}$$

This general setting is genuinely a new one whenever $\overline{\theta}_{22}$ is unknown. The problem in this case is, that, on the one hand the use of the encompassing property for consistent estimation of θ_1 requires a preliminary consistent estimation of $\overline{\theta}_{22}$. But on the other hand, the asymptotic properties of the indirect inference estimator of θ_1 will depend upon the ones of the estimator of $\overline{\theta}_{22}$. The intuitive reason for that is, that the considered estimators for $\overline{\theta}_{22}$ and θ_1 are not asymptotically independent, since the binding function $\widetilde{\beta}_1(\cdot)$ has a non zero derivative with respect to $\overline{\theta}_{22}$. We will address this issue in details in section 5. For sake of clarity, we first develop our semi-parametric indirect inference methodology within either assumption $(A4)$ joint with the encompassing property (3.4) or, alternatively, assumption $(A7)$ joint with a weakened **partial encompassing** property introduced as follows:

[10] Indeed, a particular case of this general framework is the one where $\theta_{21} = \theta_2$.

Definition 3.2 : $(B1)$ *endowed with the true unknown value* θ_1° *partially encompasses* $(\mathcal{N}_\beta)$ *for a sub-vector* β_1 *conformable to assumption* $(A7)$ *if there exists* $\overline{\theta}_{21} \in \Theta_{21}$ *such that:*

$$\beta_1^\circ = \widetilde{\beta}_1\left(\theta_1^\circ, \overline{\theta}_{21}\right) \cdot$$

From now on, we will refer to **full encompassing** as the encompassing property of definition 3.1 in order to distinguish it from **partial encompassing** (definition 3.2).

Note that as far as one is mainly concerned with the estimation of the structural parameters, the crucial issue on partial encompassing is the existence of a sub-vector β_1 conformable to definition 3.2, whichever resulting partition of β. For sake of simplicity, this convenient sub-vector β_1 will be considered hereafter as given even though the trade off consistency versus efficiency should lead to look for the largest set β_1 of components of β which still maintains consistency thanks to partial encompassing. The practical implementation of such a strategy will be discussed in more details in section 4.

Moreover, since in the case $p_{21} < p_2$, θ_{22} is not involved in the sub-vector function $\widetilde{\beta}_1(\cdot,\cdot)$, $\overline{\theta}_2$ is not unique and the condition of definition 3.2 is fulfilled independently of the value $\theta_{22} \in \Theta_{22}$. Of course, this property is not maintained in general for finite sample binding functions. Therefore, we introduce the following estimators $\widehat{\beta}_{1,T}$, $\widetilde{\beta}^s_{1,T}(\theta_1,\theta_2)$ and $\widetilde{\beta}_{1,TS}(\theta_1,\theta_2)$ respectively defined as the sub-vectors of size q_1 of the estimators $\widehat{\beta}_T$, $\widetilde{\beta}^s_T(\theta_1,\theta_2)$ and $\widetilde{\beta}_{TS}(\theta_1,\theta_2)$ defined by (2.14). Under assumptions $(A1)-(A3)$, and $(A7)$, these estimators converge to:

$$P_\circ \lim_{T\to+\infty} \widehat{\beta}_{1,T} = \beta_1^\circ,$$

$$P_* \lim_{T\to+\infty} \widetilde{\beta}^s_{1,T}(\theta_1,\theta_2) = P_* \lim_{T\to+\infty} \widetilde{\beta}_{1,TS}(\theta_1,\theta_2) = \widetilde{\beta}_1(\theta_1,\theta_{21}) \cdot$$

Since the partial-encompassing property, seen as a weakened version of the full-encompassing property that is not fulfilled in this context, does not in general ensure the consistency condition delineated by proposition (3.1), we propose to focus on another class of semi-parametric indirect estimators $\widehat{\theta}^1_{1,TS}(\overline{\theta}_{22})$ based on the sub-vector β_1 of the instrumental parameters and defined by:

$$\begin{aligned} \widehat{\theta}^1_{TS}(\overline{\theta}_{22}) &= \left(\widehat{\theta}^{1'}_{1,TS}(\overline{\theta}_{22}), \widehat{\theta}^{1'}_{21,TS}(\overline{\theta}_{22})\right)' \\ &= \underset{(\theta_1,\theta_{21})\in\Theta_1\times\Theta_{21}}{Argmin} \left[\widehat{\beta}_{1,T} - \widetilde{\beta}_{1,TS}(\theta_1,\theta_{21},\overline{\theta}_{22})\right]' \widehat{\Omega}_{1,T} \left[\widehat{\beta}_{1,T} - \widetilde{\beta}_{1,TS}(\theta_1,\theta_{21},\overline{\theta}_{22})\right], \end{aligned} \tag{3.8}$$

where $P_* \lim_{T\to+\infty} \widehat{\Omega}_{1,T} = \Omega_1$ is a positive matrix on $\mathbb{R}^{q_1}$ and $\overline{\theta}_{22}$ corresponds to the given value of the nuisance parameters θ_{22}.

It is worthwhile noting that in the case where $p_{21} < p_2$, the nuisance parameters θ_{22} are not estimated within this procedure. The issue on the estimation or the calibration of these nuisance parameters are developed in section 5. We denote $\overline{\theta}_{22}$ the value assigned to the nuisance parameters θ_{22}, that is used for performing the simulations.

In this framework, we are able to prove the following sufficient condition for the consistency of the semi-parametric II estimator $\widehat{\theta}^1_{1,TS}(\overline{\theta}_{22})$:

Proposition 3.3 : *Under assumptions* $(A1)-(A7)$ *and if* $(B1)$ *endowed with the true value* θ_1° *partially encompasses* (N_β), *then for any* $\overline{\theta}_{22} \in \Theta_{22}$, $\widehat{\theta}^1_{1,TS}(\overline{\theta}_{22})$ *is a consistent estimator of the parameters of interest* θ_1°.

Proof : *The proof of proposition* 3.3 *is straightforward since under assumptions* $(A1)-(A7)$*:*

$$P_* \lim_{T\to+\infty} \left(\widehat{\theta}^{1'}_{1,TS}(\overline{\theta}_{22}), \widehat{\theta}^{1'}_{21,TS}(\overline{\theta}_{22})\right)' = \underset{(\theta_1,\theta_{21})\in\Theta_1\times\Theta_{21}}{Argmin} \left\|\beta_1^\circ - \widetilde{\beta}_1(\theta_1,\theta_{21},\overline{\theta}_{22})\right\|_{\Omega_1},$$

where $\|\beta_1\|^2_{\Omega_1} = \beta_1'\Omega_1\beta_1$*. We have thanks to the partial-encompassing property:*

$$\begin{aligned} \widetilde{\beta}_1(\theta_1,\theta_2) &= \widetilde{\beta}_1(\theta_1,\theta_{21}), \\ \textit{and}\ \ \beta_1^\circ &= \widetilde{\beta}_1(\theta_1^\circ,\overline{\theta}_{21}), \\ \Longrightarrow\ \left(\theta_1^{\circ\prime},\overline{\theta}_{21}{}'\right)' &= \underset{(\theta_1,\theta_{21})\in\Theta_1\times\Theta_{21}}{Argmin} \left\|\beta_1^\circ - \widetilde{\beta}_1(\theta_1,\theta_{21})\right\|_{\Omega_1}, \\ &= P_* \lim_{T\to+\infty} \left(\widehat{\theta}^{1'}_{1,TS}(\overline{\theta}_{22}), \widehat{\theta}^{1'}_{21,TS}(\overline{\theta}_{22})\right)' \cdot \end{aligned}$$

For sake of simplicity, we have chosen to give a direct proof of proposition 3.3. But it is still pos3ible to see it as a corollary of a general necessary and sufficient condition for the consistency of the semi-parametric indirect inference estimator $\widehat{\theta}^1_{1,TS}(\overline{\theta}_{22})$ as in proposition 3.1. This generalization concerns the case where one is interested in indirect estimation based on a sub-vector β_1 of the instrumental parameters. More precisely, we define on the one hand the "generalized inverse" $\widetilde{\beta}_1^-$ of $\widetilde{\beta}_1$ by:

$$\widetilde{\beta}_1^-(\beta_1) = \underset{(\theta_1,\theta_{21})\in\Theta_1\times\Theta_{21}}{Argmin} \left\|\beta_1 - \widetilde{\beta}_1(\theta_1,\theta_{21})\right\|_{\Omega_1} \cdot$$

$\widetilde{\beta}_1^-$ is a function from $I\!R^{q_1}$ onto $\Theta_1\times\Theta_{21}$. On the other hand, we define $\beta_1(P)$ the sub-vector of size q_1 associated with $\beta(P)$ defined in subsection 3.2 for any probability distributions P delineated by the semi-parametric model $(A1)$; and let us denote Q_1^1 the projection operator:

$$\begin{aligned} Q_1^1\ :\ I\!R^{p_1}\times I\!R^{p_{21}} &\longrightarrow I\!R^{p_1}, \\ (\theta_1{}',\theta_{21}{}')' &\longrightarrow \theta_1 \cdot \end{aligned}$$

Then proposition 3.3 appears as a direct corollary of the following consistency criterion:

Proposition 3.4 : *Under assumptions* $(A1)-(A7)$, $\widehat{\theta}^1_{1,TS}(\overline{\theta}_{22})$ *is a consistent estimator of the parameters of interest* θ_1° *if and only if for any* P *in the family* $\mathcal{P}$ *of probability distributions delineated by the semi-parametric model* $(A1)$*:*

$$Q_1^1\left[\widetilde{\beta}_1^- \circ \beta_1(P)\right] = \widetilde{\theta_1}(P) \cdot$$

Proof : *The proof of proposition* 3.4 *is a simple extension of the proof of proposition* 3.1 *and therefore is omitted here.*

As already mentioned, under the partial-encompassing condition given in proposition 3.3, the semi-parametric indirect inference estimator is consistent whichever value $\overline{\theta}_{22}$ is used for building simulated paths.

More generally speaking, let us stress the two following ideas:

• On the one hand, the encompassing interpretation of these consistency conditions: either full-encompassing property or more generally partial-encompassing property, will be useful as far as the associated Wald encompassing test (WET) à la Mizon and Richard (1986) and its simulated version à la Gouriéroux and Monfort (1995) will provide a test of consistency (see section 4 below).

• On the other hand, it is worth mentioning that the standard principles of encompassing and WET have been slightly extended here to take into account our focus of interest, that is consistent indirect estimation through a mis-specified structural model used as a simulator. More precisely, while the standard encompassing principle was introduced by Mizon and Richard (1986) to stress that a given model, even mis-specified, produces relevant estimators as soon as it encompasses its non nested competitor, our generalized encompassing principle explains that a mis-specified structural model may produce a relevant calibration as soon as when endowed with the true value of interest θ_1°, it "encompasses" either fully or partially the moments to match.

Moreover, it is important keeping in mind that, in our general setting, the pseudo-true value of interest $\left(\theta_1^{\circ\prime}, \overline{\theta}_{21}{}'\right)'$ does not admit an intrinsic characterization but is itself estimated (under the null of the WET) through an indirect inference. In other words, our encompassing definition focuses:

$$\text{either on } \underset{(\theta_1,\theta_2)\in\Theta_1\times\Theta_2}{Min} \left\| \beta^\circ - \widetilde{\beta}(\theta_1, \theta_2) \right\|_\Omega \text{ for the full-encompassing property,}$$

$$\text{or on } \underset{(\theta_1,\theta_{21})\in\Theta_1\times\Theta_{21}}{Min} \left\| \beta_1^\circ - \widetilde{\beta}_1(\theta_1, \theta_{21}) \right\|_{\Omega_1} \text{ for the partial-encompassing property,}$$

rather than on $\beta^\circ - \widetilde{\beta}(\theta_1^*, \theta_2^*)$, for an a priori definition $\left(\theta_1^{*\prime}, \theta_2^{*\prime}\right)'$.

4 Mis-specified Structural Models and Indirect Inference

In this section we deduce the main asymptotic results, whose proofs are given in appendices 2 and 3 for both full-encompassing and partial-encompassing properties taken as null hypothesis.

4.1 Asymptotic probability distribution of semi-parametric indirect inference estimators

4.1.1 Full-encompassing semi-parametric indirect inference estimator

We focus here on the asymptotic properties of the indirect inference estimator $\widehat{\theta}_{TS}$ under the full-encompassing hypothesis $H_\circ$ according to definition 3.1. Moreover, we assume that both the true unknown value θ_1°, the pseudo-true value $\overline{\theta}_2$ used for the encompassing property and the true unknown value β° belong to the interior of the corresponding sets: $\left(\theta_1^{\circ\prime}, \overline{\theta}_2{}'\right)' \in \overset{\circ}{\Theta}$, $\beta^\circ \in \overset{\circ}{\mathcal{B}}$. We assume in addition that:

Assumption (A8):

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

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

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

(c)
$$\lim_{T\to+\infty} \underset{*}{Cov}\left\{\sqrt{T}\frac{\partial Q_T}{\partial \beta}\left(\underline{y}_T, \underline{x}_T, \beta^\circ\right),\ \sqrt{T}\frac{\partial Q_T}{\partial \beta}\left(\underline{\tilde{y}}_T^s(\theta_1^\circ, \overline{\theta}_2, z_\circ^s), \underline{x}_T, \beta^\circ\right)\right\} = K_\circ,$$

independent of the initial values $z_\circ^s,\ s = 1, \ldots, S$.

(d)
$$\sqrt{T}\frac{\partial Q_T}{\partial \beta}\left(\underline{\tilde{y}}_T^s(\theta_1^\circ, \overline{\theta}_2, z_\circ^s), \underline{x}_T, \beta^\circ\right),$$

is asymptotically normally distributed[11] with mean zero and with an asymptotic covariance matrix $I_\circ^*$ and independent of the initial values $z_\circ^s,\ s = 1, \ldots, S$.

(e)
$$J_\circ^* = \underset{T\to+\infty}{P_* \lim} \frac{\partial^2 Q_T}{\partial \beta \partial \beta'}\left(\underline{\tilde{y}}_T^s(\theta_1^\circ, \overline{\theta}_2, z_\circ^s), \underline{x}_T, \beta^\circ\right),$$

[11] Actually, we assume that the joint probability distribution of the two score vectors considered in assumptions $(A8a)$ and $(A8d)$ is asymptotically normal. Strictly speaking, this is ensured by the conjunction of $(A8a)$ and $(A8d)$ only when there is no x variable. In the general case, it might be a slightly more restrictive assumption than the conjunction of $(A8a)$ and $(A8d)$.

independent of the initial values $z_\circ^s$, $s = 1, \dots, S$.

$$(f) \qquad \lim_{T \to +\infty} \underset{*}{Cov} \left\{ \sqrt{T} \frac{\partial Q_T}{\partial \beta} \left(\underline{\tilde{y}}_T^s(\theta_1^\circ, \overline{\theta}_2, z_\circ^s), \underline{x}_T, \beta^\circ \right), \ \sqrt{T} \frac{\partial Q_T}{\partial \beta} \left(\underline{\tilde{y}}_T^\ell(\theta_1^\circ, \overline{\theta}_2, z_\circ^\ell), \underline{x}_T, \beta^\circ \right) \right\} = K_\circ^*,$$

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

$$(g) \qquad P_* \lim_{T \to +\infty} \frac{\partial \tilde{\beta}_T^s}{\partial \theta'} \left(\theta_1^\circ, \overline{\theta}_2 \right) = \frac{\partial \tilde{\beta}}{\partial \theta'} (\theta_1^\circ, \overline{\theta}_2),$$

is of full-column rank (p).

Note that in general, $I_\circ \neq I_\circ^*$, $J_\circ \neq J_\circ^*$ and $K_\circ \neq K_\circ^*$ since the structural model $(B1)$ is mis-specified. Nonetheless these equalities are fulfilled in the well-specified case. We are then able to prove the following result:

Proposition 4.1 : *Under assumptions $(A1) - (A6)/(A8)$ and the null hypothesis $H_\circ$, the indirect inference estimator $\widehat{\theta}_{TS}$ is asymptotically normal, when S is fixed and T goes to infinity:*

$$\sqrt{T} \begin{pmatrix} \widehat{\theta}_{1,TS} - \theta_1^\circ \\ \widehat{\theta}_{2,TS} - \overline{\theta}_2 \end{pmatrix} \xrightarrow[T \to +\infty]{D} \mathcal{N}\left(0, W(S, \Omega)\right),$$

where:

$$W(S,\Omega) = \left\{ \frac{\partial \tilde{\beta}'}{\partial \theta}(\theta_1^\circ, \overline{\theta}_2) \Omega \frac{\partial \tilde{\beta}}{\partial \theta'}(\theta_1^\circ, \overline{\theta}_2) \right\}^{-1} \frac{\partial \tilde{\beta}'}{\partial \theta}(\theta_1^\circ, \overline{\theta}_2) \Omega \Phi_\circ^*(S) \Omega \frac{\partial \tilde{\beta}}{\partial \theta'}(\theta_1^\circ, \overline{\theta}_2) \left\{ \frac{\partial \tilde{\beta}'}{\partial \theta}(\theta_1^\circ, \overline{\theta}_2) \Omega \frac{\partial \tilde{\beta}}{\partial \theta'}(\theta_1^\circ, \overline{\theta}_2) \right\}^{-1},$$

and with:

$$\Phi_\circ^*(S) = J_\circ^{-1} I_\circ J_\circ^{-1} + \frac{1}{S} J_\circ^{*-1} I_\circ^* J_\circ^{*-1} + \left(1 - \frac{1}{S}\right) J_\circ^{*-1} K_\circ^* J_\circ^{*-1} - J_\circ^{-1} K_\circ J_\circ^{*-1} - J_\circ^{*-1} K_\circ' J_\circ^{-1}. \qquad (4.1)$$

Proof : *see appendix A.2.*

Note that in the case where the structural model $(B1)$ is well-specified, $\Phi_\circ^*(S)$ reduces to the expression $\left(1 + \frac{1}{S}\right) J_\circ^{-1} \left(I_\circ - K_\circ\right) J_\circ^{-1}$ since $K_\circ' = K_\circ$.

The asymptotic covariance matrix depends on the metric Ω and as usual, there exists an optimal choice of the weighting matrix $\Omega^*(S)$ which minimizes $W(S, \Omega)$.

Proposition 4.2 : *Under assumptions $(A1) - (A6)/(A8)$ and the null hypothesis $H_\circ$, the optimal choice $\Omega^*(S)$ of Ω for the indirect inference estimator $\widehat{\theta}_{TS}$ is given by $\Omega^*(S) = \Phi_\circ^*(S)^{-1}$ (assuming that $\Phi_\circ^*(S)$ is non singular). The asymptotic covariance matrix is then given by:*

$$W_S^* = W(S, \Omega^*(S)) = \left\{ \frac{\partial \tilde{\beta}'}{\partial \theta}(\theta_1^\circ, \overline{\theta}_2) (\Phi_\circ^*(S))^{-1} \frac{\partial \tilde{\beta}}{\partial \theta'}(\theta_1^\circ, \overline{\theta}_2) \right\}^{-1}. \qquad (4.2)$$

Proof : *see appendix A.2.*

4.1.2 Partial-encompassing semi-parametric indirect inference estimator

We now focus on the asymptotic properties of the indirect inference estimator $\widehat{\theta}^1_{TS}(\overline{\theta}_{22})$ under the partial encompassing hypothesis $H^1_\circ$ according to definition 3.2 for a pseudo-true value $\left(\theta_1^{\circ\prime}, \overline{\theta}_{21}'\right)'$. We first maintain assumption $(A1)-(A8b)$ and we denote $\widetilde{\beta}^\circ(\overline{\theta}_{22}) = \widetilde{\beta}(\theta^\circ_1, \overline{\theta}_2)$ for the given value $\overline{\theta}_{22}$ of the nuisance parameters. We assume that $\left(\theta_1^{\circ\prime}, \overline{\theta}_2'\right)' \in \overset{\circ}{\Theta}$, $\beta^\circ \in \overset{\circ}{\mathcal{B}}$, $\widetilde{\beta}^\circ\left(\overline{\theta}_{22}\right) \in \overset{\circ}{\mathcal{B}}$ and in addition that:

Assumption (A9):

(a)
$$\lim_{T\to+\infty} \underset{*}{Cov}\left\{\sqrt{T}\frac{\partial Q_T}{\partial\beta}\left(\underline{y}_T, \underline{x}_T, \beta^\circ\right),\ \sqrt{T}\frac{\partial Q_T}{\partial\beta}\left(\underline{\widetilde{y}}^s_T(\theta^\circ_1, \overline{\theta}_2, z^s_\circ), \underline{x}_T, \widetilde{\beta}^\circ(\overline{\theta}_{22})\right)\right\} = K_\circ(\overline{\theta}_{22}),$$

independent of the initial values $z^s_\circ,\ s = 1, \ldots, S$ and for the given value $\overline{\theta}_{22}$.

(b)
$$\sqrt{T}\frac{\partial Q_T}{\partial\beta}\left(\underline{\widetilde{y}}^s_T(\theta^\circ_1, \overline{\theta}_2, z^s_\circ), \underline{x}_T, \widetilde{\beta}^\circ(\overline{\theta}_{22})\right),$$

is asymptotically normally distributed[12] with mean zero and with an asymptotic covariance matrix $I^*_\circ(\overline{\theta}_{22})$ and independent of the initial values $z^s_\circ,\ s = 1, \ldots, S$ and for the given value $\overline{\theta}_{22}$.

(c)
$$J^*_\circ(\overline{\theta}_{22}) = \underset{T\to+\infty}{P_*\lim}\frac{\partial^2 Q_T}{\partial\beta\partial\beta'}\left(\underline{\widetilde{y}}^s_T(\theta^\circ_1, \overline{\theta}_2, z^s_\circ), \underline{x}_T, \widetilde{\beta}^\circ(\overline{\theta}_{22})\right),$$

independent of the initial values $z^s_\circ,\ s = 1, \ldots, S$ and for the given value $\overline{\theta}_{22}$.

(d)
$$\lim_{T\to+\infty} \underset{*}{Cov}\left\{\sqrt{T}\frac{\partial Q_T}{\partial\beta}\left(\underline{\widetilde{y}}^s_T(\theta^\circ_1, \overline{\theta}_2, z^s_\circ), \underline{x}_T, \widetilde{\beta}^\circ(\overline{\theta}_{22})\right),\ \sqrt{T}\frac{\partial Q_T}{\partial\beta}\left(\underline{\widetilde{y}}^\ell_T(\theta^\circ_1, \overline{\theta}_2, z^\ell_\circ), \underline{x}_T, \widetilde{\beta}^\circ(\overline{\theta}_{22})\right)\right\} = K^*_\circ(\overline{\theta}_{22}),$$

independent of the initial values $z^s_\circ$ and $z^\ell_\circ$, for $s \neq \ell$ and for the given value $\overline{\theta}_{22}$.

(e)
$$\underset{T\to+\infty}{P_*\lim}\frac{\partial\widetilde{\beta}^s_{1,T}}{\partial\begin{pmatrix}\theta_1\\ \theta_{21}\end{pmatrix}'}\left(\theta^\circ_1, \overline{\theta}_2\right) = \frac{\partial\widetilde{\beta}_1}{\partial\begin{pmatrix}\theta_1\\ \theta_{21}\end{pmatrix}'}(\theta^\circ_1, \overline{\theta}_{21}),$$

is of full-column rank $(p_1 + p_{21})$. We are then able to prove the following result:

Proposition 4.3 : *Under assumptions* $(A1)-(A8b)/(A9)$ *and the null hypothesis* $H^1_\circ$*, the indirect inference estimator* $\widehat{\theta}^1_{TS}(\overline{\theta}_{22})$ *is asymptotically normal, when* S *is fixed and* T *goes to infinity:*

$$\sqrt{T}\begin{pmatrix}\widehat{\theta}^1_{1,TS}(\overline{\theta}_{22}) - \theta^\circ_1\\ \widehat{\theta}^1_{21,TS}(\overline{\theta}_{22}) - \overline{\theta}_{21}\end{pmatrix} \xrightarrow[T\to+\infty]{D} \mathcal{N}\left(0, W_1(S, \Omega_1, \overline{\theta}_{22})\right),$$

[12] As already pointed out in the previous footnote, there is a need of joint normal asymptotic distribution for the two score vectors respectively introduced by $(A8a)$ and $(A9b)$.

where:

$$W_1(S,\Omega_1,\overline{\theta}_{22}) = \left\{ \frac{\partial \widetilde{\beta}_1'}{\partial \begin{pmatrix} \theta_1 \\ \theta_{21} \end{pmatrix}} (\theta_1^\circ,\overline{\theta}_{21}) \Omega_1 \frac{\partial \widetilde{\beta}_1}{\partial \begin{pmatrix} \theta_1 \\ \theta_{21} \end{pmatrix}'} (\theta_1^\circ,\overline{\theta}_{21}) \right\}^{-1} \frac{\partial \widetilde{\beta}_1'}{\partial \begin{pmatrix} \theta_1 \\ \theta_{21} \end{pmatrix}} (\theta_1^\circ,\overline{\theta}_{21}) \Omega_1 \Phi_{\circ,1}^*(S,\overline{\theta}_{22})$$
$$\Omega_1 \frac{\partial \widetilde{\beta}_1}{\partial \begin{pmatrix} \theta_1 \\ \theta_{21} \end{pmatrix}'} (\theta_1^\circ,\overline{\theta}_{21}) \left\{ \frac{\partial \widetilde{\beta}_1'}{\partial \begin{pmatrix} \theta_1 \\ \theta_{21} \end{pmatrix}} (\theta_1^\circ,\overline{\theta}_{21}) \Omega_1 \frac{\partial \widetilde{\beta}_1}{\partial \begin{pmatrix} \theta_1 \\ \theta_{21} \end{pmatrix}'} (\theta_1^\circ,\overline{\theta}_{21}) \right\}^{-1},$$

and $\Phi_{\circ,1}^*(S,\overline{\theta}_{22})$ is the $(q_1 \times q_1)$ left-upper bloc diagonal sub-matrix of the $(q \times q)$ matrix $\Phi_\circ^*(S,\overline{\theta}_{22})$ defined by:

$$\begin{aligned} \Phi_\circ^*(S,\overline{\theta}_{22}) = \; & J_\circ^{-1} I_\circ J_\circ^{-1} + \frac{1}{S} J_\circ^{*-1}(\overline{\theta}_{22}) I_\circ^*(\overline{\theta}_{22}) J_\circ^{*-1}(\overline{\theta}_{22}) + \left(1 - \frac{1}{S}\right) J_\circ^{*-1}(\overline{\theta}_{22}) K_\circ^*(\overline{\theta}_{22}) J_\circ^{*-1}(\overline{\theta}_{22}) \\ & - J_\circ^{-1} K_\circ(\overline{\theta}_{22}) J_\circ^{*-1}(\overline{\theta}_{22}) - J_\circ^{*-1}(\overline{\theta}_{22}) K_\circ'(\overline{\theta}_{22}) J_\circ^{-1}. \end{aligned} \tag{4.3}$$

Proof : *see appendix A.3.*

The asymptotic covariance matrix depends on the metric Ω_1 and as usual, there exists an optimal choice of the weighting matrix $\Omega_1^*(S,\overline{\theta}_{22})$ which minimizes $W_1(S,\Omega_1,\overline{\theta}_{22})$.

Proposition 4.4 : *Under assumptions (A1)−(A8b)/(A9) and the null hypothesis $H_\circ^1$, the optimal choice $\Omega_1^*(S,\overline{\theta}_{22})$ of Ω_1 for the indirect inference estimator $\widehat{\theta}_{TS}^1\left(\overline{\theta}_{22}\right)$ is given by $\Omega_1^*(S,\overline{\theta}_{22}) = \Phi_{\circ,1}^*(S,\overline{\theta}_{22})^{-1}$ (assuming that $\Phi_{\circ,1}^*(S,\overline{\theta}_{22})$ is non singular). The asymptotic covariance matrix is then given by:*

$$W_{1,S}^*(\overline{\theta}_{22}) = W_1(S,\Omega_1^*(S,\overline{\theta}_{22}),\overline{\theta}_{22}) = \left\{ \frac{\partial \widetilde{\beta}_1'}{\partial \begin{pmatrix} \theta_1 \\ \theta_{21} \end{pmatrix}} (\theta_1^\circ,\overline{\theta}_{21}) \left(\Phi_{\circ,1}^*(S,\overline{\theta}_{22})\right)^{-1} \frac{\partial \widetilde{\beta}_1}{\partial \begin{pmatrix} \theta_1 \\ \theta_{21} \end{pmatrix}'} (\theta_1^\circ,\overline{\theta}_{21}) \right\}^{-1}. \tag{4.4}$$

Proof : *see appendix A.3.*

It is worthwhile to notice that propositions 4.3 and 4.4 above are not simple applications of propositions 4.1 and 4.2 to the case where the vector of structural parameters is reduced to $(\theta_1',\theta_{21}')'$ (for a given $\overline{\theta}_{22}$) and the vector of instrumental parameters is reduced to β_1. Indeed, the full set of instrumental parameters $\beta = (\beta_1',\beta_2')'$ enters the instrumental criterion in such a way that the direct estimation of β_1 and the corresponding indirect estimation of $(\theta_1',\theta_{21}')'$ cannot be easily disentangled with the evaluation of β and the corresponding θ. Of course, there are several particular circumstances where such a disentangling is straightforward ; this is the case for instance if β defines a list of just-identified and separable moment conditions.

4.2 A Wald Encompassing Test

Gouriéroux, Monfort and Renault (1993) and Gallant and Tauchen (1996) have proposed a global specification test about the structural model based on the optimal value of the objective function used in the second step of the indirect estimation method. But, similarly to the direct Pseudo Maximum Likelihood inference à la Gouriéroux, Monfort and Trognon (1984), there is a need of robustified test statistics to deal with the case where estimators are consistent despite mis-specification. Basically, one should take into account the potential discrepancy between the matrices $(I_\circ, J_\circ, K_\circ)$ and $(I_\circ^*, J_\circ^*, K_\circ^*)$ stressed in the previous subsection. Our robustified global specification test is then defined as follows:

Proposition 4.5 : *Under assumptions* $(A1)-(A6)/(A8)$ *and the null-hypothesis* $H_\circ$ *of full-encompassing of* $(\mathcal{N}_\beta)$ *by* $(B1)$ *according to definition* 3.1*, the statistic* $\xi_{T,S}$*:*

$$\xi_{T,S} = T \min_{(\theta_1,\theta_2)\in\Theta_1\times\Theta_2} \left[\widehat{\beta}_T - \frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}_T^s(\theta_1,\theta_2)\right]' \widehat{\Omega}_T^*(S) \left[\widehat{\beta}_T - \frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}_T^s(\theta_1,\theta_2)\right], \tag{4.5}$$

where $\widehat{\Omega}_T^*(S)$ *is a consistent estimator of the optimal metric* $\Omega^*(S) = \Phi_\circ^*(S)^{-1}$ *defined by* (4.1)*, is asymptotically distributed as a chi-square with* $(q-p)$ *degrees of freedom where* $q = dim\beta$ *and* $p = dim\theta$*.*

Proof : *see appendix A.*2.

Therefore, a natural specification test of asymptotic level α is associated with the critical region:

$$\mathcal{W}_\alpha = \{\xi_{T,S} > \chi^2_{1-\alpha}(q-p)\}.$$

One may recognize the expression of the so-called Simulated GET test (Generalized Encompassing Test) proposed by Gouriéroux and Monfort (1995). However the differences between the two approaches are two-fold:

⋆ On the one hand, while Gouriéroux and Monfort (1995) focuses on the comparison between two parametric models which may be both mis-specified, we do refer to a true unknown DGP and associated true unknown parameters of interest θ_1°.

⋆ On the other hand, we consider this testing procedure solely as the first step of a specification strategy which involves several additional testing procedures about partial encompassing and Hausman type specification tests.

More precisely, while the test procedure provided by proposition 4.5 robustifies the Gouriéroux, Monfort and Renault (1993) specification test by controlling the level in case of mis-specification which does not prevent the SII estimator from being consistent, of course our robustified metric does not produce a consistent test when the Gouriéroux, Monfort and Renault (1993) one does not. This is the reason why we propose the following diagnostic methodology in the two cases where the above test respectively leads one to the rejection or to the acceptance of $H_\circ$ (full encompassing).

First case: rejection of $\mathbf{H_\circ}$.

The rejection of $H_\circ$ means that the set of moments which are matched through the instrumental vector of parameters β is too large and captures the mis-specification of the structural model. Facing the evidence of mis-specification of her structural model, the econometrician usually looks for a larger model, through a battery of now standard specification tools (relaxing restrictions with diagnostics, using model choice criteria, encompassing theory, semi-non-parametric expansion...). But the main point we want to stress here is that our Semi-parametric Indirect Inference (SII) approach provides an alternative solution by keeping the mis-specified structural model and looking for a fine tuning of the instrumental model. The basic idea is that "given that the (structural) model is false" the instrumental model $Q_T\left(\underline{y}_T, \underline{x}_T, \beta\right)$ should be examined only in the "dimension" β_1. In this case, the estimation of the structural parameters of interest will not be contaminated by the mis-specification because, while the value of (3.1) is not zero, the partial encompassing condition $\beta_1^\circ = \widetilde{\beta}_1(\theta_1^\circ, \overline{\theta}_{21})$ (see definition 3.2) is fulfilled for a convenient $\overline{\theta}_{21}$. In other words, fine tuning means looking for a reduction of β through partial encompassing tests, in an ascending procedure. Of course, ascending means here reducing the set of instrumental parameters which reduces the number of calibrated features in a general sense: either β defines a list of moments to match and β_1 is a well-suited sub-list or more generally, the occurrence of the characteristics β_2 of observed and simulated paths in $Q_T\left(\underline{y}_T, \underline{x}_T, \beta\right)$ are neutralized in the sense of subsection 3.2. Therefore we define a partial encompassing test, viewed as a robustified specification test as follows:

Proposition 4.6 : *Under assumptions* $(A1) - (A8b)/(A9)$ *and the null-hypothesis* $H_\circ^1$ *of partial encompassing of* $(\mathcal{N}_\beta)$ *by* $(B1)$ *according to definition* 3.2, *the statistic* $\xi_{T,S}^1(\overline{\theta}_{22})$*:*

$$\xi_{T,S}^1(\overline{\theta}_{22}) = T \min_{(\theta_1,\theta_{21})\in\Theta_1\times\Theta_{21}} \left[\widehat{\beta}_{1,T} - \frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}_{1,T}^s(\theta_1,\theta_{21},\overline{\theta}_{22})\right]' \widehat{\Omega}_{1,T}^*(S) \left[\widehat{\beta}_{1,T} - \frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}_{1,T}^s(\theta_1,\theta_{21},\overline{\theta}_{22})\right], \tag{4.6}$$

where $\widehat{\Omega}_{1,T}^*(S)$ *is a consistent estimator of the optimal metric* $\Omega_1^*(S,\overline{\theta}_{22}) = \Phi_{\circ,1}^*(S,\overline{\theta}_{22})^{-1}$ *defined by* (4.3), *is asymptotically distributed as a chi-square with* $(q_1 - p_1 - p_{21})$ *degrees of freedom where* $q_1 = dim\beta_1,\ p_1 = dim\theta_1,\ p_{21} = dim\theta_{21}$.

Proof : *see appendix A.*3.

Therefore, the associated specification test of asymptotic level α is defined by the following critical region:

$$\mathcal{W}_\alpha^1 = \{\xi_{T,S}^1(\overline{\theta}_{22}) > \chi_{1-\alpha}^2\,(q_1 - p_1 - p_{21})\}.$$

Roughly speaking, the goal of our ascending approach is to look for the largest possible sub-vector β_1 of β leading to an acceptance of $H_\circ^1$ (partial encompassing). In the latter case, one is led to deal with a similar issue as discussed below (second case: acceptance of $H_\circ$). The only unsolved case would correspond to the one where any trial run of particular partial encompassing would lead to reject. In such a case the pair (structural model, instrumental model) is inadequate to produce consistent estimators of the structural parameters of interest. Therefore, this pair has to be modified. Typically, this should imply the specification of a new structural model (either a larger one or a non-nested one) which in turn will suggest in general some modifications on the instrumental model. Indeed, the main message of our semi-parametric encompassing concept is precisely the duality of the two model choices: since consistency is ensured when the structural model encompasses the instrumental one, the two choices should be coordinated to get consistent estimators. In other words, the instrumental model should remain true to the right structural ideas and not be a large set of moments prompted by an automatic statistical process, like for instance semi-non-parametric expansion.

Second case: acceptance of $\mathbf{H}_{\circ}$.

As already pointed out, our robustified specification testing strategy is not consistent when the Gouriéroux, Monfort and Renault (1993) global specification test is not. More precisely, a fallacious acceptance of $H_{\circ}$ might be produced by the already mentioned pitfall which is specific to simulation-based inference, when data simulated from a wrong DGP give a spurious perfect fit: $\theta^{*} = (\theta_1^{*\prime}, \theta_2^{*\prime})^{\prime}$ solution to $\beta(P_{\circ}) = \widetilde{\beta}(\theta_1^{*}, \theta_2^{*})$ although $\theta_1^{*} \neq \theta_1^{\circ}$. Indeed, Tauchen (1997) was the first to emphasize on the fact that the specification test of the Indirect Inference (or equivalently efficient moment matching à la Gallant and Tauchen (1996)) has no power against some alternatives. He actually concludes that "without very strong a priori knowledge, the only way to avoid this situation is to take a flexible, more non-parametric approach to the specification of the score generator".

Our version of this point of view consists here in enlarging the vector β of instrumental parameters to check that acceptance of the encompassing hypothesis $H_{\circ}$ is maintained. Typically, a semi-parametric score generator produces a vector β whose size grows infinitely with the sample size. However, we would like to mitigate this point of view in three respects.

First, the point we take in this paper is that any structural model is mis-specified and therefore "does miserably" in various dimensions, particularly because it is necessary a joint hypothesis about a hopefully well-founded Economic Theory but also an unfortunately ad hoc statistical specification. Thus a sufficiently "flexible more non-parametric approach" will always succeed in rejecting the structural model. But one should keep in mind that the focus of interest is not really the specification error of our structural model but the consistency of the estimation of the structural parameters. Therefore, the quest for a larger instrumental model able to prove that the structural model is mis-specified is in some circumstances irrelevant. These circumstances are delineated as the set of possible DGP which produce consistent SII (for a given instrumental model) despite mis-specification. These DGP are in the implicit null hypothesis of the test (namely the encompassing assumption $H_{\circ}$) while they are in the alternative with a Tauchen (1997) point of view.

Our second point is that the choice of the instrumental model should help the econometrician to answer the fundamental question "How to live with mis-specification if you must?" (Maasoumi (1990)). The problem is that, even though one acknowledges only some stylized facts as pointed out for instance by Bansal, Gallant, Hussey and Tauchen (1995): "an equilibrium model is too smooth to produce realistic nonlinearity at the weekly frequency", nobody suggests to abandon the equilibrium models. By the way, the same authors conclude their article by noticing that "the findings about an equilibrium model being too smooth left the reader alone in front of the central question of the usefulness of the structural model, if one excludes the possibility of isolating a few selected dimensions along which it does well and along which it could be used". In order to isolate such dimensions, one should precisely think about the structural interpretation of the instrumental parameters. This may enter in conflict with the objective of using a sufficiently "flexible more non-parametric approach" by automatic expansion of the instrumental model.

Third if one is afraid of accepting $H_\circ$ within the second pitfall, that is a fallacious perfect fit and $\theta_1^* \neq \theta_1^\circ$, it is always possible to question this value of θ_1^* by the following trial run. Let us imagine that we have at our disposal several candidates of pair (structural model, instrumental model). One may perform SII and the corresponding encompassing test for these various pairs. Even if all these tests lead to accept the null hypothesis of consistency of the SII estimator, it is very likely that in case of spurious fit, some different values of θ_1^* will appear. This is a well-adapted warning about the possibly zero power drawback of the encompassing test. Indeed, one may even build Hausman type tests about this issue.

5 Semi-parametric Indirect Inference with Nuisance Parameters

We address in this section the issue on the choice of the value $\overline{\theta}_{22}$ of the nuisance parameters θ_{22} and its estimation. We first wish to emphasize on the fact that this question has to be addressed with two very different perspectives, namely:

• Either the encompassing property as defined by definition 3.2 allows for an arbitrary choice of $\overline{\theta}_{22}$. In this case, thanks to this additional degree of freedom, one has to assess a value according to some extra-specified loss function.

• Or, the crucial indirect identification assumption requires, according to (3.7), the use of a precise value $\overline{\theta}_{22}$ of θ_{22}. Moreover, as already explained, this value is unknown and has to be consistently estimated in a first step. We will address respectively in the two subsections below these two different issues.

5.1 Case of innocuous nuisance parameters

Under the partial encompassing property as introduced in definition 3.2, our SII methodology provides a consistent estimator $\widehat{\theta}^1_{TS}(\overline{\theta}_{22})$ to the pseudo-true value of interest $\left(\theta_1^{\circ\prime}, \overline{\theta}_{21}{}'\right)'$ and is asymptotically $\sqrt{T}$-normal as shown in propositions 3.3, 4.3, 4.4. Moreover, the statistics $\xi^1_{T,S}(\overline{\theta}_{22})$ is asymptotically distributed as a chi-square distribution with $q_1 - p_1 - p_{21}$ degrees of freedom as laid out in proposition 4.6. We want to stress here that, in case of innocuous nuisance parameters, these results remain unchanged whenever $\overline{\theta}_{22}$ is replaced by a consistent estimator $\widehat{\theta}_{22,TS}$ such that $\sqrt{T}\left(\widehat{\theta}_{22,TS} - \overline{\theta}_{22}\right) = O_{P_*}(1)$[13] (see appendices $A.1$ and $A.3$ for the proofs). Hence the question about the evaluation of the nuisance parameters θ_{22} remains, especially, when the full-encompassing property is not fulfilled so that the joint indirect estimation of $(\theta_1', \theta_{21}')'$ with that of the nuisance parameters θ_{22} leads to an inconsistent indirect inference estimator $\widehat{\theta}_{1,TS}$.

In this respect, as long as, under the partial encompassing assumption (definition 3.2), the consistency of the SII estimator $\widehat{\theta}^1_{1,TS}(\overline{\theta}_{22})$ does not depend on the value $\overline{\theta}_{22}$ (or more generally on the estimator $\widehat{\theta}_{22,TS}$), we are able to say that, with respect to the consistency of the SII estimator $\widehat{\theta}^1_{1,TS}(\overline{\theta}_{22})$, one can set, a priori, whichever value $\overline{\theta}_{22}$ he wishes to impose on the nuisance parameters θ_{22}. But it is clear that other features generated by the structural model $(B1)$ do depend on the value $\overline{\theta}_{22}$, for instance the asymptotic probability distribution of the SII estimator $\widehat{\theta}^1_{1,TS}(\overline{\theta}_{22})$ (for statistical considerations), or the dimensions β_2 of the instrumental model (for more structural considerations). In this respect, one always has in mind to assign a value $\overline{\theta}_{22}$ that minimizes some desired general loss function $\delta(P_\circ, \theta_{22})$ that is:

$$\overline{\theta}_{22} = \underset{\theta_{22} \in \Theta_{22}}{Argmin}\, \delta(P_\circ, \theta_{22}), \tag{5.1}$$

[13] Actually in case of innocuous nuisance parameters, the consistency of the SII estimator $\widehat{\theta}_{1,TS}$ requires the weak consistency of the nuisance parameters: $P_* \lim_{T \to +\infty} \widehat{\theta}_{22,TS} = \overline{\theta}_{22}$ and the asymptotic normal distribution of the SII estimator $\widehat{\theta}_{1,TS}$ requires that $\sqrt{T}\left(\widehat{\theta}_{22,TS} - \overline{\theta}_{22}\right) = O_{P_*}(1)$.

where $P_\circ$ corresponds to the true unknown probability distribution of $\{(y_t, x_t), t \in \mathbb{Z}\}$ according to $(A1)$. In practice, the nuisance parameters $\overline{\theta}_{22}$ will be consistently estimated by the estimator $\widehat{\theta}_{22,TS}$ defined as follows:

$$\widehat{\theta}_{22,TS} = \underset{\theta_{22}\in\Theta_{22}}{Argmin}\, \delta_{TS}(\underline{y}_T, \underline{x}_T, \theta_{22}), \tag{5.2}$$

where $P_* \lim_{T\to+\infty} \underset{\theta_{22}\in\Theta_{22}}{Sup} \left|\delta_{TS}(\underline{y}_T, \underline{x}_T, \theta_{22}) - \delta(P_\circ, \theta_{22})\right| = 0$. $\widehat{\theta}_{22,TS}$ is therefore consistent to the value $\overline{\theta}_{22}$ of the nuisance parameters θ_{22} and is assumed to be such that $\sqrt{T}\left(\widehat{\theta}_{22,TS} - \overline{\theta}_{22}\right) = O_{P_*}(1)$. The subscript S means here, that when necessary, one can build simulated paths of the endogenous variables, thanks to the structural model $(B1)$, to compute the estimated loss function $\delta_{TS}(\underline{y}_T, \underline{x}_T, \theta_{22})$. For instance, as can be seen in the sequel, we may consider the following estimated loss function:

$$\delta_{TS}(\underline{y}_T, \underline{x}_T, \theta_{22}) = \left\|\psi_T(\underline{y}_T, \underline{x}_T) - \frac{1}{S}\sum_{s=1}^{S} \psi_T(\underline{\widetilde{y}}_T^s(\widehat{\theta}_{TS}^1(\theta_{22}^*), \theta_{22}), \underline{x}_T)\right\|_{p(\Psi)}, \tag{5.3}$$

where:

$$\begin{aligned} P_\circ \lim_{T\to+\infty} \left[\psi_T(\underline{y}_T, \underline{x}_T) - \psi_\infty(P_\circ)\right] &= 0, \\ P_* \lim_{T\to+\infty} \underset{\theta\in\Theta}{Sup} \left\|\psi_T(\underline{\widetilde{y}}_T^s(\theta), \underline{x}_T) - \psi_\infty(P_\theta)\right\|_{p(\Psi)} &= 0\cdot \end{aligned}$$

ψ_∞ is an operator defined from the set $\mathcal{P}$ of probability distributions on $(\mathcal{X}\times\mathcal{Y})^{\mathbb{Z}}$ onto $\mathbb{R}^{p(\psi)}$. θ_{22}^* is some initial values assigned to the nuisance parameters θ_{22} in order to produce in a first step estimation a consistent SII estimator $\widehat{\theta}_{TS}^1(\theta_{22}^*)$ of the pseudo-true value $\left(\theta_1^{\circ\prime}, \overline{\theta}_{21}{}'\right)'$. $\|\cdot\|_{p(\Psi)}$ is some norm on $\mathbb{R}^{p(\psi)}$.

In this case the loss function $\delta(P_\circ, \theta_{22})$ is simply defined by:

$$\delta(P_\circ, \theta_{22}) = \left\|\psi_\infty(P_\circ) - \psi_\infty(P_{(\theta_1^{\circ\prime}, \overline{\theta}_{21}{}', \theta_{22}{}')'})\right\|_{p(\Psi)}.$$

As an illustration, we suggest to use the following natural loss function. We consider β_2° and $\widetilde{\beta}_2(\theta_1, \theta_2)$ respectively the pseudo-true value and the binding function associated with the parameters β_2 defined by the instrumental model (2.13). As already mentioned in section 3, β_2° and $\widetilde{\beta}_2(\theta_1, \theta_2)$ are consistently estimated by $\widehat{\beta}_{2,T}$ and $\widetilde{\beta}_{2,TS}(\theta_1, \theta_2)$.

We define the value $\overline{\theta}_{22}$ of the nuisance parameters as the solution to the following minimization program:

$$\overline{\theta}_{22} = \underset{\theta_{22}\in\Theta_{22}}{Argmin} \left(\beta_2^\circ - \widetilde{\beta}_2(\theta_1^\circ, \overline{\theta}_{21}, \theta_{22})\right)'\Omega_2\left(\beta_2^\circ - \widetilde{\beta}_2(\theta_1^\circ, \overline{\theta}_{21}, \theta_{22})\right), \text{[14]} \tag{5.4}$$

where Ω_2 is a positive matrix on $\mathbb{R}^{q_2}$ and $q_2 = dim\beta_2$. In order to estimate the parameters $\overline{\theta}_{22}$, we define the estimator $\widehat{\theta}_{22,TS}$ as follows:

$$\widehat{\theta}_{22,TS} = \underset{\theta_{22}\in\Theta_{22}}{Argmin} \left(\widehat{\beta}_{2,T} - \widetilde{\beta}_{2,TS}(\widehat{\theta}_{TS}^1(\theta_{22}^*), \theta_{22})\right)'\Omega_2\left(\widehat{\beta}_{2,T} - \widetilde{\beta}_{2,TS}(\widehat{\theta}_{TS}^1(\theta_{22}^*), \theta_{22})\right), \text{[15]} \tag{5.5}$$

[14] Note that in the case where $\overline{\theta}_{22}$ is not unique, one can always use a more restrictive loss function so that $\overline{\theta}_{22}$ is unique.

[15] Under usual regularity conditions this estimator is consistent to the value $\overline{\theta}_{22}$ of the nuisance parameters θ_{22} and such that $\sqrt{T}\left(\widehat{\theta}_{22,TS} - \overline{\theta}_{22}\right) = O_{P_*}(1)$.

for a given initial value θ_{22}^*.
This loss function illustrate the wish of reproducing the dimensions of interest associated with β_2 under the constraint that the SII estimator $\widehat{\theta}_{TS}^1(\overline{\theta}_{22})$ is consistent to the pseudo-true unknown value $\left(\theta_1^{\circ\prime}, \overline{\theta}_{21}^{\prime}\right)^{\prime}$ of the structural parameters of interest. In other words, since the nuisance parameters have no genuine meaning, one possible way of selecting their value $\overline{\theta}_{22}$, is to perform simulation exercises where $\theta_1 = \widehat{\theta}_{1,TS}^1(\theta_{22}^*)$, $\theta_{21} = \widehat{\theta}_{21,TS}^1(\theta_{22}^*)$. θ_{22} is then calibrated in order to minimize the discrepancy criterion between the empirical moments and the simulated ones associated with β_2 as defined by (5.5).

5.2 Case of harmful nuisance parameters

We state in this subsection the general SII results in the case of harmful nuisance parameters. We first maintain assumptions $(A1) - (A6)$ and assume in addition $H_\circ^1$:[16]

- $(B1)$ endowed with the true value θ_1° partially encompasses $(\mathcal{N}_\beta)$, i.e.: there exists: $\overline{\theta}_{21}, \overline{\theta}_{22} \in \Theta_{21} \times \Theta_{22}$ such that:
$$\beta_1^\circ = \widetilde{\beta}_1\left(\theta_1^\circ, \overline{\theta}_{21}, \overline{\theta}_{22}\right) \tag{5.6}$$
- and for the previous value $\overline{\theta}_{22}$, (3.7) is fulfilled, namely: $\widetilde{\beta}_1\left(\theta_1, \theta_{21}, \overline{\theta}_{22}\right) = \widetilde{\beta}_1\left(\theta_1^*, \theta_{21}^*, \overline{\theta}_{22}\right) \Longrightarrow (\theta_1, \theta_{21}) = (\theta_1^*, \theta_{21}^*)\cdot$

Assumption (A10):
We have at our disposal a first step consistent estimator $\widehat{\theta}_{22,TS}$ of $\overline{\theta}_{22}$ such that:

(a) $P_* \lim\limits_{T\to+\infty} \widehat{\theta}_{22,TS} = \overline{\theta}_{22}$,

(b) $\sqrt{T}\left(\widehat{\theta}_{22,TS} - \overline{\theta}_{22}\right) \xrightarrow[T\to+\infty]{\mathrm{D}} \mathcal{N}\left(0, \overline{\Phi}_{22}\right)\cdot$[17]

Assumption (A11):

$$\underset{\theta_1,\theta_{21}\in\Theta_1\times\Theta_{21}}{Sup} \left\| \widetilde{\beta}_{1,TS}\left(\theta_1, \theta_{21}, \widehat{\theta}_{22,TS}\right) - \widetilde{\beta}_1\left(\theta_1, \theta_{21}, \overline{\theta}_{22}\right)\right\|_{\Omega_1} \xrightarrow[T\to+\infty]{P_*} 0\cdot$$

We recall that the indirect inference estimator based on the sub-vector β_1 of the instrumental parameters is defined by:

$$\begin{aligned}
\widehat{\theta}_{TS}^1(\overline{\theta}_{22}) &= \left(\widehat{\theta}_{1,TS}^{\prime}\left(\overline{\theta}_{22}\right), \widehat{\theta}_{21,TS}^{\prime}\left(\overline{\theta}_{22}\right)\right)^{\prime}, \\
&= \underset{\theta_1,\theta_{21}\in\Theta_1\times\Theta_{21}}{Argmin} \left[\widehat{\beta}_{1,T} - \widetilde{\beta}_{1,TS}\left(\theta_1, \theta_{21}, \widehat{\theta}_{22,TS}\right)\right]^{\prime} \widehat{\Omega}_{1,T} \left[\widehat{\beta}_{1,T} - \widetilde{\beta}_{1,TS}\left(\theta_1, \theta_{21}, \widehat{\theta}_{22,TS}\right)\right],
\end{aligned} \tag{5.7}$$

[16] We still refer here to $H_\circ^1$ as the partial encompassing property although the setting is, as already explained, different.
[17] Note that $(A10b)$ implies $(A10a)$.

where $P_* \lim_{T\to+\infty} \widehat{\Omega}_{1,T} = \Omega_1$ is a positive matrix on $I\!R^{q_1}$. Note that with a slight abuse of notations, we will refer to the indirect estimator $\widehat{\theta}^1_{TS}\left(\overline{\theta}_{22}\right)$ rather than $\widehat{\theta}^1_{TS}\left(\widehat{\theta}_{22,TS}\right)$. We are now able to prove the following generalized consistency property.[18]

Proposition 5.1 : *Under assumptions* $(A1)-(A6),(A10a),(A11)$ *and* $H^1_\circ : (B1)$ *endowed with the true value* θ°_1 *partially encompasses* $(\mathcal{N}_\beta)$ *according to* (3.7) *and* (5.6) *in case of harmful nuisance parameters, the indirect estimator* $\widehat{\theta}^1_{1,TS}\left(\overline{\theta}_{22}\right)$ *is consistent to* θ°_1.

Proof : *See appendix A.*1.

In order to derive the asymptotic distribution of the SII estimator in case of harmful nuisance parameters, we first maintain assumptions $(A1)-(A6)/(A8a)-(A8b)/(A9a)-(A9d)/(A10b)$ and modify $(A9e)$ to $(A9e)'$:

$(A9e)'$

- $$P_* \lim_{T\to+\infty} \frac{\partial \widetilde{\beta}^s_{1,T}}{\partial \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}'}\left(\theta^\circ_1, \overline{\theta}_2\right) = \frac{\partial \widetilde{\beta}_1}{\partial \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}'}(\theta^\circ_1, \overline{\theta}_2),$$
  is of full-column rank $(p_1 + p_2)$.
- $$\lim_{T\to+\infty} Cov\left[\sqrt{T}\frac{\partial Q_T}{\partial \beta}\left(\underline{y}_T, \underline{x}_T, \beta^\circ\right), \sqrt{T}\left(\widehat{\theta}_{22,TS} - \overline{\theta}_{22}\right)\right] = L_\circ(\overline{\theta}_{22}),$$
- $$\lim_{T\to+\infty} Cov\left[\sqrt{T}\frac{\partial Q_T}{\partial \beta}\left(\underline{\widetilde{y}}^s_T(\theta^\circ_1, \overline{\theta}_{22}, z^s_\circ), \underline{x}_T, \widetilde{\beta}^\circ(\overline{\theta}_{22})\right), \sqrt{T}\left(\widehat{\theta}_{22,TS} - \overline{\theta}_{22}\right)\right] = L^*_\circ(\overline{\theta}_{22}).$$

We are now able to prove the following result[19]:

Proposition 5.2 : *Under assumptions* $(A1)-(A6)/(A8a)-(A8b)/(A9)'/(A10b)$ *and the null hypothesis* $H^1_\circ$, *the indirect inference estimator* $\widehat{\theta}^1_{TS}(\overline{\theta}_{22})$ *is asymptotically normal, when S is fixed and T goes to infinity:*

$$\sqrt{T}\begin{pmatrix} \widehat{\theta}^1_{1,TS}(\overline{\theta}_{22}) - \theta^\circ_1 \\ \widehat{\theta}^1_{21,TS}(\overline{\theta}_{22}) - \overline{\theta}_{21} \end{pmatrix} \xrightarrow[T\to+\infty]{D} \mathcal{N}\left(0, W_1(S, \Omega_1, \overline{\theta}_{22}, \overline{\Phi}_{22})\right),$$

where:

$$W_1(S, \Omega_1, \overline{\theta}_{22}, \overline{\Phi}_{22}) = \left\{\frac{\partial \widetilde{\beta}'_1}{\partial \begin{pmatrix} \theta_1 \\ \theta_{21} \end{pmatrix}}(\theta^\circ_1, \overline{\theta}_2)\Omega_1 \frac{\partial \widetilde{\beta}_1}{\partial \begin{pmatrix} \theta_1 \\ \theta_{21} \end{pmatrix}'}(\theta^\circ_1, \overline{\theta}_2)\right\}^{-1} \frac{\partial \widetilde{\beta}'_1}{\partial \begin{pmatrix} \theta_1 \\ \theta_{21} \end{pmatrix}}(\theta^\circ_1, \overline{\theta}_2)\Omega_1 \Phi^*_{\circ,1}(S, \overline{\theta}_{22}, \overline{\Phi}_{22})$$
$$\Omega_1 \frac{\partial \widetilde{\beta}_1}{\partial \begin{pmatrix} \theta_1 \\ \theta_{21} \end{pmatrix}'}(\theta^\circ_1, \overline{\theta}_2)\left\{\frac{\partial \widetilde{\beta}'_1}{\partial \begin{pmatrix} \theta_1 \\ \theta_{21} \end{pmatrix}}(\theta^\circ_1, \overline{\theta}_2)\Omega_1 \frac{\partial \widetilde{\beta}_1}{\partial \begin{pmatrix} \theta_1 \\ \theta_{21} \end{pmatrix}'}(\theta^\circ_1, \overline{\theta}_2)\right\}^{-1},$$

[18]We will focus here on sufficient (partial encompassing) conditions for consistency. However the necessary and sufficient conditions framework developed in section 3 can also be extended in case of harmful nuisance parameters.

[19]We will refer to $(A9)'$ as the set of assumptions $(A9a)-(A9d)/(A9e)'$.

$\Phi^*_{\circ,1}(S,\overline{\theta}_{22},\overline{\Phi}_{22}) = \Phi^*_{\circ,1}(S,\overline{\theta}_{22}) + \Psi^*_{\circ,1}(S,\overline{\theta}_{22}) + \Psi^{*'}_{\circ,1}(S,\overline{\theta}_{22}) + \Gamma^*_{\circ,1}(S,\overline{\theta}_{22}) + \Gamma^{*'}_{\circ,1}(S,\overline{\theta}_{22}) + \dfrac{\partial\widetilde{\beta}_1}{\partial\theta'_{22}}(\theta^\circ_1,\overline{\theta}_2)\overline{\Phi}_{22}\dfrac{\partial\widetilde{\beta}'_1}{\partial\theta_{22}}(\theta^\circ_1,\overline{\theta}_2)$.

- $\Phi^*_{\circ,1}(S,\overline{\theta}_{22})$ *is the* $(q_1\times q_1)$ *left-upper bloc diagonal sub-matrix of the* $(q\times q)$ *matrix* $\Phi^*_\circ(S,\overline{\theta}_{22})$ *defined by:*

$$\begin{aligned}\Phi^*_\circ(S,\overline{\theta}_{22}) &= J_\circ^{-1}I_\circ J_\circ^{-1} + \frac{1}{S}J^{*^{-1}}_\circ(\overline{\theta}_{22})I^*_\circ(\overline{\theta}_{22})J^{*^{-1}}_\circ(\overline{\theta}_{22}) + \left(1-\frac{1}{S}\right)J^{*^{-1}}_\circ(\overline{\theta}_{22})K^*_\circ(\overline{\theta}_{22})J^{*^{-1}}_\circ(\overline{\theta}_{22})\\ &-J_\circ^{-1}K_\circ(\overline{\theta}_{22})J^{*^{-1}}_\circ(\overline{\theta}_{22}) - J^{*^{-1}}_\circ(\overline{\theta}_{22})K'_\circ(\overline{\theta}_{22})J_\circ^{-1}.\end{aligned}$$

- $\Psi^*_{\circ,1}(S,\overline{\theta}_{22})$ *is the* $(q_1\times q_1)$ *upper bloc sub-matrix of the* $(q\times q_1)$ *matrix* $\Psi^*_\circ(S,\overline{\theta}_{22})$ *defined by:*

$$\Psi^*_\circ(S,\overline{\theta}_{22}) = J_\circ^{-1}L_\circ(\overline{\theta}_{22})\frac{\partial\widetilde{\beta}'_1}{\partial\theta_{22}}(\theta^\circ_1,\overline{\theta}_2),$$

- $\Gamma^*_{\circ,1}(S,\overline{\theta}_{22})$ *is the* $(q_1\times q_1)$ *upper bloc sub-matrix of the* $(q\times q_1)$ *matrix* $\Gamma^*_\circ(S,\overline{\theta}_{22})$ *defined by:*

$$\Gamma^*_\circ(S,\overline{\theta}_{22}) = -J^{*^{-1}}_\circ(\overline{\theta}_{22})L^*_\circ(\overline{\theta}_{22})\frac{\partial\widetilde{\beta}'_1}{\partial\theta_{22}}(\theta^\circ_1,\overline{\theta}_2).$$

Proof : *see appendix A.4.*

The asymptotic covariance matrix depends on the metric Ω_1 and as usual, there exists an optimal choice of the weighting matrix $\Omega^*_1(S,\overline{\theta}_{22},ø6\Phi_{22})$ which minimizes $W_1(S,\Omega_1,\overline{\theta}_{22},\overline{\Phi}_{22})$.

Proposition 5.3 : *Under assumptions* $(A1)-(A6)/(A8a)-(A8b)/(A9)'/(A10b)$ *and the null hypothesis* $H^1_\circ$, *the optimal choice* $\Omega^*_1(S,\overline{\theta}_{22},\overline{\Phi}_{22})$ *of* Ω_1 *for the indirect inference estimator* $\widehat{\theta}^1_{TS}\left(\overline{\theta}_{22}\right)$ *is given by* $\Omega^*_1(S,\overline{\theta}_{22},\overline{\Phi}_{22}) = \Phi^*_{\circ,1}(S,\overline{\theta}_{22},\overline{\Phi}_{22})^{-1}$ *(assuming that* $\Phi^*_{\circ,1}(S,\overline{\theta}_{22},\overline{\Phi}_{22})$ *is non singular). The asymptotic covariance matrix is then given by:*

$$W^*_{1,S}(\overline{\theta}_{22},\overline{\Phi}_{22}) = \left\{\frac{\partial\widetilde{\beta}'_1}{\partial\begin{pmatrix}\theta_1\\ \theta_{21}\end{pmatrix}}(\theta^\circ_1,\overline{\theta}_2)\left(\Phi^*_{\circ,1}(S,\overline{\theta}_{22},\overline{\Phi}_{22})\right)^{-1}\frac{\partial\widetilde{\beta}_1}{\partial\begin{pmatrix}\theta_1\\ \theta_{21}\end{pmatrix}'}(\theta^\circ_1,\overline{\theta}_2)\right\}^{-1}. \tag{5.8}$$

Proof : *see appendix A.4.*

We now focus on the modified test statistics $\xi^1_{T,S}(\overline{\theta}_{22},\overline{\Phi}_{22})$ used in the second step of our ascending procedure of tests described in subsection 4.2.

Proposition 5.4 : *Under assumptions* $(A1)-(A6)/(A8a)-(A8b)/(A9)'/(A10b)$ *and the null-hypothesis* $H^1_\circ$ *of partial encompassing of* $(\mathcal{N}_\beta)$ *by* $(B1)$ *according to definition* 3.6, *the statistic* $\xi^1_{T,S}(\overline{\theta}_{22},\overline{\Phi}_{22})$*:*

$$\begin{aligned}\xi^1_{T,S}(\overline{\theta}_{22},\overline{\Phi}_{22}) = \;& T\min_{(\theta_1,\theta_{21})\in\Theta_1\times\Theta_{21}}\left[\widehat{\beta}_{1,T}-\frac{1}{S}\sum_{s=1}^S\widetilde{\beta}^s_{1,T}(\theta_1,\theta_{21},\widehat{\theta}_{22,TS})\right]'\widehat{\Omega}^*_{1,T}(S)\\ &\left[\widehat{\beta}_{1,T}-\frac{1}{S}\sum_{s=1}^S\widetilde{\beta}^s_{1,T}(\theta_1,\theta_{21},\widehat{\theta}_{22,TS})\right],\end{aligned} \tag{5.9}$$

where $\widehat{\Omega}^{*}_{1,T}(S)$ *is a consistent estimator of the optimal metric* $\Omega_1^*(S,\overline{\theta}_{22},\overline{\Phi}_{22}) = \Phi^{*}_{\circ,1}(S,\overline{\theta}_{22},\overline{\Phi}_{22})^{-1}$ *defined in proposition* 5.3, *is asymptotically distributed as a chi-square with* $(q_1 - p_1 - p_{21})$ *degrees of freedom where* $q_1 = dim\beta_1,\ p_1 = dim\theta_1,\ p_{21} = dim\theta_{21}$.

Proof : *see appendix* A.4.

Therefore, the associated specification test of asymptotic level α is defined by the following critical region:

$$\mathcal{W}^1_{\alpha} = \{\xi^1_{T,S}(\overline{\theta}_{22},\overline{\Phi}_{22}) > \chi^2_{1-\alpha}\left(q_1 - p_1 - p_{21}\right)\}.$$

It is worth mentioning that the proofs of propositions $5.2 - 5.4$ are given directly in appendix A.4. However, one can also derived these results by applying the general theory proposed by Dridi (2000), while noticing that the SII estimator corresponds to the SALS estimator deduced from the estimating equations:

$$\widetilde{\beta}_1(\theta_1,\theta_{21},\overline{\theta}_{22}) - \beta_1^{\circ} = 0 \Longrightarrow (\theta_1',\theta_{21}')' = \left(\theta_1^{\circ'},\overline{\theta}_{21}'\right)',$$

and $\overline{\theta}_{22}$ has been replaced by $\widehat{\theta}_{22,TS}$. We have decided here to give rather a direct proof in order to take into account the mis-specification issues on the simulator. Besides this illustrates the usefulness of the SALS approach as a complementary approach to the SII method which seeks to identify the relevant estimating equations to be used through the ascending procedure of tests. In the next section, we illustrate our SII methodology through examples based on the popular Stochastic Volatility models.

6 Stochastic Volatility and Asymmetries

Empirical financial studies have found definite evidence that the stock market returns present strong conditional heteroskedasticity patterns at high frequency data level. And while the Economic Theory provides little guidance on the selection of an appropriate model and estimation strategy for the conditional variance, it has now become essential to answer such issues, especially when one is interested in valuing financial equities through general asset pricing models.

In order to answer the previous challenge, the stochastic volatility model (SV hereafter) has been proposed by Clark (1973), Tauchen and Pitts (1983), Taylor (1986-1994), Ghysels, Harvey and Renault (1995) among many authors. These models appear as an alternative specification to the famous Auto-regressive Conditionally Heteroskedastic (ARCH) model as introduced by Engle (1982) and Bollerslev (1986). The main difference between the two models relies on whether the volatility of the process is observable or not. More precisely, the SV model introduces unobservable latent factors, which account for broad general features of the financial market data (persistent volatility, volatility clustering effect, leverage effect, asymmetries, leptokurtosis...). Even though ARCH models are more tractable in the uni-variate case, the SV model proposes several improvements with respect to the ARCH specification.

First as pointed out by Andersen and Sorensen (1995), "multi-variate ARCH models induce a proliferation of parameters that must be handled in an arguably ad hoc manner", whereas SV models introduce low dimensional unobservable factors. Second SV models as proposed by Meddahi and Renault (1997) are closed under temporal aggregation whereas standard ARCH models are not.

These are the reasons why we focus in this subsection on SV models $\{y_t, t \in \mathbb{Z}\}$ defined as follows:

$$\left\{ \begin{array}{rcl} y_t & = & \sigma_{t-1}\varepsilon_t, \\ \sigma_t^2 & = & \omega + \gamma\sigma_{t-1}^2 + \nu_t, \end{array} \right. \tag{6.1}$$

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 semi-parametric specification (6.1), the innovation processes $\{\varepsilon_t, t \in \mathbb{Z}\}$ and $\{\nu_t, t \in \mathbb{Z}\}$ are assumed to share the following properties:

$$\begin{array}{lll} E\left[\varepsilon_t/\, I_{t-1}\right] = 0, & E\left[\varepsilon_t^2/\, I_{t-1}\right] = 1, & E\left[\varepsilon_t^3/\, I_{t-1}\right] = \mu_3, \\ E\left[\varepsilon_t^4/\, I_{t-1}\right] = \mu_4, & E\left[\nu_t/\, I_{t-1}\right] = 0, & E\left[\nu_t^2/\, I_{t-1}\right] = \eta^2, \text{ [20]} \\ E\left[\varepsilon_t\nu_t/\, I_{t-1}\right] = \rho\eta, & & \end{array} \tag{6.2}$$

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 found strong evidence that the stock market returns have an important asymmetric behavior. Within the framework delineated by the specification $(6.1) - (6.2)$, this stylized fact can be explained by the skewness of the standardized innovation

[20] This semi-parametric SV model is due to Meddahi and Renault (1997). Of course, we could have focused on log-stochastic volatility models specifications à la Andersen (1994), Taylor (1994) and Harvey, Ruiz and Shephard (1994) but as can be seen in the sequel, there are several reasons why we do not use the latter specification.

process $\{\varepsilon_t, t \in \mathbb{Z}\}$ ($\mu_3 \neq 0$) and by the so-called leverage effect ($\rho < 0$), which corresponds according to Black (1976), to the negative correlation between innovations to volatility process $\{\nu_t, t \in \mathbb{Z}\}$ and innovations to return process $\{\varepsilon_t, t \in \mathbb{Z}\}$ (see appendix A.5. for an in depth discussion of this issue).

These two sources of asymmetries may lead the econometrician to build mis-specified structural model, especially when according to the common practice, she wrongly predisposes one specification rather than the other. That is, either neglecting the leverage effect and imposing $\rho = 0$ or neglecting the skewness of the standardized innovation process $\{\varepsilon_t, t \in \mathbb{Z}\}$ and imposing $\mu_3 = 0$, while these restrictions are respectively not fulfilled.

Of course, one can always argue, that it is possible to avoid this kind of mis-specifications by relaxing the fallacious constraint $\rho = 0$ or $\mu_3 = 0$. But in our opinion, this objection is irrelevant since until further developments in the econometric modeling, one never knows, in practice, how to improve a priori the model specification. A good illustration of this point is precisely the aforementioned confusion between the leverage effect and the skewness of the standardized innovation process $\{\varepsilon_t, t \in \mathbb{Z}\}$. In this case, we are able to give a new insight on the sources of the asymmetric behavior of the stock markets returns. Indeed, it is always easy to claim a posteriori that one can avoid the mis-specification by relaxing the fallacious constraint $\rho = 0$ or $\mu_3 = 0$. In this respect, the examples given below should be regarded as illustrations and applications of our SII methodology and which examine the effects of mis-specifications in the asymmetries.

Furthermore within the semi-parametric SV specification (6.1) – (6.2) 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). This may lead to another type of deeper mis-specifications, which are scarcely avoidable. Indeed, within this semi-parametric setting, the theory provides no insight on what the p.d.f. of the joint process $\{(\varepsilon_t, \nu_t), t \in \mathbb{Z}\}$ could be. This is even more upsetting since when one is also interested in the estimation of the asymmetry parameters (ρ, μ_3), no direct estimation, such as for instance the Generalized Method of Moments, is available.

In this framework one has, first, to choose in a rather arbitrarily ad hoc manner a specification for the p.d.f. of the joint process $\{(\varepsilon_t, \nu_t), t \in \mathbb{Z}\}$ and , second, to perform a SII generally with a mis-specified structural model being used as a simulator. This corresponds to our basic outlook, that is, consistent indirect estimation of some parameters of interest despite a mis-specified simulator.

Note that in our examples there are two types of mis-specifications, that is, mis-specifications in the asymmetries and mis-specifications in the p.d.f. of the joint process $\{(\varepsilon_t, \nu_t), t \in \mathbb{Z}\}$.

We denote $\delta = (\omega, \gamma, \eta^2, \mu_3, \rho, \mu_4)'$ the structural unknown parameters associated with $(6.1) - (6.2)$. We define the family $\mathcal{P}$ of probability distributions compatible with the semi-parametric SV model

$(6.1)-(6.2)$. That is, there exists an application $\widetilde{\delta}(\cdot)$ from $\mathcal{P}$ onto a part $\Delta=\widetilde{\delta}(\mathcal{P})$ such that:

$$
\begin{aligned}
&\widetilde{\delta}: \mathcal{P} \longrightarrow \Delta, \\
&\qquad P \longrightarrow \widetilde{\delta}=(\widetilde{\omega},\widetilde{\gamma},\widetilde{\eta}^2,\widetilde{\mu}_3,\widetilde{\rho},\widetilde{\mu}_4)', \\
&\forall P \in \mathcal{P},\ \forall\ \{y_t, t\in \mathbb{Z}\} \in L_4\left(\mathbb{R},\mathcal{B}(\mathbb{R}),\lambda\right)^{\mathbb{Z}} \rightsquigarrow P, \text{ [21]} \\
&\Rightarrow \exists\ \{(\sigma_t,\nu_t), t\in\mathbb{Z}\} \text{ such that:} \\
&y_t = \sigma_{t-1}\varepsilon_t, \\
&\sigma_t^2 = \widetilde{\omega}(P)+\widetilde{\gamma}(P)\sigma_{t-1}^2+\nu_t,
\end{aligned}
\tag{6.3}
$$

with:

$$
\begin{aligned}
&\underset{P}{E}\left[\varepsilon_t/\ I_{t-1}\right]=0, && \underset{P}{E}\left[\varepsilon_t^2/\ I_{t-1}\right]=1, \quad \underset{P}{E}\left[\varepsilon_t^3/\ I_{t-1}\right]=\widetilde{\mu}_3(P), \\
&\underset{P}{E}\left[\varepsilon_t^4/\ I_{t-1}\right]=\widetilde{\mu}_4(P), && \underset{P}{E}\left[\nu_t/\ I_{t-1}\right]=0, \quad \underset{P}{E}\left[\nu_t^2/\ I_{t-1}\right]=\widetilde{\eta}^2(P), \\
&\underset{P}{E}\left[\varepsilon_t\nu_t/\ I_{t-1}\right]=\widetilde{\rho}(P)\widetilde{\eta}(P),
\end{aligned}
\tag{6.4}
$$

where the information set $I_t=\sigma\left(\varepsilon_t,\varepsilon_\tau,\nu_\tau,\tau<t\right)$ is the σ-field generated by $(\varepsilon_t,\varepsilon_\tau,\nu_\tau,\tau<t)$, $0<\widetilde{\omega}(P)$ and $0<\widetilde{\gamma}(P)<1$. That is for each $P\in\mathcal{P}$ and each stochastic process $\{y_t,t\in\mathbb{Z}\}$ such that its p.d.f. is P, there exists $\widetilde{\delta}(P)\in\Delta$ such that the stochastic process $\{y_t,t\in\mathbb{Z}\}$ belongs to the class of SV models as delineated by $(6.1)-(6.2)$ for the value $\delta=\widetilde{\delta}(P)$ of the unknown structural parameters.

We focus now on the effects of neglecting the skewness parameter μ_3. The data consist in the observations of a stochastic process $\{y_t, t\in\mathbb{Z}\}$, $t=1,...,T$. We denote $P_\circ$ the true unknown p.d.f. of $\{y_t,t\in\mathbb{Z}\}$.

Assumption (A11): [22]

(i) $P_\circ$ belongs to the family $\mathcal{P}$ of probability distributions.

(ii) We define $\delta^\circ=\widetilde{\delta}\left(P_\circ\right)$ and we assume that $\delta^\circ=\left(\omega^\circ,\gamma^\circ,\eta^{\circ 2},\mu_3^\circ,0,\mu_4^\circ\right)'\in\overset{\circ}{\Delta}$ and $\underset{\circ}{E}\left[\varepsilon_t^2\nu_t/\ I_{t-1}\right]=0$.[23]

In other words according to $(A11)$, the data are generated by a conditionally skewed and leptokurtic SV model.

As previously seen, there are two sources of asymmetric responses of the stock market returns: the skewness of the innovations to returns ($\mu_3\neq 0$) and the leverage effect ($\rho<0$). In this context, a common but nonetheless wrong practice consists in predisposing the leverage effect ρ while neglecting the skewness of the innovations to returns ($\mu_3\neq 0$). The econometrician focuses in this case on SV models defined as follows.

[21] Note that $\Delta=\widetilde{\delta}(\mathcal{P})=\mathbb{R}_+^*\times]0,1[\times\mathbb{R}_+^*\times\mathbb{R}\times\mathbb{R}\times\mathbb{R}_+^*$ and that $L_4\left(\mathbb{R},\mathcal{B}(\mathbb{R}),\lambda\right)$ is the set of measurable variables with finite fourth order moments. λ corresponds to the Lebesgues measure and $\{y_t,t\in\mathbb{Z}\}\rightsquigarrow P$ means that the p.d.f. of $\{y_t,t\in\mathbb{Z}\}$ is P.

[22] Note that assumption $(A11)$ does correspond to the condition delineated by assumption $(A1)$.

[23] Note that this symmetry assumption could be implied by the following independence assumption $\{\varepsilon_t,t\in\mathbb{Z}\}\perp\!\!\!\perp\{\nu_t,t\in\mathbb{Z}\}$. In that case, we have trivially $\rho^\circ=0$ since $\underset{\circ}{E}\left[\varepsilon_t\nu_t/\ I_{t-1}\right]=\rho^\circ\eta^\circ=\underset{\circ}{E}\left[\varepsilon_t/\ I_{t-1}\right]\underset{\circ}{E}\left[\nu_t/\ I_{t-1}\right]=0$.

Nominal assumption (B2):

(i)

$$\begin{cases} y_t &= \sigma_{t-1}\varepsilon_t, \\ \sigma_t^2 &= \omega + \gamma\sigma_{t-1}^2 + \nu_t, \end{cases} \tag{6.5}$$

where $0 < \gamma < 1$ and $0 < \omega$ and the following assumptions on the innovation processes $\{\varepsilon_t, \nu_t, t \in \mathbb{Z}\}$ are made:

$$\begin{array}{lll} E\left[\varepsilon_t / I_{t-1}\right] = 0, & E\left[\varepsilon_t^2 / I_{t-1}\right] = 1, & E\left[\varepsilon_t^3 / I_{t-1}\right] = 0, \\ E\left[\varepsilon_t^4 / I_{t-1}\right] = \mu_4, & E\left[\nu_t / I_{t-1}\right] = 0, & E\left[\nu_t^2 / I_{t-1}\right] = \eta^2,\ ^{24} \\ E\left[\varepsilon_t \nu_t / I_{t-1}\right] = \rho\eta, & E\left[\varepsilon_t^2 \nu_t / I_{t-1}\right] = 0, & \end{array} \tag{6.6}$$

where the information set $I_t = \sigma\left(\varepsilon_t, \varepsilon_\tau, \nu_\tau, \tau < t\right)$ is the σ-field generated by $\left(\varepsilon_t, \varepsilon_\tau, \nu_\tau, \tau < t\right)$. The parameters of interest θ_1 is defined by $\theta_1 = (\omega, \gamma, \eta^2, \rho, \mu_4)'$. We define the function $\widetilde{\theta}_1(\cdot)$ from the set $\mathcal{P}$ onto $\Theta_1 = \widetilde{\theta}_1\left(\mathcal{P}\right)$ by:

$$\begin{array}{rcl} \widetilde{\theta}_1 & : & \mathcal{P} \longrightarrow \Theta_1, \\ & & P \longrightarrow \widetilde{\theta}_1\left(P\right) = \left(\widetilde{\omega}\left(P\right), \widetilde{\gamma}\left(P\right), \widetilde{\eta}^2\left(P\right), \widetilde{\rho}\left(P\right), \widetilde{\mu}_4\left(P\right)\right)'. \end{array} {}^{25}$$

With a slight abuse of notations we can write $\widetilde{\delta}\left(P\right) = \left(\widetilde{\theta}_1\left(P\right)', \widetilde{\mu}_3\left(P\right)\right)'$.

(ii) As already pointed out, in order to identify and estimate the parameters of interest $\theta_1^\circ = \left(\omega^\circ, \gamma^\circ, \eta^{\circ 2}, \rho^\circ, \mu_4^\circ\right)'$ $(\rho^\circ = 0)$, the econometrician makes further assumptions on the law of motion of the joint process $\{(\varepsilon_t, \nu_t), t \in \mathbb{Z}\}$. This may require additional nuisance parameters $\theta_2 \in \Theta_2 \subset \mathbb{R}^{p_2}$. [26] We denote $\theta = (\theta_1', \theta_2')'$ the structural unknown parameters, $\theta \in \Theta = \Theta_1 \times \Theta_2$. The structural model delineated by assumption $(B2)$ is mis-specified, first, with respect to the skewness of the standardized innovations $(\mu_3^\circ \neq 0)$ and second, in general, with respect to the joint p.d.f. of $\{(\varepsilon_t, \nu_t), t \in \mathbb{Z}\}$ (θ_2 say).

In order to perform a SII of the true unknown value θ_1° of the structural parameters of interest θ_1, one has to specify a convenient instrumental model according to proposition (5.1). A possible choice of the instrumental model can be built on an ARCH(q_1) specification since it provides a natural framework to capture the aforementioned features of the data (volatility clusters, asym-

[24] The symmetry assumption $E\left[\varepsilon_t^2 \nu_t / I_{t-1}\right] = 0$ 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 $E\left[\xi_t / I_{t-1}\right] = 0$ and $\{\xi_t, t \in \mathbb{Z}\} \perp\!\!\!\perp \{\varepsilon_t, t \in \mathbb{Z}\}$.

[25] The definition of this function follows immediately from the specification (6.1) – (6.2) and the definition of $\mathcal{P}$. Moreover $\widetilde{\theta}_1\left(P\right)$ can be characterized through the condition (2.11), that is:

$$\underset{P}{E}\, g\left(y_t, y_{t-1}, \ldots, y_{t-K}, \theta_1\right) = 0 \quad \Longrightarrow \quad \theta_1 = \widetilde{\theta}_1\left(P\right),$$

where g is obviously defined by the specification (6.1) – (6.2).

[26] Note that this does not prevent the p.d.f. as well as the support of the joint process $\{(\varepsilon_t, \nu_t), t \in \mathbb{Z}\}$ from depending on the structural parameters of interest θ_1.

metries, leptokurtosis):

$$\begin{cases} y_t &= \sqrt{h_t} z_t, \\ h_t &= \beta_\circ + \sum_{i=1}^{q_1} \beta_i y_{t-i}^2, \end{cases} \tag{6.7}$$

with $q_1 \geq 1$, $E[z_t/\ J_{t-1}] = 0$, $Var[z_t/\ J_{t-1}] = 1$ and where the information set $J_t = \sigma(y_\tau, \tau < t)$ is the σ-field generated by $(y_\tau, \tau < t)$.

We define the instrumental model $(\mathcal{N}_{\beta_1})$ through the following moment conditions (6.8) associated with the ARCH(q_1) specification (6.7):

$$\begin{aligned} E\left[y_t^2 - \beta_{1,\circ} - \sum_{i=1}^{q_1} \beta_{1,i} y_{t-i}^2\right] &= 0, \\ E\left[\left(y_t^2 - \beta_{1,\circ} - \sum_{i=1}^{q_1} \beta_{1,i} y_{t-i}^2\right) y_{t-j}^2\right] &= 0, \qquad j = 1, \ldots, q_1, \\ E\left[y_t^2\right] &= \beta_{1,q_1+1}, \\ E\left[y_t^2 y_{t-1}^2\right] &= \beta_{1,q_1+2}, \\ E\left[y_t^2 y_{t-2}^2\right] &= \beta_{1,q_1+3}, \\ E\left[y_t^2 y_{t-1}\right] &= \beta_{1,q_1+4}, \\ E\left[y_t^4\right] &= \beta_{1,q_1+5}. \end{aligned} \tag{6.8}$$

This choice of moments is performed with respect to structural ideas, in particular, leverage effect and kurtosis $(\beta_{1,q_1+4}, \beta_{1,q_1+5})$. [27] The number of moment conditions $(q_1 + 6)$, is exactly the number of instrumental parameters $\beta_1 = (\beta_{1,\circ}, \ldots, \beta_{1,q_1+5})' \in \mathcal{B}_1 \subset I\!R^{q_1+6}$. So that we are in a just-identified framework concerning the instrumental model $(\mathcal{N}_{\beta_1})$. Moreover, we are able to prove that the moment conditions (6.8) uniquely define the instrumental parameters β_1 (see appendix A.5.).

We associate with the instrumental model $(\mathcal{N}_{\beta_1})$ the following instrumental criterion:

$$Q_T^1\left(\underline{y}_T, \beta_1\right) = \left\{\frac{1}{T}\sum_{t=q_1+1}^{T} g_1\left(y_t, \ldots, y_{t-q_1}, \beta_1\right)\right\}' \left\{\frac{1}{T}\sum_{t=q_1+1}^{T} g_1\left(y_t, \ldots, y_{t-q_1}, \beta_1\right)\right\}. \tag{6.9}$$ [28]

[27] The moment conditions associated respectively with $(\beta_{1,q_1+1}, \beta_{1,q_1+2}, \beta_{1,q_1+3})$ have been added here just in order to ensure the indirect identification of the structural parameters of interest θ_1 (see appendix A.5. and lemma 6.1 in the sequel).

[28] In this just-identified setting, there is no need to introduce a weighting matrix for the instrumental GMM criterion (see appendix A.5. for more details).

$g_1(y_t,\ldots,y_{t-q_1},\beta_1)$ is the function from $\mathbb{R}^{2q_1+7}$ onto $\mathbb{R}^{q_1+6}$ defined as follows:

$$g_1(y_t,\ldots,y_{t-q_1},\beta_1)=\begin{bmatrix} y_t^2-\beta_{1,\circ}-\sum_{i=1}^{q_1}\beta_{1,i}y_{t-i}^2 \\ \left(y_t^2-\beta_{1,\circ}-\sum_{i=1}^{q_1}\beta_{1,i}y_{t-i}^2\right)y_{t-1}^2 \\ \cdot \\ \cdot \\ \cdot \\ \left(y_t^2-\beta_{1,\circ}-\sum_{i=1}^{q_1}\beta_{1,i}y_{t-i}^2\right)y_{t-q_1}^2 \\ y_t^2-\beta_{1,q_1+1} \\ y_t^2y_{t-1}^2-\beta_{1,q_1+2} \\ y_t^2y_{t-2}^2-\beta_{1,q_1+3} \\ y_t^2y_{t-1}-\beta_{1,q_1+4} \\ y_t^4-\beta_{1,q_1+5} \end{bmatrix}. \tag{6.10}$$

We assume the following law of large numbers and uniform law of large numbers:

Assumption (A12):

$$(A12)\quad \begin{aligned} &P_\circ\lim_{T\to+\infty}\left[Q_T^1\left(\underline{y}_T,\beta_1\right)-\left\|\underset{\circ}{E}\,g_1(y_t,\ldots,y_{t-q_1},\beta_1)\right\|_2^2\right] &&=0,\\ \forall\theta\in\Theta,\quad &P_*\lim_{T\to+\infty}\,\underset{\beta_1\in\mathcal{B}_1}{Sup}\left|Q_T^1\left(\underline{\widetilde{y}}_T^s(\theta,z_\circ^s),\beta_1\right)-\left\|\underset{\theta}{E}\,g_1\left(\widetilde{y}_t^s(\theta,z_\circ^s),\ldots,\widetilde{y}_{t-q_1}^s(\theta,z_\circ^s),\beta_1\right)\right\|_2^2\right| &&=0,\end{aligned}$$

where $\{\widetilde{y}_\circ^s(\theta,z_\circ^s),\ldots,\widetilde{y}_T^s(\theta,z_\circ^s)\}$ correspond to simulated paths conditionally on $z_\circ^s$ for $s=1,\ldots,S$ and for any values θ of the structural parameters.

In this context, we first prove the two following lemmas useful for establishing the desired consistency property.

Lemma 6.1 :

• *The functions:*

$$\begin{array}{lll} \mathcal{B}_1 & \to & \mathbb{R}^+ \\ \beta_1 & \to & \left\|\underset{\circ}{E}\,g_1(y_t,\ldots,y_{t-q_1},\beta_1)\right\|_2^2, \\ \beta_1 & \to & \left\|\underset{\theta}{E}\,g_1\left(\widetilde{y}_t^s(\theta,z_\circ^s),\ldots,\widetilde{y}_{t-q_1}^s(\theta,z_\circ^s),\beta_1\right)\right\|_2^2, \end{array}$$

are non stochastic twice differentiable functions not depending on the initial conditions $z_\circ^s$ and with a unique minimum with respect to β_1: β_1° and $\widetilde{\beta}_1(\theta)$ defined by:

$$\begin{aligned} \left\|\underset{\circ}{E}\,g_1(y_t,\ldots,y_{t-q_1},\beta_1^\circ)\right\|_2^2 &= 0,\\ \left\|\underset{\theta}{E}\,g_1\left(\widetilde{y}_t^s(\theta,z_\circ^s),\ldots,\widetilde{y}_{t-q_1}^s(\theta,z_\circ^s),\widetilde{\beta}_1(\theta)\right)\right\|_2^2 &= 0\cdot \end{aligned}$$

• *Moreover the function* $\widetilde{\beta}_1(\cdot)$ *is partially locally identified with respect to* θ_1 *at the point* θ_1°*, that is:*

$$\forall \theta_1 \in \Theta_1,\ \forall \overline{\theta}_2 \in \Theta_2, \quad \widetilde{\beta}_1(\theta_1, \overline{\theta}_2) = \widetilde{\beta}_1(\theta_1^\circ, \overline{\theta}_2) \Longrightarrow \theta_1 = \theta_1^\circ \cdot$$

Proof : *see appendix A.5.*

In order to indirectly identify the additional nuisance parameters θ_2, the econometrician is generally led to introduce the following additional instrumental criterion:

$$Q_T^2\left(\underline{y}_T, \beta_2\right) \text{ where } \beta_2 \in \mathcal{B}_2 \text{ a compact subset of } I\!R^{q_2},^{29} \tag{6.11}$$

and such that:

Assumption (A13):

$$(A13) \qquad \begin{aligned} &P_\circ \lim_{T\to+\infty} \left[Q_T^2\left(\underline{y}_T, \beta_2\right) - q_\circ^2(\beta_2)\right] &&= 0, \\ &P_* \lim_{T\to+\infty} \underset{\theta\in\Theta}{Sup} \left|Q_T^2\left(\underline{\widetilde{y}}_T^s(\theta, z_\circ^s), \beta_2\right) - q_M^2(\theta, \beta_2)\right| &&= 0, \end{aligned}$$

$q_\circ^2(\beta_2)$ and $q_M^2(\theta, \beta_2)$ are assumed to be non stochastic twice differentiable functions not depending on the initial conditions $z_\circ^s$ and with a unique minimum with respect to β_2. Let β_2° and $\widetilde{\beta}_2(\theta_1, \theta_2)$ be respectively the minimum of $q_\circ^2(\beta_2)$ and $q_M^2(\theta, \beta_2)$, that is:

Assumption (A14):

$$(A14) \qquad \begin{aligned} \beta_2^\circ = \beta(P_\circ) &= \underset{\beta_2\in\mathcal{B}_2}{Argmin}\, q_\circ^2(\beta_2), \\ \widetilde{\beta}_2(\theta_1, \theta_2) &= \underset{\beta_2\in\mathcal{B}_2}{Argmin}\, q_M(\theta_1, \theta_2, \beta_2)\cdot \end{aligned}$$

We denote $q = q_1 + q_2 + 6$, and we define the instrumental criterion $Q_T\left(\underline{y}_T, \beta\right)$ by:

$$Q_T\left(\underline{y}_T, \beta\right) = Q_T^1\left(\underline{y}_T, \beta_1\right) + Q_T^2\left(\underline{y}_T, \beta_2\right), \tag{6.12}$$

where $\beta = (\beta_1', \beta_2')' \in \mathcal{B} = \mathcal{B}_1 \times \mathcal{B}_2 \subset I\!R^q$.

We make the following identification assumption:

Assumption (A15):

the function $\widetilde{\beta}(\cdot)$ is one to one. We already know that under assumptions $(A12) - (A14)$:

$$P_\circ \lim_{T\to+\infty} \widehat{\beta}_T = \beta^\circ,$$

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

[29] A possible choice for the instrumental criterion (6.11) can be set through a GMM type instrumental criterion defining β_2.

We assume in addition that the latter convergence is uniform in θ, that is for $s = 1, \dots, S$:
Assumption (A16):

$$(A16) \qquad P_* \lim_{T\to+\infty} \underset{(\theta_1',\theta_2')'\in\Theta}{Sup} \left\| \widetilde{\beta}_T^s(\theta_1,\theta_2) - \widetilde{\beta}(\theta_1,\theta_2) \right\|_q = 0 \cdot$$

Lemma 6.2 : *Under assumptions $(A12)-(A16)$, the conditions delineated by assumptions $(A2)-(A5)$ are fulfilled.*

Proof : *this is immediately deduced from the definitions of assumptions $(A12) - (A16)$ and $(A2) - (A5)$.*

We are now able to prove one of the main results of this section.

Proposition 6.1 : *Under assumptions $(A11) - (A16)$ and for each $\overline{\theta}_2 \in \Theta_2$, the structural model $(B2)$ endowed with the true value θ_1° partially encompasses (N_β) (with respect to β_1) that is:*

$$\beta_1^\circ = \widetilde{\beta}_1(\theta_1^\circ, \overline{\theta}_2) \cdot$$

Proof : *see appendix $A.5$.*

Proposition 6.2 : *Under assumptions $(A11) - (A16)$, the SII estimator $\widehat{\theta}_{1,TS}^1(\overline{\theta}_{22})$ as defined by* (5.7) *is consistent to the true unknown value $\theta_1^\circ = \left(\omega^\circ, \gamma^\circ, \eta^{\circ 2}, 0, \mu_4^\circ\right)'$.*

Proof : *By conjunction of proposition* (3.3) *with proposition* (6.1)*, the result of proposition* (6.2) *is straightforward.*

Actually it is worth noticing that the addition of the directions β_2 has be done solely in order to ensure the identification and thus the two-steps estimation of the nuisance parameters θ_{22} while still maintaining a consistent estimation of θ_1°. If one is only interested in the consistent estimation of θ_1°, one can fix any value to the nuisance parameters θ_{22}, this enables to get rid of the additional β_2 and associated instrumental criterion. However and as already mentioned one may desire to minimize some given loss function inv/lving β_2 while choosing θ_{22}.
Propositions (6.1) and (6.2) are, in our opinion good examples illustrating our SII methodology: the econometrician has basically to focus on the dimensions of almost correct specification as delineated by our partial encompassing definition (5.6). Moreover, the joint estimation of the structural parameters of interest θ_1 with that of the nuisance parameters θ_2 generally leads to an inconsistent SII estimator $\widehat{\theta}_{1,TS}$ of the true unknown value θ_1°, since one has to introduce additional moments or more generally additional instrumental parameters, as delineated by assumptions $(A12) - (A14)$, which no longer satisfy the desired encompassing property for consistency. Indeed under assumptions $(A11) - (A16)$ and $(B2)$, and when one focuses on indirect estimation simultaneously about

θ_1° and $\overline{\theta}_2$, one has to use the additional instrumental characteristics associated with β_2. Therefore the objective minimum (2.15) applied to the SV model is in general not reached at zero, so that the indirect estimator is in general inconsistent for the true unknown value θ_1°. More precisely, there always exists a weighting matrix Ω such that the consistency property is violated. It is also possible to build an example where the indirect estimator based on the whole instrumental criterion (β say) is always inconsistent. This is achieved as soon as one uses as instrumental criterion $\mathcal{N}_{\beta_1}$ and adds one of the following instrumental moment conditions:

$$\beta_{q_1+6}^{j} = E\left[\left(y_t^2 - \beta_{1,\circ} - \sum_{i=1}^{q_1} \beta_{1,i} y_{t-i}^2\right) y_{t-j}\right], \quad j = 1, \ldots, q_1.$$

And while Andersen and Sorensen (1995) advocates: "As the sample expands one should exploit additional moment restrictions. However, in small samples... the use of additional information can be harmful", we stress here that the use of additional information, as for instance by means of SNP score generator à la Bansal, Gallant, Hussey and Tauchen (1995), is always harmful when one acknowledges the potential mis-specification in the structural model and seeks to consistently estimate some components of the structural parameters.

Furthermore, we could have proposed an example based on log-stochastic volatility model specification but there are, at least, two reasons why we do not.

First, in this "control experiment" framework and apart from the lognormal case, no closed form expression for the moments of interest can be derived. So that one has to rely on Monte Carlo experiments to assess the consistency property. This does not mean that in our case, the simulations are not performed. They are indeed but the ease in the computations offered by the SV model à la Meddahi and Renault enables us to prove our consistency results without using such Monte Carlo experiments.

Second, even though in the lognormal case, closed form expression can be derived (see Jacquier, Polson and Rossi (1994)), we would have lost the semi-parametric property of our results. That is, in the case of lognormal SV model, the only sources of mis-specifications would have to come from the asymmetries.

Of course an analogous example can be build where the previous specification of the structural model ($B2$) corresponds to the DGP and vice versa. The same kind of consistency results are established and deduced from partial encompassing property.

Finally, we have focused on cases where the actual leverage effect is null ($\rho^\circ = 0$) and there are asymmetries in the innovation process ($\mu_3^\circ \neq 0$), however the previous partial encompassing and therefore consistency results extend whenever $\rho^\circ \neq 0$. In the latter case, the only requirement is that there exits $\overline{\theta}_2$ such that $\underset{\theta_1^\circ,\overline{\theta}_2}{E}(\sigma_t) = \underset{\circ}{E}(\sigma_t)$.

7 Concluding Remarks

In this paper, we have proposed an extension to the Indirect Inference methodology to semi-parametric settings and shown how the Semi-parametric Indirect Inference works on basic examples using SV models. Besides the introduction of a new notion of Partial Encompassing that focuses on Pseudo True Values of Interest, robustifies the usual Indirect Inference and enables WET as well as Hausman tests procedures, the main messages of this paper are two-fold:

• First, in order to build consistent SII estimators of the parameters of interest, one has to focus on a parsimonious instrumental model which basically does not capture in some sense the mis-specified part of the simulated paths.

• Second, as long as one acknowledges the likely mis-specifications in the structural model but wishes to consistently estimate some parameters of interest, one should avoid the use of SNP score generator à la Gallant and Tauchen (1996), which in this case would vainly lead to reject the structural model, as well as inconsistent estimators. Finally, let us stress that the building of this Semi-parametric Indirect Inference theory has been suggested by the now increasing literature referring to the so-called **Calibration** methodology. Despite its lack of statistical formalization, we think that this methodology shares the concepts of our SII and should precisely be formalized through our SII methodology (see Broze, Dridi and Renault (1999)).

References

Andersen, T. G. (1994), "Stochastic Auto-regressive Volatility: a Framework for Volatility Modeling", *Mathematical Finance*, 4, 75-102.

Andersen, T. G. and B. E. Sorensen (1996), "GMM Estimation of a Stochastic Volatility Model: a Monte Carlo Study", *Journal of Business Economics and Statistics*, 14, 328-352.

Andrews, D. W. K. (1991), "Heteroskedasticity and Auto-correlation Consistent Covariance Matrix Estimation", *Econometrica*, 59, 817-858.

Andrews, D. W. K. (1999), "Consistent Moment Selection Procedure for Generalized Method of Moments Estimation", *Econometrica*, 67, 543-564.

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), "Non-parametric Estimation of Structural Models for High-Frequency Currency Market Data", *Journal of Econometrics*, 66, 251-287.

Bergstrom, A. R. (1985), "Non-parametric Functions in a Hilbert Space", *Econometric Theory*, 1, 7-26.

Bierens, H. J. and N. R. Swanson (2000), "The Econometrics Consequences of the Ceteris Paribus Condition in Economic Theory", *Journal of Econometrics*, 95 (2), 223-253.

Black, F. (1976), "Studies in Stock Price Volatility Changes", Proceedings of the 1976 Business Meeting of the Business and Economic Statistics Section, American Statistical Association, 177-181.

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

Broze, L., R. Dridi and E. Renault (1999), "Calibration of Structural Models by Semi-parametric Indirect Inference", Mimeo Université de Toulouse.

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

Dalhaus, R. and W. Wefelmeyer (1996), "Asymptotically Optimal Estimation in Misspecified Time Series Models", *The Annals of Statistics*, 24, 952-974.

Dhaene, G., C. Gouriéroux and O. Scaillet (1998), "Instrumental Models and Indirect Encompassing", *Econometrica*, 66, 673-688.

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

Dridi, R. (2000), "Simulated Asymptotic Least Squares Theory", Mimeo LSE.

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

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

Florens, J-P., D. Hendry and J-F. Richard (1996), "Encompassing and Specificity", *Econometric Theory*, 12, 620-656.

Gallant, R. A., R. Hussey and G. Tauchen (1997), "Estimating SV Models with Diagnostic", *Journal of Econometrics*, 81, 159-192.

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

Ghysels, E., A. Harvey and E. Renault (1995), "Stochastic Volatility", *G.S. Maddala and C.R. Rao eds, North Holland, Handbook of statistics, Statistical Methods in Finance*, 14, 119-191.

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

Gouriéroux, C. and A. Monfort (1996), "Simulation-Based Econometric Methods", CORE Lectures, Oxford: Oxford University Press.

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

Gouriéroux, C., A. Monfort and C. Tenreiro (1994), "Kernel M-Estimator and Functional Residual Plots", Mimeo CREST.

Gouriéroux, C., A. Monfort and A. Trognon (1984), "Pseudo Maximum Likelihood Methods: theory", *Econometrica*, 52, 681-700.

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

Harvey, A. C., E. Ruiz and N. Shephard (1994), "Multi-variate Stochastic Variance Models", *Review of Economic Studies*, 61, 247-264.

Hendry, D. F. and J-F. Richard (1982), "On the Formulation of Empirical Models in Dynamic Econometrics", *Journal of Econometrics*, 20, 3-33.

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

Jacquier, E., N. G. Polson and P. E. Rossi (1994), "Bayesian Analysis of Stochastic Volatility Models", *Journal of Business Economics and Statistics*, 12, 371-417.

Maasoumi, E. (1990), "How to Live with Mis-specification if You Must?", *Journal of Econometrics*, 44, 67-86.

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.

Mizon, G. E. and J-F. Richard (1986), "The Encompassing Principle and its Applications to Testing Non-Nested Hypotheses", *Econometrica*, 54, 657-678.

Monfort, A. (1996), "A Reappraisal of Mis-specified Econometric Models", *Econometric Theory*, 12, 597–619.

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 Auto-correlation Consistent Covariance Matrix", *Econometrica*, 55, 703-708.

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

Phillips, P. C. B. (1989), "Partially Identified Models", *Econometric Theory*, 5, 181-240.

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

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. (1982), "Maximum Likelihood Estimation of Mis-specified Models", *Econometrica*, 50, 1-25.

Appendices

A.1. Proof of proposition 3.3 when $\overline{\theta}_{22}$ is replaced by a $\sqrt{\mathbf{T}}$-consistent estimator $\widehat{\theta}_{\mathbf{22,TS}}$:

$$\widehat{\theta}^{1}_{TS}\left(\widehat{\theta}_{22,TS}\right) = \underset{\theta_1,\theta_{21}\in\Theta_1\times\Theta_{21}}{Argmin} \left\|\widehat{\beta}_{1,T} - \widetilde{\beta}_{1,TS}\left(\theta_1,\theta_{21},\widehat{\theta}_{22,TS}\right)\right\|_{\widehat{\Omega}_{1,T}}.$$

Under assumption $(A5)$ we have:

$$\underset{T\to+\infty}{P_*\lim}\,\underset{\theta\in\Theta}{Sup}\left\|\widetilde{\beta}_{1,TS}(\theta) - \widetilde{\beta}_1(\theta)\right\|_{q_1} = 0,$$

$$\iff \forall\varepsilon>0,\ \eta>0,\ \exists T_{\varepsilon,\eta}\ /\ \forall T\geq T_{\varepsilon,\eta},\ \forall\theta\in\Theta : P_*\left[\left\|\widetilde{\beta}_{1,TS}(\theta) - \widetilde{\beta}_1(\theta)\right\|_{q_1} > \eta\right] < \varepsilon,$$

since with probability approaching one $\widehat{\theta}_{22,TS}$ belongs to Θ_{22} (for consistent to $\overline{\theta}_{22}\in\Theta_{22}$),

$$\implies \forall\varepsilon>0,\ \eta>0,\ \exists T_{\varepsilon,\eta}/\ \forall T\geq T_{\varepsilon,\eta},\ \forall\theta_1,\theta_{21}\in\Theta_1\times\Theta_{21} :$$
$$P_*\left[\left\|\widetilde{\beta}_{1,TS}(\theta_1,\theta_{21},\widehat{\theta}_{22,TS}) - \widetilde{\beta}_1(\theta_1,\theta_{21},\widehat{\theta}_{22,TS})\right\|_{q_1} > \eta\right] < \varepsilon,$$

$$\implies \underset{\theta_1,\theta_{21}\in\Theta_1\times\Theta_{21}}{Sup}\left\|\widetilde{\beta}_{1,TS}(\theta_1,\theta_{21},\widehat{\theta}_{22,TS}) - \widetilde{\beta}_1(\theta_1,\theta_{21},\widehat{\theta}_{22,TS})\right\|_{q_1} \xrightarrow[T\to+\infty]{P_*} 0\cdot$$

Case of innocuous nuisance parameters

Under the assumption of partial encompassing and $(A7)$, $\widetilde{\beta}_1\left(\theta_1,\theta_{21},\widehat{\theta}_{22,TS}\right) = \widetilde{\beta}_1\left(\theta_1,\theta_{21}\right)$. Therefore we obtain: $\underset{\theta_1,\theta_{21}\in\Theta_1\times\Theta_{21}}{Sup}\left\|\widetilde{\beta}_{1,TS}(\theta_1,\theta_{21},\widehat{\theta}_{22,TS}) - \widetilde{\beta}_1(\theta_1,\theta_{21})\right\|_{q_1} \xrightarrow[T\to+\infty]{P_*} 0$. Thus, by virtue of the uniform convergence in probability of the associated criterion, we have $\underset{T\to+\infty}{P_*\lim}\,\widehat{\theta}^{1}_{TS}\left(\widehat{\theta}_{22,TS}\right) = \theta^{1^*} = \underset{\theta_1,\theta_{21}\in\Theta_1\times\Theta_{21}}{Argmin}\left\|\beta^{\circ}_1 - \widetilde{\beta}_1\left(\theta_1,\theta_{21}\right)\right\|_{\Omega_1}$.

We already know that under the partial encompassing property $\theta^{1^*} = \left(\theta^{\circ\prime}_1,\overline{\theta}_{21}{}'\right)'$ (see proof of proposition 3.3 in the text), therefore $\underset{T\to+\infty}{P_*\lim}\,\widehat{\theta}_{1,TS}\left(\widehat{\theta}_{22,TS}\right) = \theta^{\circ}_1$.

Case of harmful nuisance parameters

Under assumptions $(A5)$ and $(A11)$, we have:

$$\underset{T\to+\infty}{P_*\lim}\,\underset{\theta\in\Theta}{Sup}\left\|\widetilde{\beta}_{1,TS}\left(\theta\right) - \widetilde{\beta}_1\left(\theta\right)\right\|_{q_1} = 0,$$

$$\underset{T\to+\infty}{P_*\lim}\,\underset{\theta_1,\theta_{21}\in\Theta_1\times\Theta_{21}}{Sup}\left\|\widetilde{\beta}_1\left(\theta_1,\theta_{21},\widehat{\theta}_{22,TS}\right) - \widetilde{\beta}_1\left(\theta_1,\theta_{21},\overline{\theta}_{22}\right)\right\|_{q_1} = 0,$$

$$\iff \forall\varepsilon>0,\eta>0,\exists T_{\varepsilon,\eta}/\forall T\geq T_{\varepsilon,\eta},\forall\theta\in\Theta,$$

$$P_*\left[\left\|\widetilde{\beta}_{1,TS}\left(\theta\right) - \widetilde{\beta}_1\left(\theta\right)\right\|_{q_1} > \frac{\eta}{2}\right] < \frac{\varepsilon}{2},$$

$$P_*\left[\left\|\widetilde{\beta}_1\left(\theta_1,\theta_{21},\widehat{\theta}_{22,TS}\right) - \widetilde{\beta}_1\left(\theta_1,\theta_{21},\overline{\theta}_{22}\right)\right\|_{q_1} > \frac{\eta}{2}\right] < \frac{\varepsilon}{2},$$

$$\Longrightarrow \quad \forall\varepsilon > 0, \eta > 0, \exists T_{\varepsilon,\eta} / \forall T \geq T_{\varepsilon,\eta}, \forall \theta_1, \theta_{21} \in \Theta_1 \times \Theta_{21},$$

$$P_* \left[\left\| \widetilde{\beta}_{1,TS} \left(\theta_1, \theta_{21}, \widehat{\theta}_{22,TS} \right) - \widetilde{\beta}_1 \left(\theta_1, \theta_{21}, \widehat{\theta}_{22,TS} \right) \right\|_{q_1} > \frac{\eta}{2} \right] < \frac{\varepsilon}{2},$$

$$P_* \left[\left\| \widetilde{\beta}_1 \left(\theta_1, \theta_{21}, \widehat{\theta}_{22,TS} \right) - \widetilde{\beta}_1 \left(\theta_1, \theta_{21}, \overline{\theta}_{22} \right) \right\|_{q_1} > \frac{\eta}{2} \right] < \frac{\varepsilon}{2},$$

Using the triangular inequality, we obtain:

$$\forall\varepsilon > 0, \eta > 0, \exists T_{\varepsilon,\eta} / \forall T \geq T_{\varepsilon,\eta}, \forall \theta_1, \theta_{21} \in \Theta_1 \times \Theta_{21},$$

$$P_* \left[\left\| \widetilde{\beta}_{1,TS} \left(\theta_1, \theta_{21}, \widehat{\theta}_{22,TS} \right) - \widetilde{\beta}_1 \left(\theta_1, \theta_{21}, \overline{\theta}_{22} \right) \right\|_{q_1} > \eta \right] < \varepsilon,$$

$$\Longrightarrow \quad \underset{\theta_1, \theta_{21} \in \Theta_1 \times \Theta_{21}}{Sup} \left\| \widetilde{\beta}_{1,TS} \left(\theta_1, \theta_{21}, \widehat{\theta}_{22,TS} \right) - \widetilde{\beta}_1 \left(\theta_1, \theta_{21}, \overline{\theta}_{22} \right) \right\|_{q_1} \xrightarrow[T \to +\infty]{P_*} 0 \cdot$$

Thus, by virtue of the uniform convergence in probability of the associated criterion, we have $P_* \lim_{T \to +\infty} \widehat{\theta}^1_{TS} \left(\widehat{\theta}_{22,TS} \right) = \theta^{1^*} = \underset{\theta_1, \theta_{21} \in \Theta_1 \times \Theta_{21}}{Argmin} \left\| \beta_1^\circ - \widetilde{\beta}_1 \left(\theta_1, \theta_{21}, \overline{\theta}_{22} \right) \right\|_{\Omega_1}$. Under the partial encompassing property (5.6), $\theta^{1^*} = \underset{\theta_1, \theta_{21} \in \Theta_1 \times \Theta_{21}}{Argmin} \left\| \widetilde{\beta}_1 \left(\theta_1^\circ, \overline{\theta}_{21}, \overline{\theta}_{22} \right) - \widetilde{\beta}_1 \left(\theta_1, \theta_{21}, \overline{\theta}_{22} \right) \right\|_{\Omega_1}$. And under the partial indirect identification (3.7), we obtain $\widetilde{\beta}_1 \left(\theta_1^\circ, \overline{\theta}_{21}, \overline{\theta}_{22} \right) = \widetilde{\beta}_1 \left(\theta_1^*, \theta_{21}^*, \overline{\theta}_{22} \right) \Longrightarrow \theta^{1^*} = \left(\theta_1^{\circ\prime}, \overline{\theta}_{21}' \right)'$. This proves the consistency of the SII estimator in case of harmful nuisance parameters.

A.2. Proofs of Propositions 4.1, 4.2, 4.5:

First order conditions for the indirect estimator $\widehat{\theta}_{\mathbf{TS}}$:

The first order conditions corresponding to the optimization problem:

$$\underset{(\theta_1, \theta_2) \in \Theta_1 \times \Theta_2}{Min} \left[\widehat{\beta}_T - \frac{1}{S} \sum_{s=1}^S \widetilde{\beta}_T^s(\theta_1, \theta_2) \right]' \widehat{\Omega}_T \left[\widehat{\beta}_T - \frac{1}{S} \sum_{s=1}^S \widetilde{\beta}_T^s(\theta_1, \theta_2) \right]$$

are:

$$\frac{1}{S} \sum_{s=1}^S \frac{\partial \widetilde{\beta}_T^{s\prime}}{\partial \theta} (\widehat{\theta}_{1,TS}, \widehat{\theta}_{2,TS}) \widehat{\Omega}_T \sqrt{T} \left[\widehat{\beta}_T - \frac{1}{S} \sum_{s=1}^S \widetilde{\beta}_T^s(\widehat{\theta}_{1,TS}, \widehat{\theta}_{2,TS}) \right] = 0 \cdot$$

The expansion of the first order conditions around the limit value $\left(\theta_1^{\circ\prime}, \overline{\theta}_{21}{}' \right)'$ gives:

$$\frac{1}{S} \sum_{s=1}^S \frac{\partial \widetilde{\beta}_T^{s\prime}}{\partial \theta} (\theta_1^\circ, \overline{\theta}_2) \Omega \sqrt{T} \left[\widehat{\beta}_T - \frac{1}{S} \sum_{s=1}^S \widetilde{\beta}_T^s(\theta_1^\circ, \overline{\theta}_2) - \frac{1}{S} \sum_{s=1}^S \frac{\partial \widetilde{\beta}_T^s}{\partial \theta'} (\theta_1^\circ, \overline{\theta}_2) \begin{pmatrix} \widehat{\theta}_{1,TS} - \theta_1^\circ \\ \widehat{\theta}_{2,TS} - \overline{\theta}_2 \end{pmatrix} \right] = o_{P_*}(1),$$

which leads to:

$$\sqrt{T} \begin{pmatrix} \widehat{\theta}_{1,TS} - \theta_1^\circ \\ \widehat{\theta}_{2,TS} - \overline{\theta}_2 \end{pmatrix} = \left\{ \frac{\partial \widetilde{\beta}'}{\partial \theta} (\theta_1^\circ, \overline{\theta}_2) \Omega \frac{\partial \widetilde{\beta}}{\partial \theta'} (\theta_1^\circ, \overline{\theta}_2) \right\}^{-1} \times \frac{\partial \widetilde{\beta}'}{\partial \theta} (\theta_1^\circ, \overline{\theta}_2) \Omega \sqrt{T} \left\{ \widehat{\beta}_T - \frac{1}{S} \sum_{s=1}^S \widetilde{\beta}_T^s(\theta_1^\circ, \overline{\theta}_2) \right\} + o_{P_*}(1),$$

since under assumption $(A8g)$ we have:

$$P_* \lim_{T \to +\infty} \frac{1}{S} \sum_{s=1}^S \frac{\partial \widetilde{\beta}_T^{s\prime}}{\partial \theta} (\theta_1^\circ, \overline{\theta}_2) = \frac{\partial \widetilde{\beta}'}{\partial \theta} (\theta_1^\circ, \overline{\theta}_2),$$

$$\frac{\partial \widetilde{\beta}}{\partial \theta'} (\theta_1^\circ, \overline{\theta}_2) \text{ is of full-column rank } (p) \cdot$$

Expansions of $\widehat{\beta}_{\mathbf{T}}$ and $\widetilde{\beta}^{\mathbf{s}}_{\mathbf{T}}(\theta^{\circ}_{\mathbf{1}}, \overline{\theta}_{\mathbf{2}})$:

We begin with the first-order conditions on the instrumental criterion:

$$\frac{\partial Q_T}{\partial \beta}\left(\underline{y}_T, \underline{x}_T, \widehat{\beta}_T\right) = 0\cdot$$

The expansion of the latter equation around the limit value β° gives:

$$\sqrt{T}\frac{\partial Q_T}{\partial \beta}\left(\underline{y}_T, \underline{x}_T, \beta^{\circ}\right) + \frac{\partial^2 Q_T}{\partial \beta \partial \beta'}\left(\underline{y}_T, \underline{x}_T, \beta^{\circ}\right)\sqrt{T}\left[\widehat{\beta}_T - \beta^{\circ}\right] = o_{P_{\circ}}(1),$$

which leads to:

$$\sqrt{T}\left[\widehat{\beta}_T - \beta^{\circ}\right] = -J_{\circ}^{-1}\sqrt{T}\frac{\partial Q_T}{\partial \beta}\left(\underline{y}_T, \underline{x}_T, \beta^{\circ}\right) + o_{P_{\circ}}(1)\cdot$$

We have by using the same arguments:

$$\sqrt{T}\left[\widetilde{\beta}^s_T(\theta^{\circ}_1, \overline{\theta}_2) - \widetilde{\beta}(\theta^{\circ}_1, \overline{\theta}_2)\right] = -J_{\circ}^{*-1}\sqrt{T}\frac{\partial Q_T}{\partial \beta}\left(\underline{\widetilde{y}}^s_T(\theta^{\circ}_1, \overline{\theta}_2, z^s_{\circ}), \underline{x}_T, \widetilde{\beta}(\theta^{\circ}_1, \overline{\theta}_2)\right) + o_{P_*}(1),$$

and thanks to the full-encompassing hypothesis $H_{\circ}$: $\beta^{\circ} = \widetilde{\beta}(\theta^{\circ}_1, \overline{\theta}_2)$ we are led to:

$$\sqrt{T}\left[\widetilde{\beta}^s_T(\theta^{\circ}_1, \overline{\theta}_2) - \beta^{\circ}\right] = -J_{\circ}^{*-1}\sqrt{T}\frac{\partial Q_T}{\partial \beta}\left(\underline{\widetilde{y}}^s_T(\theta^{\circ}_1, \overline{\theta}_2, z^s_{\circ}), \underline{x}_T, \beta^{\circ}\right) + o_{P_*}(1)\cdot$$

Asymptotic distribution of $\sqrt{\mathbf{T}}\left[\widehat{\beta}_{\mathbf{T}} - \frac{\mathbf{1}}{\mathbf{S}}\sum_{\mathbf{s=1}}^{\mathbf{S}}\widetilde{\beta}^{\mathbf{s}}_{\mathbf{T}}(\theta^{\circ}_{\mathbf{1}}, \overline{\theta}_{\mathbf{2}})\right]$:

$$\begin{aligned}\sqrt{T}\left[\widehat{\beta}_T - \frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}^s_T(\theta^{\circ}_1, \overline{\theta}_2)\right] &= -J_{\circ}^{-1}\sqrt{T}\frac{\partial Q_T}{\partial \beta}\left(\underline{y}_T, \underline{x}_T, \beta^{\circ}\right) + \\ J_{\circ}^{*-1}\sqrt{T}\frac{1}{S}\sum_{s=1}^{S}\frac{\partial Q_T}{\partial \beta}&\left(\underline{\widetilde{y}}^s_T\left(\theta^{\circ}_1, \overline{\theta}_2, z^s_{\circ}\right), \underline{x}_T, \beta^{\circ}\right) + o_{P_*}(1)\cdot\end{aligned}$$

Under assumptions $(A1)-(A8)$, $\sqrt{T}\left[\widehat{\beta}_T - \frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}^s_T(\theta^{\circ}_1, \overline{\theta}_2)\right]$ is asymptotically normally distributed with mean zero and a covariance matrix given by $\Phi^*_{\circ}(S)$:

$$\Phi^*_{\circ}(S) = J_{\circ}^{-1}I_{\circ}J_{\circ}^{-1} + \frac{1}{S}J_{\circ}^{*-1}I^*_{\circ}J_{\circ}^{*-1} + \left(1 - \frac{1}{S}\right)J_{\circ}^{*-1}K^*_{\circ}J_{\circ}^{*-1} - J_{\circ}^{-1}K_{\circ}J_{\circ}^{*-1} - J_{\circ}^{*-1}K'_{\circ}J_{\circ}^{-1},$$

and the result of Proposition 4.1 follows. As usual the optimal choice of the matrix Ω which minimizes the asymptotic variance of the indirect inference estimator is $\Omega^* = \Phi^*_{\circ}(S)^{-1}$ and the result of Proposition 4.2 follows.

Proof of Proposition 4.5:

The optimal value of the objective function is:

$$\xi_{T,S} = T\left[\widehat{\beta}_T - \frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}^s_T(\widehat{\theta}_{1,TS}, \widehat{\theta}_{2,TS})\right]'\widehat{\Omega}^*_T\left[\widehat{\beta}_T - \frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}^s_T(\widehat{\theta}_{1,TS}, \widehat{\theta}_{2,TS})\right],$$

where $\left(\widehat{\theta}'_{1,TS}, \widehat{\theta}'_{2,TS}\right)'$ corresponds to the optimal indirect inference estimator. The first order expansion of $\xi_{T,S}$ around the limit value $\left(\theta^{\circ\prime}_1, \overline{\theta}_2{}'\right)'$ gives:

$$\begin{aligned}\xi_{T,S} = \; & T\left[\widehat{\beta}_T - \frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}^s_T(\theta^{\circ}_1, \overline{\theta}_2) - \frac{\partial \widetilde{\beta}}{\partial \theta'}(\theta^{\circ}_1, \overline{\theta}_2)\begin{pmatrix}\widehat{\theta}_{1,TS} - \theta^{\circ}_1 \\ \widehat{\theta}_{2,TS} - \overline{\theta}_2\end{pmatrix}\right]'\Omega^*\left[\widehat{\beta}_T - \frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}^s_T(\theta^{\circ}_1, \overline{\theta}_2) - \right. \\ & \left.\frac{\partial \widetilde{\beta}}{\partial \theta'}(\theta^{\circ}_1, \overline{\theta}_2)\begin{pmatrix}\widehat{\theta}_{1,TS} - \theta^{\circ}_1 \\ \widehat{\theta}_{2,TS} - \overline{\theta}_2\end{pmatrix}\right] + o_{P_*}(1)\cdot\end{aligned}$$

By using the asymptotic expansion of $\sqrt{T}\begin{pmatrix}\widehat{\theta}_{1,TS}-\theta_1^{\circ}\\ \widehat{\theta}_{2,TS}-\overline{\theta}_2\end{pmatrix}$ around the limit value $\left(\theta_1^{\circ\prime},\overline{\theta}_2^{\prime}\right)^{\prime}$ previously given, we get:

$$\begin{aligned}\frac{\partial\widetilde{\beta}}{\partial\theta'}(\theta_1^{\circ},\overline{\theta}_2)\sqrt{T}\begin{pmatrix}\widehat{\theta}_{1,TS}-\theta_1^{\circ}\\ \widehat{\theta}_{2,TS}-\overline{\theta}_2\end{pmatrix} &= \frac{\partial\widetilde{\beta}}{\partial\theta'}(\theta_1^{\circ},\overline{\theta}_2)\left\{\frac{\partial\widetilde{\beta}'}{\partial\theta}(\theta_1^{\circ},\overline{\theta}_2)\times\right.\\ &\left.\Omega^*\frac{\partial\widetilde{\beta}}{\partial\theta'}(\theta_1^{\circ},\overline{\theta}_2)\right\}^{-1}\frac{\partial\widetilde{\beta}'}{\partial\theta}(\theta_1^{\circ},\overline{\theta}_2)\Omega^*\sqrt{T}\left[\widehat{\beta}_T-\frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}_T^s(\theta_1^{\circ},\overline{\theta}_2)\right]+o_{P_*}(1),\end{aligned}$$

and thus:

$$\begin{aligned}\sqrt{T}\left[\widehat{\beta}_T-\frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}_T^s(\theta_1^{\circ},\overline{\theta}_2)-\frac{\partial\widetilde{\beta}}{\partial\theta'}(\theta_1^{\circ},\overline{\theta}_2)\begin{pmatrix}\widehat{\theta}_{1,TS}-\theta_1^{\circ}\\ \widehat{\theta}_{2,TS}-\overline{\theta}_2\end{pmatrix}\right] &= [Id_q-M]\sqrt{T}\left[\widehat{\beta}_T-\frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}_T^s(\theta_1^{\circ},\overline{\theta}_2)\right]\\ &+o_{P_*}(1),\end{aligned}$$

where M is the orthogonal projector on the space spanned by the columns of $\frac{\partial\widetilde{\beta}}{\partial\theta'}(\theta_1^{\circ},\overline{\theta}_2)$ for the inner product Ω^* that is:

$$M=\frac{\partial\widetilde{\beta}}{\partial\theta'}(\theta_1^{\circ},\overline{\theta}_2)\left\{\frac{\partial\widetilde{\beta}'}{\partial\theta}(\theta_1^{\circ},\overline{\theta}_2)\Omega^*\frac{\partial\widetilde{\beta}}{\partial\theta'}(\theta_1^{\circ},\overline{\theta}_2)\right\}^{-1}\frac{\partial\widetilde{\beta}'}{\partial\theta}(\theta_1^{\circ},\overline{\theta}_2)\Omega^*\cdot$$

With these notations, the statistic $\xi_{T,S}$ is equal to:

$$\xi_{T,S}=T\left[\widehat{\beta}_T-\frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}_T^s(\theta_1^{\circ},\overline{\theta}_2)\right]'(Id_q-M)'\Omega^*(Id_q-M)\left[\widehat{\beta}_T-\frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}_T^s(\theta_1^{\circ},\overline{\theta}_2)\right]+o_{P_*}(1)\cdot$$

As previously seen $\sqrt{T}\left[\widehat{\beta}_T-\frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}_T^s(\theta_1^{\circ},\overline{\theta}_2)\right]\xrightarrow[T\to+\infty]{D}\mathcal{N}\left[0,\ \Omega^{*-1}\right]$ and $\frac{\partial\widetilde{\beta}}{\partial\theta'}(\theta_1^{\circ},\overline{\theta}_2)$ is of full-column rank (p) which implies that:

$$T\left[\widehat{\beta}_T-\frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}_T^s(\theta_1^{\circ},\overline{\theta}_2)\right]'(Id_q-M)'\Omega^*(Id_q-M)\left[\widehat{\beta}_T-\frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}_T^s(\theta_1^{\circ},\overline{\theta}_2)\right]\xrightarrow[T\to+\infty]{D}\chi^2(q-p),$$

and the result of Proposition 4.5 follows.

A.3. Proofs of Propositions 4.3, 4.4, 4.6:

First order conditions for the indirect estimator $\widehat{\theta}^{1}_{1,TS}(\overline{\theta}_{22})$:

The first order conditions corresponding to the optimization problem:

$$\underset{(\theta_1,\theta_{21})\in\Theta_1\times\Theta_{21}}{Min}\left[\widehat{\beta}_{1,T}-\frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}_{1,T}^s(\theta_1,\theta_{21},\widehat{\theta}_{22,TS})\right]'\widehat{\Omega}_{1,T}\left[\widehat{\beta}_{1,T}-\frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}_{1,T}^s(\theta_1,\theta_{21},\widehat{\theta}_{22,TS})\right],$$

where $\widehat{\theta}_{22,TS}$ is a consistent estimator of the value $\overline{\theta}_{22}$ of the nuisance parameters θ_{22} and such that $\sqrt{T}\left(\widehat{\theta}_{22,TS}-\overline{\theta}_{22}\right)=O_{P_*}(1)$, are:

$$\begin{aligned}&\frac{1}{S}\sum_{s=1}^{S}\frac{\partial\widetilde{\beta}_{1,T}^{s\ \prime}}{\partial\begin{pmatrix}\theta_1\\ \theta_{21}\end{pmatrix}}(\widehat{\theta}^1_{1,TS}(\overline{\theta}_{22}),\widehat{\theta}^1_{21,TS}(\overline{\theta}_{22}),\widehat{\theta}_{22,TS})\widehat{\Omega}_{1,T}\sqrt{T}\left[\widehat{\beta}_{1,T}-\frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}_{1,T}^s(\widehat{\theta}^1_{1,TS}(\overline{\theta}_{22}),\widehat{\theta}^1_{21,TS}(\overline{\theta}_{22}),\widehat{\theta}_{22,TS})\right]\\ &=0\cdot\end{aligned}$$

The expansion of the first order conditions around the limit value $\left(\theta_1^{\circ'}, \overline{\theta}_2'\right)'$ gives:

$$\frac{1}{S}\sum_{s=1}^{S}\frac{\partial\widetilde{\beta}_{1,T}^{s}{}'}{\partial\begin{pmatrix}\theta_1\\ \theta_{21}\end{pmatrix}}(\theta_1^\circ,\overline{\theta}_2)\Omega_1\sqrt{T}\left[\widehat{\beta}_{1,T}-\frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}_{1,T}^{s}(\theta_1^\circ,\overline{\theta}_2)-\frac{1}{S}\sum_{s=1}^{S}\frac{\partial\widetilde{\beta}_{1,T}^{s}}{\partial\begin{pmatrix}\theta_1\\ \theta_{21}\end{pmatrix}'}(\theta_1^\circ,\overline{\theta}_2)\right.$$
$$\left.\begin{pmatrix}\widehat{\theta}_{1,TS}^{1}(\overline{\theta}_{22})-\theta_1^\circ\\ \widehat{\theta}_{21,TS}^{1}(\overline{\theta}_{22})-\overline{\theta}_{21}\end{pmatrix}\right]=o_{P_*}(1),$$

since under $H_\circ^1$, $\dfrac{\partial\widetilde{\beta}_1}{\partial\theta_{22}'}(\theta_1,\theta_2)=0$. This leads to:

$$\sqrt{T}\begin{bmatrix}\widehat{\theta}_{1,TS}^{1}(\overline{\theta}_{22})-\theta_1^\circ\\ \widehat{\theta}_{21,TS}^{1}(\overline{\theta}_{22})-\overline{\theta}_{21}\end{bmatrix}=\left[\frac{\partial\widetilde{\beta}_1'}{\partial\begin{bmatrix}\theta_1\\ \theta_{21}\end{bmatrix}}(\theta_1^\circ,\overline{\theta}_{21})\Omega_1\frac{\partial\widetilde{\beta}_1}{\partial\begin{bmatrix}\theta_1\\ \theta_{21}\end{bmatrix}'}(\theta_1^\circ,\overline{\theta}_{21})\right]^{-1}$$
$$\frac{\partial\widetilde{\beta}_1'}{\partial\begin{bmatrix}\theta_1\\ \theta_{21}\end{bmatrix}}(\theta_1^\circ,\overline{\theta}_{21})\Omega_1\sqrt{T}\left[\widehat{\beta}_{1,T}-\frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}_{1,T}^{s}(\theta_1^\circ,\overline{\theta}_2)\right]+o_{P_*}(1),$$

since under assumption $(A9e)$, we have:

$$\underset{T\to+\infty}{P_*\lim}\frac{1}{S}\sum_{s=1}^{S}\frac{\partial\widetilde{\beta}_{1,T}^{s}{}'}{\partial\begin{pmatrix}\theta_1\\ \theta_{21}\end{pmatrix}}(\theta_1^\circ,\overline{\theta}_2)=\frac{\partial\widetilde{\beta}_1'}{\partial\begin{pmatrix}\theta_1\\ \theta_{21}\end{pmatrix}}(\theta_1^\circ,\overline{\theta}_2),$$

and under $H_\circ^1$:

$$\frac{\partial\widetilde{\beta}_1}{\partial\begin{pmatrix}\theta_1\\ \theta_{21}\end{pmatrix}'}(\theta_1^\circ,\overline{\theta}_2)=\frac{\partial\widetilde{\beta}_1}{\partial\begin{pmatrix}\theta_1\\ \theta_{21}\end{pmatrix}'}(\theta_1^\circ,\overline{\theta}_{21}),$$

is of full-column rank (p_1+p_{21}) thanks to $(A9e)$.

Expansions of $\widehat{\beta}_{1,T}$ and $\widetilde{\beta}_{1,T}^{s}(\theta_1^\circ,\overline{\theta}_2)$:

We begin with the expansion of the first-order conditions on the instrumental model around the limit value β°:

$$\frac{\partial Q_T}{\partial\beta}\left(\underline{y}_T,\underline{x}_T,\widehat{\beta}_T\right)=0\cdot$$

The expansion of the latter equation around the limit value β° gives:

$$\sqrt{T}\frac{\partial Q_T}{\partial\beta}\left(\underline{y}_T,\underline{x}_T,\beta^\circ\right)+\frac{\partial^2 Q_T}{\partial\beta\partial\beta'}\left(\underline{y}_T,\underline{x}_T,\beta^\circ\right)\sqrt{T}\left[\widehat{\beta}_T-\beta^\circ\right]=o_{P_\circ}(1),$$

which leads to:

$$\sqrt{T}\left[\widehat{\beta}_T-\beta^\circ\right]=-J_\circ^{-1}\sqrt{T}\frac{\partial Q_T}{\partial\beta}\left(\underline{y}_T,\underline{x}_T,\beta^\circ\right)+o_{P_\circ}(1)\cdot$$

We have by using the same argument:

$$\sqrt{T}\left[\widetilde{\beta}_T^{s}(\theta_1^\circ,\overline{\theta}_2)-\widetilde{\beta}^\circ(\overline{\theta}_{22})\right]=-J_\circ^{*\,-1}(\overline{\theta}_{22})\sqrt{T}\frac{\partial Q_T}{\partial\beta}\left(\underline{\widetilde{y}}_T^{s}(\theta_1^\circ,\overline{\theta}_2,z_\circ^s),\underline{x}_T,\widetilde{\beta}^\circ(\overline{\theta}_{22})\right)+o_{P_*}(1)\cdot$$

Asymptotic distribution of $\sqrt{\mathbf{T}}\left[\widehat{\beta}_{\mathbf{1,T}} - \frac{\mathbf{1}}{\mathbf{S}}\sum_{\mathbf{s=1}}^{\mathbf{S}} \widetilde{\beta}^{\mathbf{s}}_{\mathbf{1,T}}(\theta^\circ_{\mathbf{1}}, \overline{\theta}_{\mathbf{2}})\right]$:

We have:

$$\sqrt{T}\left[\widehat{\beta}_T - \beta^\circ - \frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}^s_T(\theta^\circ_1, \overline{\theta}_2) + \widetilde{\beta}^\circ(\overline{\theta}_{22})\right] = -J_\circ^{-1}\sqrt{T}\frac{\partial Q_T}{\partial \beta}\left(\underline{y}_T, \underline{x}_T, \beta^\circ\right)$$
$$+J_\circ^{*-1}(\overline{\theta}_{22})\sqrt{T}\frac{1}{S}\sum_{s=1}^{S}\frac{\partial Q_T}{\partial \beta}\left(\underline{\widetilde{y}}^s_T\left(\theta^\circ_1, \overline{\theta}_2, z^s_\circ\right), \underline{x}_T, \widetilde{\beta}^\circ(\overline{\theta}_{22})\right) + o_{P_*}(1)\cdot$$

The statistic $\sqrt{T}\left[\widehat{\beta}_T - \beta^\circ - \frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}^s_T(\theta^\circ_1, \overline{\theta}_2) + \widetilde{\beta}^\circ(\overline{\theta}_{22})\right]$ is, under assumptions $(A1)-(A8b)/(A9)$, asymptotically normally distributed with mean zero and a covariance matrix given by $\Phi^*_\circ(S, \overline{\theta}_{22})$:

$$\Phi^*_\circ(S, \overline{\theta}_{22}) = J_\circ^{-1}I_\circ J_\circ^{-1} + \frac{1}{S}J_\circ^{*-1}(\overline{\theta}_{22})I^*_\circ(\overline{\theta}_{22})J_\circ^{*-1}(\overline{\theta}_{22}) + \left(1 - \frac{1}{S}\right)J_\circ^{*-1}(\overline{\theta}_{22})K^*_\circ(\overline{\theta}_{22})J_\circ^{*-1}(\overline{\theta}_{22})$$
$$-J_\circ^{-1}K_\circ(\overline{\theta}_{22})J_\circ^{*-1}(\overline{\theta}_{22}) - J_\circ^{*-1}(\overline{\theta}_{22})K'_\circ(\overline{\theta}_{22})J_\circ^{-1}.$$

Let $\Phi^*_{\circ,1}(S, \overline{\theta}_{22})$ be the $(q_1 \times q_1)$ left-upper bloc diagonal sub-matrix of the $(q \times q)$ matrix $\Phi^*_\circ(S, \overline{\theta}_{22})$. We have thanks to the partial-encompassing hypothesis $H^1_\circ$:
$\beta^\circ_1 = \widetilde{\beta}_1(\theta^\circ_1, \overline{\theta}_{21})$ which leads to:

$$\sqrt{T}\left[\widehat{\beta}_{1,T} - \beta^\circ_1 - \frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}^s_{1,T}(\theta^\circ_1, \overline{\theta}_{21}) + \widetilde{\beta}_1(\theta^\circ_1, \overline{\theta}_{21})\right] = \sqrt{T}\left[\widehat{\beta}_{1,T} - \frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}^s_{1,T}(\theta^\circ_1, \overline{\theta}_{21})\right]\cdot$$

The statistic $\sqrt{T}\left[\widehat{\beta}_{1,T} - \frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}^s_{1,T}(\theta^\circ_1, \overline{\theta}_{21})\right]$ is asymptotically normally distributed with mean zero and a covariance matrix given by $\Phi^*_{\circ,1}(S, \overline{\theta}_{22})$ and the result of Proposition 4.3 follows. As usual the optimal choice of the matrix Ω_1 which minimizes the asymptotic covariance of the indirect inference estimator based on the sub-vector binding function is $\Omega^*_1(\overline{\theta}_{22}) = \Phi^*_{\circ,1}(S, \overline{\theta}_{22})^{-1}$ and the result of Proposition 4.4 follows.

Proof of Proposition 4.6:

The optimal value of the objective function is:

$$\xi^1_{T,S}(\overline{\theta}_{22}) = \; T\left[\widehat{\beta}_{1,T} - \frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}^s_{1,T}(\widehat{\theta}^1_{1,TS}(\overline{\theta}_{22}), \widehat{\theta}^1_{21,TS}(\overline{\theta}_{22}), \widehat{\theta}_{22,TS})\right]'\widehat{\Omega}^*_{1,T}$$
$$\left[\widehat{\beta}_{1,T} - \frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}^s_{1,T}(\widehat{\theta}^1_{1,TS}(\overline{\theta}_{22}), \widehat{\theta}^1_{21,TS}(\overline{\theta}_{22}), \widehat{\theta}_{22,TS})\right],$$

where $\left(\widehat{\theta}^{1'}_{1,TS}(\overline{\theta}_{22}), \widehat{\theta}^{1'}_{21,TS}(\overline{\theta}_{22})\right)'$ corresponds to the optimal indirect inference estimator and $\widehat{\theta}_{22,TS}$ to a consistent estimator of $\overline{\theta}_{22}$ and such that $\sqrt{T}\left(\widehat{\theta}_{22,TS} - \overline{\theta}_{22}\right) = O_{P_*}(1)$. The first order expansion of $\xi^1_{T,S}(\overline{\theta}_{22})$ around the limit value $\left(\theta_1^{\circ'}, \overline{\theta}_2'\right)'$ gives:

$$\xi^1_{T,S}(\overline{\theta}_{22}) = T\left[\widehat{\beta}_{1,T} - \frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}^s_{1,T}(\theta_1^\circ, \overline{\theta}_2) - \frac{\partial \widetilde{\beta}_1}{\partial \begin{pmatrix} \theta_1 \\ \theta_{21} \end{pmatrix}'}(\theta_1^\circ, \overline{\theta}_{21}) \begin{pmatrix} \widehat{\theta}^1_{1,TS}(\overline{\theta}_{22}) - \theta_1^\circ \\ \widehat{\theta}^1_{21,TS}(\overline{\theta}_{22}) - \overline{\theta}_{21} \end{pmatrix}\right]' \times$$

$$\Omega_1^*(\overline{\theta}_{22})\left[\widehat{\beta}_{1,T} - \frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}^s_{1,T}(\theta_1^\circ, \overline{\theta}_2) - \frac{\partial \widetilde{\beta}_1}{\partial \begin{pmatrix} \theta_1 \\ \theta_{21} \end{pmatrix}'}(\theta_1^\circ, \overline{\theta}_{21}) \begin{pmatrix} \widehat{\theta}^1_{1,TS}(\overline{\theta}_{22}) - \theta_1^\circ \\ \widehat{\theta}^1_{21,TS}(\overline{\theta}_{22}) - \overline{\theta}_{21} \end{pmatrix}\right] + o_{P_*}(1),$$

since under $H^1_\circ$, $\dfrac{\partial \widetilde{\beta}_1}{\partial \theta_{22}'}(\theta_1, \theta_2) = 0$. By using the asymptotic expansion of $\sqrt{T}\begin{pmatrix} \widehat{\theta}^1_{1,TS}(\overline{\theta}_{22}) - \theta_1^\circ \\ \widehat{\theta}^1_{21,TS}(\overline{\theta}_{22}) - \overline{\theta}_{21} \end{pmatrix}$ around the limit value $\left(\theta_1^{\circ'}, \overline{\theta}_{21}'\right)'$ previously given, we get:

$$\frac{\partial \widetilde{\beta}_1}{\partial \begin{pmatrix} \theta_1 \\ \theta_{21} \end{pmatrix}'}(\theta_1^\circ, \overline{\theta}_{21})\sqrt{T}\begin{pmatrix} \widehat{\theta}^1_{1,TS}(\overline{\theta}_{22}) - \theta_1^\circ \\ \widehat{\theta}^1_{21,TS}(\overline{\theta}_{22}) - \overline{\theta}_{21} \end{pmatrix} = \frac{\partial \widetilde{\beta}_1}{\partial \begin{pmatrix} \theta_1 \\ \theta_{21} \end{pmatrix}'}(\theta_1^\circ, \overline{\theta}_{21})\left[\frac{\partial \widetilde{\beta}_1'}{\partial \begin{pmatrix} \theta_1 \\ \theta_{21} \end{pmatrix}}(\theta_1^\circ, \overline{\theta}_{21})\Omega_1^*(\overline{\theta}_{22})\right.$$

$$\left.\frac{\partial \widetilde{\beta}_1}{\partial \begin{pmatrix} \theta_1 \\ \theta_{21} \end{pmatrix}'}(\theta_1^\circ, \overline{\theta}_{21})\right]^{-1} \frac{\partial \widetilde{\beta}_1'}{\partial \begin{pmatrix} \theta_1 \\ \theta_{21} \end{pmatrix}}(\theta_1^\circ, \overline{\theta}_{21})\Omega_1^*(\overline{\theta}_{22}) \times \sqrt{T}\left[\widehat{\beta}_{1,T} - \frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}^s_{1,T}(\theta_1^\circ, \overline{\theta}_{21})\right] + o_{P_*}(1),$$

and thus:

$$\sqrt{T}\left[\widehat{\beta}_{1,T} - \frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}^s_{1,T}(\theta_1^\circ, \overline{\theta}_2) - \frac{\partial \widetilde{\beta}_1}{\partial \begin{pmatrix} \theta_1 \\ \theta_{21} \end{pmatrix}'}(\theta_1^\circ, \overline{\theta}_{21}) \begin{pmatrix} \widehat{\theta}^1_{1,TS}(\overline{\theta}_{22}) - \theta_1^\circ \\ \widehat{\theta}^1_{21,TS}(\overline{\theta}_{22}) - \overline{\theta}_{21} \end{pmatrix}\right] = [Id_{q_1} - M_1] \times$$

$$\sqrt{T}\left[\widehat{\beta}_{1,T} - \frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}^s_{1,T}(\theta_1^\circ, \overline{\theta}_2)\right] + o_{P_*}(1),$$

where M_1 is the orthogonal projector on the space spanned by the columns of $\dfrac{\partial \widetilde{\beta}_1}{\partial \begin{pmatrix} \theta_1 \\ \theta_{21} \end{pmatrix}'}(\theta_1^\circ, \overline{\theta}_{21})$ for the inner product $\Omega_1^*(\overline{\theta}_{22})$ that is:

$$M_1 = \frac{\partial \widetilde{\beta}_1}{\partial \begin{pmatrix} \theta_1 \\ \theta_{21} \end{pmatrix}'}(\theta_1^\circ, \overline{\theta}_{21})\left\{\frac{\partial \widetilde{\beta}_1'}{\partial \begin{pmatrix} \theta_1 \\ \theta_{21} \end{pmatrix}}(\theta_1^\circ, \overline{\theta}_{21})\Omega_1^*(\overline{\theta}_{22})\frac{\partial \widetilde{\beta}_1}{\partial \begin{pmatrix} \theta_1 \\ \theta_{21} \end{pmatrix}'}(\theta_1^\circ, \overline{\theta}_{21})\right\}^{-1} \frac{\partial \widetilde{\beta}_1'}{\partial \begin{pmatrix} \theta_1 \\ \theta_{21} \end{pmatrix}}(\theta_1^\circ, \overline{\theta}_{21})\Omega_1^*(\overline{\theta}_{22}) \cdot$$

With these notations the statistic $\xi^1_{T,S}(\overline{\theta}_{22})$ is equal to:

$$\begin{aligned} \xi^1_{T,S}(\overline{\theta}_{22}) &= T\left[\widehat{\beta}_{1,T} - \frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}^s_{1,T}(\theta_1^\circ, \overline{\theta}_2)\right]'(Id_{q_1} - M_1)'\Omega_1^*(\overline{\theta}_{22})\left(Id_{q_1} - M_1\right)\left[\widehat{\beta}_{1,T} - \frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}^s_{1,T}(\theta_1^\circ, \overline{\theta}_2)\right] \\ &\quad + o_{P_*}(1) \cdot \end{aligned}$$

As previously seen we have $\sqrt{T}\left[\widehat{\beta}_{1,T}-\frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}_{1,T}^{s}(\theta_1^{\circ},\overline{\theta}_2)\right] \xrightarrow[T\to+\infty]{D} \mathcal{N}\left(0,\ \Omega_1^{*}(\overline{\theta}_{22})^{-1}\right)$ and $\frac{\partial\widetilde{\beta}_1}{\partial\begin{pmatrix}\theta_1\\ \theta_{21}\end{pmatrix}'}(\theta_1^{\circ},\overline{\theta}_{21})$ is of full-column rank (p_1+p_{21}) which implies that:

$$
\begin{aligned}
&T\left[\widehat{\beta}_{1,T}-\frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}_{1,T}^{s}(\theta_1^{\circ},\overline{\theta}_2)\right]'(Id_{q_1}-M_1)'\Omega_1^{*}(\overline{\theta}_{22})\,(Id_{q_1}-M_1)\left[\widehat{\beta}_{1,T}-\frac{1}{S}\sum_{s=1}^{S}\widetilde{\beta}_{1,T}^{s}(\theta_1^{\circ},\overline{\theta}_2)\right]\\
&\xrightarrow[T\to+\infty]{D} \chi^2(q_1-p_1-p_{21}),
\end{aligned}
$$

and the result of Proposition 4.6 follows.

A.4. Proofs of proposition 6.1 and lemma 6.1:

Asymmetries: leverage effect versus skewness of the standardized innovations:
For sake of simplicity, we focus on SV models of the following form:

$$
\begin{aligned}
y_t &= \alpha y_{t-1}+\sigma_{t-1}\varepsilon_t, \quad |\alpha|<1,\\
\sigma_t^2 &= \omega+\gamma\sigma_{t-1}^2+\nu_t, \quad 0<\omega,\ 0<\gamma<1,
\end{aligned}
$$

and we assume in addition that:

$$
\begin{aligned}
&E\left[\varepsilon_t/I_{t-1}\right]=0, \qquad E\left[\varepsilon_t^2/I_{t-1}\right]=1, \quad E\left[\varepsilon_t^3/I_{t-1}\right]=\mu_3,\\
&E\left[\varepsilon_t^4/I_{t-1}\right]=\mu_4, \qquad E\left[\nu_t/I_{t-1}\right]=0, \quad E\left[\nu_t^2/I_{t-1}\right]=\eta^2,\\
&E\left[\varepsilon_t\nu_t/I_{t-1}\right]=\rho\eta,
\end{aligned}
$$

that is an $AR(1)$ process with SV innovations. However, the analysis can be extended to any $ARMA(p,q)$. In this context, we have:

$$
\begin{aligned}
&\underset{\theta}{E}\left[y_t\right]=0,\\
&\underset{\theta}{E}\left[y_t^3\right]=\frac{\mu_3}{1-\alpha^3}\underset{\theta}{E}\left[\sigma_t^3\right]+\frac{3\alpha\rho\eta}{(1-\alpha^3)\,(1-\alpha\gamma)}\underset{\theta}{E}\left[\sigma_t\right]\cdot
\end{aligned}
$$

Therefore there are two ways to capture the asymmetries: on the one hand the skewness of the standardized innovations ($\mu_3 \neq 0$) and on the other hand the leverage effect ($\rho \neq 0$).

$$\underset{\theta}{E}\left[y_t^3\right] = \underset{\theta}{E}\left[\alpha^3 y_{t-1}^3 + 3\alpha^2 y_{t-1}^2 \sigma_{t-1}\varepsilon_t + 3\alpha y_{t-1}\sigma_{t-1}^2\varepsilon_t^2 + \sigma_{t-1}^3\varepsilon_t^3\right],$$

$$(1-\alpha^3)\underset{\theta}{E}\left[y_t^3\right] = 3\alpha^2 \underset{\theta}{E}\left[y_{t-1}^2\sigma_{t-1}\varepsilon_t\right] + 3\alpha \underset{\theta}{E}\left[y_{t-1}\sigma_{t-1}^2\varepsilon_t^2\right] + \underset{\theta}{E}\left[\sigma_{t-1}^3\varepsilon_t^3\right],$$

$$\underset{\theta}{E}\left[y_t^3\right] = \frac{1}{1-\alpha^3}\left\{3\alpha^2 \underset{\theta}{E}\left[y_{t-1}^2\sigma_{t-1}\underset{\theta}{E}\left(\varepsilon_t/I_{t-1}\right)\right] + 3\alpha \underset{\theta}{E}\left[y_{t-1}\sigma_{t-1}^2\underset{\theta}{E}\left(\varepsilon_t^2/I_{t-1}\right)\right] + \right.$$
$$\left.\underset{\theta}{E}\left[\sigma_{t-1}^3\underset{\theta}{E}\left(\varepsilon_t^3/I_{t-1}\right)\right]\right\},$$

$$\underset{\theta}{E}\left[y_t^3\right] = \frac{1}{1-\alpha^3}\left\{3\alpha \underset{\theta}{E}\left[y_{t-1}\sigma_{t-1}^2\right] + \mu_3 \underset{\theta}{E}\left[\sigma_{t-1}^3\right]\right\}.$$

$$\underset{\theta}{E}\left[y_{t-1}\sigma_{t-1}^2\right] = \underset{\theta}{E}\left[(\alpha y_{t-2} + \sigma_{t-2}\varepsilon_{t-1})\left(\omega + \gamma\sigma_{t-2}^2 + \nu_{t-1}\right)\right],$$

$$\underset{\theta}{E}\left[y_{t-1}\sigma_{t-1}^2\right] = \alpha\gamma \underset{\theta}{E}\left[y_{t-2}\sigma_{t-2}^2\right] + \underset{\theta}{E}\left[\sigma_{t-2}\varepsilon_{t-1}\nu_{t-1}\right],$$

$$\underset{\theta}{E}\left[y_{t-1}\sigma_{t-1}^2\right] = \frac{\rho\eta}{1-\alpha\gamma}\underset{\theta}{E}\left[\sigma_{t-2}\right],$$

therefore, we have: $\underset{\theta}{E}\left[y_t^3\right] = \dfrac{1}{1-\alpha^3}\left\{\dfrac{3\alpha\rho\eta}{1-\alpha\gamma}\underset{\theta}{E}\left[\sigma_t\right] + \mu_3 \underset{\theta}{E}\left[\sigma_t^3\right]\right\}.$

Identification of β_1:

In order to show that β_1 is identified and in light of (6.8), we just need to prove that $\beta_{11} = (\beta_{1,\circ},\ldots,\beta_{1,q_1})'$ is identified.
We first compute the moment of interest both under the DGP and the structural mis-specified

model.

	DGP	*Structural Misspecified SV Model*
$E\left[\sigma^2\right] =$	$\dfrac{\omega^{\circ}}{1-\gamma^{\circ}},$	$\dfrac{\omega}{1-\gamma},$
$E\left[\sigma^4\right] =$	$\dfrac{\omega^{\circ 2}}{\left(1-\gamma^{\circ}\right)^2}+\dfrac{\eta^{\circ 2}}{1-\gamma^{\circ 2}},$	$\dfrac{\omega^2}{\left(1-\gamma\right)^2}+\dfrac{\eta^2}{1-\gamma^2},$
$E\left[y^2\right] =$	$\dfrac{\omega^{\circ}}{1-\gamma^{\circ}},$	$\dfrac{\omega}{1-\gamma},$
$E\left[y^4\right] =$	$\mu_4^{\circ}\left[\dfrac{\omega^{\circ 2}}{\left(1-\gamma^{\circ}\right)^2}+\dfrac{\eta^{\circ 2}}{1-\gamma^{\circ 2}}\right],$	$\mu_4\left[\dfrac{\omega^2}{\left(1-\gamma\right)^2}+\dfrac{\eta^2}{1-\gamma^2}\right],$
$E\left[y_t^2 y_{t-K}^2\right] =$	$\dfrac{\omega^{\circ 2}}{\left(1-\gamma^{\circ}\right)^2}+\gamma^{\circ K}\dfrac{\eta^{\circ 2}}{1-\gamma^{\circ 2}},$	$\dfrac{\omega^2}{\left(1-\gamma\right)^2}+\gamma^K\dfrac{\eta^2}{1-\gamma^2},$
$E\left[y_t^2 y_{t-1}\right] =$	$0,$	$\rho\eta \underset{\theta}{E}\left[\sigma_t\right],$
$E\left[y_t^3\right] =$	$\mu_3^{\circ}\underset{\circ}{E}\left[\sigma_t^3\right],$	$0\cdot$

We now introduce the $(q_1+1)\times(q_1+1)$ matrix $\Sigma_{q_1}(\theta_1)$ as follows:

$$\Sigma_{q_1}(\theta_1)=\begin{bmatrix} 1 & x & x & \dots & x & \dots & x \\ x & \mu_4(u+v) & u+\gamma v & \dots & u+\gamma^{k-1}v & \dots & u+\gamma^{q_1-1}v \\ x & u+\gamma v & \mu_4(u+v) & \dots & u+\gamma^{k-2}v & \dots & u+\gamma^{q_1-2}v \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ x & u+\gamma^{k-1}v & u+\gamma^{k-2}v & \dots & \mu_4(u+v) & \dots & u+\gamma^{q_1-k}v \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x & u+\gamma^{q_1-1}v & u+\gamma^{q_1-2}v & \dots & u+\gamma^{q_1-k}v & \dots & \mu_4(u+v) \end{bmatrix},$$

where $x=\dfrac{\omega}{1-\gamma}$, $u=x^2$, $v=\dfrac{\eta^2}{1-\gamma^2}$. Then for each $\theta_1\in\Theta_1$, where Θ_1 is the space domain of the parameters of interest θ_1, $\Sigma_{q_1}(\theta_1)$ is non singular for $q_1\geq 2$. (The proof is obtained by induction reasoning on the size of $\Sigma_{q_1}(\theta_1)$, see Dridi (1999) for more details.) $a_{q_1}(\theta_1)$ by:

$$a_{q_1}(\theta_1)=\begin{bmatrix}\dfrac{\omega}{1-\gamma}\\[2ex] \dfrac{\omega^2}{(1-\gamma)^2}+\gamma\dfrac{\eta^2}{1-\gamma^2}\\ \vdots\\ \dfrac{\omega^2}{(1-\gamma)^2}+\gamma^k\dfrac{\eta^2}{1-\gamma^2}\\ \vdots\\ \dfrac{\omega^2}{(1-\gamma)^2}+\gamma^{q_1}\dfrac{\eta^2}{1-\gamma^2}\end{bmatrix}.$$

Then the first q_1+1 components $\widetilde{\beta}_{11}(\theta_1)$ of the binding function $\widetilde{\beta}(\theta)$ are given under assumptions $(A18)-(A19)$ and $(B2)$ by:

$$\Sigma_{q_1}(\theta_1)\,\widetilde{\beta}_{11}(\theta)=a_{q_1}(\theta_1),$$

$$\iff\ \forall\theta_1\in\Theta_1,\quad \widetilde{\beta}_{11}(\theta)=\widetilde{\beta}_{11}(\theta_1)=\Sigma_{q_1}^{-1}(\theta_1)\,a_{q_1}(\theta_1)\cdot$$

We, thus, have $\widetilde{\beta}_1(\theta)=\left(\widetilde{\beta}'_{11}(\theta_1),\widetilde{\beta}'_{12}(\theta)\right)'$ with:

$$\widetilde{\beta}_{11}(\theta_1)=\Sigma_{q_1}^{-1}(\theta_1)\,a_{q_1}(\theta_1),$$

$$\widetilde{\beta}_{12}(\theta)=\begin{bmatrix}\dfrac{\omega}{1-\gamma}\\ \dfrac{\omega^2}{(1-\gamma)^2}+\gamma\dfrac{\eta^2}{1-\gamma^2}\\ \dfrac{\omega^2}{(1-\gamma)^2}+\gamma^2\dfrac{\eta^2}{1-\gamma^2}\\ \rho\eta\underset{\theta}{E}\,\sigma_t\\ \mu_4\left(\dfrac{\omega^2}{(1-\gamma)^2}+\dfrac{\eta^2}{1-\gamma^2}\right)\end{bmatrix}.$$

Using exactly the same types of argument, we also have that the first q_1+1 components β°_{11} of the instrumental pseudo-true value β° are given under assumptions $(A18)-(A19)$ by:

$$\Sigma_{q_1}(\theta^\circ_1)\,\beta^\circ_{11}=a_{q_1}(\theta^\circ_1),$$

$$\iff\quad \beta^\circ_{11}=\widetilde{\beta}_{11}(\theta^\circ_1)=\Sigma_{q_1}^{-1}(\theta^\circ_1)\,a_{q_1}(\theta^\circ_1)\cdot$$

We, thus, have $\beta_1^\circ = \left(\beta_{11}^{\circ\prime}, \beta_{12}^{\circ\prime}\right)'$ with:

$$\beta_{11}^\circ = \Sigma_{q_1}^{-1}\left(\theta_1^\circ\right) a_{q_1}\left(\theta_1^\circ\right),$$

$$\beta_{12}^\circ = \begin{bmatrix} \dfrac{\omega^\circ}{1-\gamma^\circ} \\ \dfrac{\omega^{\circ 2}}{\left(1-\gamma^\circ\right)^2} + \gamma^\circ \dfrac{\eta^{\circ 2}}{1-\gamma^{\circ 2}} \\ \dfrac{\omega^{\circ 2}}{\left(1-\gamma^\circ\right)^2} + \gamma^{\circ 2} \dfrac{\eta^{\circ 2}}{1-\gamma^{\circ 2}} \\ 0 \\ \mu_4^\circ \left(\dfrac{\omega^{\circ 2}}{\left(1-\gamma^\circ\right)^2} + \dfrac{\eta^{\circ 2}}{1-\gamma^{\circ 2}} \right) \end{bmatrix}.$$

This proves both identification of the instrumental pseudo true value β_1° and of the binding function $\widetilde{\beta}_1(\theta)$. The differentiability of the binding functions is obvious.

Proof of lemma 6.1:

Under assumptions $(A19)$ and $(B2)$, we have that the functions:

$$\begin{array}{lcl} \mathcal{B}_1 & \rightarrow & I\!R^+ \\ \beta_1 & \rightarrow & \left\| \underset{\circ}{E}\, g_1\left(y_t, \ldots, y_{t-q_1}, \beta_1\right) \right\|_2^2, \\ \beta_1 & \rightarrow & \left\| \underset{\theta}{E}\, g_1\left(\widetilde{y}_t^s(\theta, z_\circ^s), \ldots, \widetilde{y}_{t-q_1}^s(\theta, z_\circ^s), \beta_1\right) \right\|_2^2, \end{array}$$

are non stochastic differentiable functions not depending on the initial conditions $z_\circ^s$. The uniqueness of each minimum with respect to β_1 follows from the fact that for $\beta_1^\circ = \left(\widetilde{\beta}_{11}'\left(\theta_1^\circ\right), \beta_{12}^{\circ\prime}\right)'$ and $\widetilde{\beta}_1(\theta) = \left(\widetilde{\beta}_{11}'\left(\theta_1\right), \widetilde{\beta}_{12}'(\theta)\right)'$, each objective function has a zero value (thus is minimal) and that the previous values β_1° and $\widetilde{\beta}_1(\theta)$ are uniquely identified (see previous proof).

Partial Indirect Identification:

The moment conditions defining $\widetilde{\beta}_1(\theta)$ are:

$$\widetilde{\beta}_{1,\circ}\left(\theta_1\right) + x \sum_{i=1}^{q_1} \widetilde{\beta}_{1,i}\left(\theta_1\right) = x, \qquad \widetilde{\beta}_{1,\circ}\left(\theta_1\right) + \sum_{i=1}^{q_1} \left(x^2 + \gamma^{|i-j|} v\right) \widetilde{\beta}_{1,i}\left(\theta_1\right) = x^2 + \gamma^j v, \quad j = 1, \ldots, q_1,$$

$$\widetilde{\beta}_{1,q_1+1}\left(\theta_1\right) = x, \qquad \widetilde{\beta}_{1,q_1+2}\left(\theta_1\right) = x^2 + \gamma v,$$

$$\widetilde{\beta}_{1,q_1+3}\left(\theta_1\right) = x^2 + \gamma^2 v, \qquad \widetilde{\beta}_{1,q_1+4}(\theta) = \rho\eta \underset{\theta}{E}\, \sigma_t,$$

$$\widetilde{\beta}_{1,q_1+5}\left(\theta_1\right) = \mu_4\left(x^2 + v\right),$$

where $x = \dfrac{\omega}{1-\gamma}$ and $v = \dfrac{\eta^2}{1-\gamma^2}$.

Let $\theta_1 \in \Theta_1$ and $\overline{\theta}_2 \in \Theta_2$, we are now able to show that $\widetilde{\beta}_1\left(\theta_1, \overline{\theta}_2\right) = \widetilde{\beta}_1\left(\theta_1^\circ, \overline{\theta}_2\right) \Longrightarrow \theta_1 = \theta_1^\circ$.

$$\widetilde{\beta}_1\left(\theta_1, \overline{\theta}_2\right) = \widetilde{\beta}_1\left(\theta_1^\circ, \overline{\theta}_2\right),$$

$$\Longrightarrow \begin{cases} \widetilde{\beta}_{1,q_1+1}(\theta_1) = \widetilde{\beta}_{1,q_1+1}(\theta_1^\circ), \\ \widetilde{\beta}_{1,q_1+2}(\theta_1) = \widetilde{\beta}_{1,q_1+2}(\theta_1^\circ), \\ \widetilde{\beta}_{1,q_1+3}(\theta_1) = \widetilde{\beta}_{1,q_1+3}(\theta_1^\circ), \\ \widetilde{\beta}_{1,q_1+4}\left(\theta_1, \overline{\theta}_2\right) = \widetilde{\beta}_{1,q_1+4}\left(\theta_1^\circ, \overline{\theta}_2\right), \\ \widetilde{\beta}_{1,q_1+5}(\theta_1) = \widetilde{\beta}_{1,q_1+5}(\theta_1^\circ). \end{cases}$$

The previous system can be written as follows:

$$\begin{array}{l} x = x^\circ, \\ x^2 + \gamma v = x^{\circ^2} + \gamma^\circ v^\circ, \\ x^2 + \gamma^2 v = x^{\circ^2} + \gamma^{\circ^2} v^\circ, \\ \rho\eta\Psi\left(\theta_1, \overline{\theta}_2\right) = 0, \\ \mu_4\left(x^2 + v\right) = \mu_4^\circ\left(x^{\circ^2} + v^\circ\right), \end{array}$$

where $\Psi\left(\theta_1, \overline{\theta}_2\right) = \underset{\theta_1, \overline{\theta}_2}{E} \sigma_t > 0$. Since $\gamma, \gamma^\circ, v, v^\circ$ are strictly positive numbers, this implies that:

$$\begin{array}{l} x = x^\circ, \\ \gamma v = \gamma^\circ v^\circ, \\ \gamma^2 v = \gamma^{\circ^2} v^\circ, \\ \rho = 0, \\ \mu_4\left(x^2 + v\right) = \mu_4^\circ\left(x^{\circ^2} + v^\circ\right), \end{array}$$

therefore we obtain:

$$\begin{array}{l} \omega = \omega^\circ, \\ \gamma = \gamma^\circ, \\ \eta^2 = \eta^{\circ^2}, \\ \rho = \rho^\circ = 0, \\ \mu_4 = \mu_4^\circ. \end{array}$$

Consistency of the SII estimator:

As previously seen, we know that $\beta_1^\circ = \left(\beta_{11}^{\circ\prime}, \beta_{12}^{\circ\prime}\right)'$ with:

$$\beta_{11}^\circ = \Sigma_{q_1}^{-1}\left(\theta_1^\circ\right) a_{q_1}\left(\theta_1^\circ\right) = \widetilde{\beta}_{11}^\circ\left(\theta_1^\circ\right),$$

$$\beta_{12}^\circ = \begin{bmatrix} \dfrac{\omega^\circ}{1-\gamma^\circ} \\ \\ \dfrac{\omega^{\circ 2}}{\left(1-\gamma^\circ\right)^2} + \gamma^\circ \dfrac{\eta^{\circ 2}}{1-\gamma^{\circ 2}} \\ \\ \dfrac{\omega^{\circ 2}}{\left(1-\gamma^\circ\right)^2} + \gamma^{\circ 2} \dfrac{\eta^{\circ 2}}{1-\gamma^{\circ 2}} \\ \\ 0 \\ \\ \mu_4^\circ \left(\dfrac{\omega^{\circ 2}}{\left(1-\gamma^\circ\right)^2} + \dfrac{\eta^{\circ 2}}{1-\gamma^{\circ 2}} \right) \end{bmatrix} = \widetilde{\beta}_{12}\left(\theta_1^\circ, \overline{\theta}_2\right),$$

for all $\overline{\theta}_2 \in \Theta_2$. Therefore $\forall \overline{\theta}_2 \in \Theta_2,\ \beta_1^\circ = \widetilde{\beta}_{12}\left(\theta_1^\circ, \overline{\theta}_2\right)$ which implies both the partial encompassing property and the desired consistency property.