# A Simple Example of An Indirect Estimator With Discontinuous Limit Theory in the MA(1) Model

Stelios Arvanitis
Athens University of Economics and Business

January 21, 2013

**Abstract**

Indirect estimators usually emerge from two-step optimization procedures. Each step in such a procedure may induce complexities in the asymptotic theory of the estimator. In this note we provide with a simple example in which the one defined by the inversion of the binding function has a "discontinuous" limit theory even in cases where the auxiliary estimator does not. This example lives in the framework of estimation of the $\mathrm{MA}(1)$ parameter. The "discontinuities" involve the dependence of the rate of convergence on the parameter, the non continuity of the limit distribution w.r.t. the parameter and the estimator's *non regularity*.

KEYWORDS: Indirect estimator, binding function, indirect identification, $\mathrm{MA}(1)$ process, multiplicative structure, martingale difference $\mathrm{CLT}$, $\mathrm{CLT}$ to stochastic integrals, sequence of local alternatives, rate of convergence dependent on the parameter, discontinuous weak limits, non regularity.

## 1 Introduction

Indirect estimators usually emerge from two-step optimization procedures. They were formally introduced by Gourieroux, Monfort and Renault [4]. They are defined as (potentially measurable selections of approximate) minimizers of criteria (inversion criterion) that are functions of an auxiliary estimator (denoted by $\beta_n$), itself derived as an extremum estimator. The latter minimizes a criterion function (auxiliary criterion), that reflects (part of) the structure of a *possibly* misspecified auxiliary model. The inversion criterion, depends on a function connecting the underlying statistical models and termed as *the binding function*. Minimization of the inversion criterion, which usually has the form of a stochastic norm, essentially inverts the binding function or some appropriate *approximation*.

Each step of any such procedure may induce complexities in the asymptotic theory of the indirect estimator. In this note we provide with a simple example in which the indirect estimator defined by the inversion of the binding function has a rate of convergence that depends on the parameter and a "discontinuous" limit theory due to the properties of the binding function, even in cases where the auxiliary estimator does not. This example lives in the framework of estimation of the $\mathrm{MA}(1)$ parameter when a set of $\mathrm{AR}(1)$ processes is considered as the auxiliary and $\beta_n$ is simply the $\mathrm{OLSE}$. We derive the relevant theory by considering limits w.r.t. sequences of local alternatives to the parameter of interest. In this respect we also manage to illustrate the dependence of limits on the choice of these sequences. The "discontinuities" of the limit theory involve the dependence of the convergence rate on the parameter, the non continuity of the limit distribution w.r.t. the parameter,[1] and the *non regularity* of the indirect estimator at hand.[2]

In the following section we define the model and derive some initial weak limits for useful sequences of random elements. We assume multiplicativity for the structuring sequence of the $\mathrm{MA}(1)$ process. This enables a plethora of specifications. By considering two *illustrative* cases we manage to obtain limits to useful random elements that either follow the Normal distribution or are vectors of stochastic integrals. In section three we define the auxiliary estimator and derive its $\sqrt{n}$ convergence to the binding function in each of these cases. In both we obtain continuous weak limits to Normal distributions or to functions of the aforementioned integrals and regularity.

The properties of the binding function imply that indirect identification is possible *only if* the parameter space is a subset of any one of the two elements of a particular covering of the real numbers. The intersection of these two is the set $\{-1, 1\}$ which constitutes the boundary of *non-invertibility* and additionally in this framework, the boundary of *indirect identification* for the particular model. In section four, we define the indirect estimator by considering as parameter space any of the previous sets and derive its relevant weak limit[3] w.r.t. local alternatives for any element of the parameter space. These reveal the aforementioned

[1] This concept of continuity is considered w.r.t. the weak topology on the relevant set of probability measures and the topology of the underlying parameter space (which in our case is Euclidean).

[2] For a definition of regularity (restricted in cases where the rate of convergence is $\sqrt{n}$) see van der Vaart [11], e.g. page 115.

[3] By abuse of terminology we refer to the weak limit of a sequence of random elements instead to that of the sequence of the relevant distributions.

discontinuities when the parameter of interest lies on (or appropriately converges to an element of) the boundary of indirect identification.

## 2 The Model and Some Initial Limit Theory

Let $(\Omega, \mathcal{F}, P)$ denote a complete probability space and $(\mathcal{F}_t)_{t\in\mathbb{Z}}$ a filtration of subalgebras. The symbol $\rightsquigarrow$ denotes convergence in distribution under $P$. Consider $(z_t)_{t\in\mathbb{Z}}$ a sequence of $\mathrm{iid}$ real valued random variables defined on $\Omega$ such that $\mathbf{E}z_0^2 = 1$. Let also $(v_t)_{t\in\mathbb{Z}}$ denote a sequence of random variables *adopted* to $(\mathcal{F}_t)_{t\in\mathbb{Z}}$. For an arbitrary real number $\theta$ define the $\mathrm{MA}\,(1)$ process $((y_t)_{t\in\mathbb{Z}})$ by

$$\begin{aligned} y_t &= \varepsilon_t + \theta\varepsilon_{t-1}, \\ \varepsilon_t &= z_t v_t,\ t \in \mathbb{Z}. \end{aligned} \tag{1}$$

The multiplicative structure in the second equation of (1) encompasses a variety of structuring sequences $((z_t)_{t\in\mathbb{Z}}, (v_t)_{t\in\mathbb{Z}})$ for the $\mathrm{MA}\,(1)$ process. Each of the following mutually exclusive assumptions describes potential properties for these. They are solely used here as merely *illustrative* and obviously *non exhaustive* cases. *In both we will henceforth assume that* $\mathbf{E}z_0 = 0$ *and that* $z_t$ *is independent of* $\mathcal{F}_t$ *and* $\mathcal{F}_t = \sigma(v_t, v_{t-1}, \ldots)$ *for any* $t \in \mathbb{Z}$. The first assumption restricts the properties of $(v_t)_{t\in\mathbb{Z}}$ and requires the existence of certain joint moments, so that the $\mathrm{CLT}$ for stationary, ergodic, finite variance martingale difference sequences is applicable.

**Assumption A.1** $(v_t)_{t\in\mathbb{Z}}$ *is stationary and ergodic. Moreover* $\sigma^2 \doteqdot \mathbf{E}\left(v_0^2\right) < +\infty$, $\mathbf{E}\left(z_{-1}^2 v_{-1}^2 v_0^2\right) < +\infty$, $\mathbf{E}\left(z_{-2}^2 v_{-2}^2 v_0^2\right) < +\infty$ *and* $\mathbf{E}\left(\left|z_{-1}z_{-2}v_{-1}v_{-2}\right| v_0^2\right) < +\infty$.

This enables inter alia a plethora of conditionally heteroskedastic specifications for the white noise process $(\varepsilon_t)_{t\in\mathbb{Z}}$. For example $v_t^2$ could be specified as the conditional variance of any of the $\mathrm{GARCH}$-type or stochastic volatility processes for which the stationarity, ergodicity and the moment conditions hold.[4] The second assumption allows for the use of $\mathrm{CLT}$ to stochastic integrals via the characterization of the $(v_t)_{t\in\mathbb{Z}}$ sequence as a random walk "killed" before $t = 1$. In this respect we allow the conditional (adapted to $(\mathcal{F}_t)_{t\in\mathbb{Z}}$) volatility to be a non stationary process.

[4]For a very simple (and restricted) example let $v_t^2$ satisfy the $\mathrm{GARCH}\,(1,1)$ recurrence equation, $\mathbf{E}z_0^4 < +\infty$ and $\mathbf{E}\left(az_0^2 + \beta\right)^2 < 1$ where $a$ denotes the $\mathrm{ARCH}$ and $\beta$ the $\mathrm{GARCH}$ parameter respectively.

**Assumption A.2** *$v_t = 0$ for all $t \leq 0$. Moreover $v_t = v_{t-1} + u_t$ for all $t > 0$ where the sequence $(u_t)_{t\in\mathbb{Z}}$ has zero means, and for $r > 2$ is $L^r$ uniformly (w.r.t. $t$) bounded and $L^2-\mathrm{NED}$ of size $-\frac{1}{2}$ on an $a$-mixing process of size $-\frac{r}{r-2}$. Moreover for $U_t = (u_t, u_{t-1}, u_{t-2})$, $\frac{1}{n}\mathbf{E}\left(\sum_{t=1}^{n} U_t' U_t\right) \to \Omega$ which is positive definite. Finally, $z_t$ is independent from $v_s$ for any $t, s \geq 0$ and moreover $\mathbf{E}\left(|z_0|^p\right) < +\infty$ for some $p > 2$.*

Given these the following lemma provides with well known results concerning the weak limits of random elements that affect the asymptotic theory of the estimators to be examined in the following sections.

**Lemma 2.1** **a**. *Under assumption A.1*

$$\frac{1}{n}\sum\nolimits_{t=1}^{n} \varepsilon_t^2 \rightsquigarrow \sigma^2, \text{ and}$$

$$\frac{1}{\sqrt{n}}\sum\nolimits_{t=1}^{n} \begin{pmatrix} \varepsilon_t \varepsilon_{t-1} \\ \varepsilon_t \varepsilon_{t-2} \end{pmatrix} \rightsquigarrow N\left(0_{2\times 1}, V\right)$$

*for the symmetric* $V = \begin{pmatrix} \mathbf{E}\left(z_{-1}^2 v_{-1}^2 v_0^2\right) & \mathbf{E}\left(z_{-1} z_{-2} v_{-1} v_{-2} v_0^2\right) \\ \cdot & \mathbf{E}\left(z_{-2}^2 v_{-2}^2 v_0^2\right) \end{pmatrix}$. **b**. *Under assumption A.2*

$$A_n^{-1}\sum\nolimits_{t=1}^{n} \begin{pmatrix} \varepsilon_t \varepsilon_{t-1} \\ \varepsilon_t \varepsilon_{t-2} \\ \varepsilon_t^2 \end{pmatrix} \rightsquigarrow \begin{pmatrix} \int_0^1 B_1(s) B_2(s)\, dB_2^*(s) \\ \int_0^1 B_1(s) B_3(s)\, dB_3^*(s) \\ \int_0^1 B_1^2(s)\, ds \end{pmatrix}$$

*where $A_n = \mathrm{diag}\left(n^{3/2}, n^{3/2}, n^2\right)$ and $\mathbf{B} = (B_1, B_2, B_3)$, $\mathbf{B}^* = (B_1^*, B_2^*, B_3^*)$ denote mutually independent vector Brownian motions defined on $[0,1]$ with covariance matrices $\Omega s$ and $\mathrm{id}_{3\times 3}\, s$ respectively for any $s \in [0,1]$.*

**Proof.** **a**. Under assumption A.1 $(\varepsilon_t)_{t\in\mathbb{Z}}$, $(\varepsilon_t \varepsilon_{t-1})_{t\in\mathbb{Z}}$, $(\varepsilon_t \varepsilon_{t-2})_{t\in\mathbb{Z}}$ are stationary ergodic (see Proposition 2.1.1 of [10]). The first result follows from Birkhoff's LLN since $\mathbf{E}\varepsilon_0^2 = \sigma^2$. Furthermore for any non zero $(\lambda_1, \lambda_2) \in \mathbb{R}^2$, $(\lambda_1 \varepsilon_t \varepsilon_{t-1} + \lambda_2 \varepsilon_t \varepsilon_{t-2})_{t\in\mathbb{Z}}$ is a stationary ergodic martingale difference sequence with variance $\lambda_1^2 \mathbf{E}\left(z_{-1}^2 v_{-1}^2 v_0^2\right) + \lambda_2^2 \mathbf{E}\left(z_{-2}^2 v_{-2}^2 v_0^2\right) + 2\lambda_1\lambda_2 \mathbf{E}\left(z_{-1} z_{-2} v_{-1} v_{-2} v_0^2\right)$. Hence the homonymous CLT is applicable (see for example the more general Theorem 24.3 of Davidson [3]). Then the result follows by the Cramer-Wold device (see Theorem 25.5 of Davidson [3]). **b**. First notice that due to the assumed behavior of $(z_t)_{t\in\mathbb{Z}}$, assumption A.2 and Corollary 29.19 of Davidson [3] we have for $x_n(s) \doteqdot n^{-1/2}\sum_{t=1}^{[ns]} (z_t, z_t z_{t-1}, z_t z_{t-2}, U_t)'$, $s \in [0,1]$, that

$x_n^* \rightsquigarrow (\mathbf{B}^*, \mathbf{B})$ in $\mathbb{D}([0,1], \mathbb{R}^6)$ where $[\cdot]$ denotes the integer part function.[5] Let $\langle\cdot\rangle$ denote quadratic variation and notice that $A_n^{-1}\sum_{t=1}^{[ns]}(\varepsilon_t\varepsilon_{t-1}, \varepsilon_t\varepsilon_{t-2}, \varepsilon_t^2)'$ can be expressed as

$$\int_0^s \left(N_n(s) N_{1,n}(s) dN_{1,n}^*(s), N_n(s) N_{2,n}(s) dN_{2,n}^*(s), (N_n(s))^2 d\langle N_n^*\rangle(s)\right)'$$

for $N_n(s) = \sum_{t=1}^{[ns]} \frac{u_t}{\sqrt{n}}$, $N_{i,n}(s) = \sum_{t=1}^{[ns]} \frac{u_{t-i}}{\sqrt{n}}$, $N_n^*(s) = \frac{z_0}{\sqrt{n}} + \sum_{t=0}^{[ns]} \frac{z_t}{\sqrt{n}}$, $N_{i,n}^*(s) = \frac{z_0 z_{-i}}{\sqrt{n}} + \sum_{t=1}^{[ns]} \frac{z_t z_{t-i}}{\sqrt{n}}$ for $i = 1, 2$ which define semimartingales. Due to Theorem 1 of Protter [7] and the integration by parts formula for the Ito's integral (see the proof of Corollary (Integration by Parts) of Protter [7]) the processes $N_n N_{1,n}$, $N_n N_{2,n}$, $N_n^2$, $\langle N_n^*\rangle$ and any linear combination of those and the previous are also semimartingales. Consider $\mu, \lambda$ non zero elements of $\mathbb{R}^3$. The previous limit result and the continuous mapping theorem imply that

$$\begin{aligned} & \left(\lambda_1 N_n N_{1,n} + \lambda_2 N_n N_{2,n} + \lambda_3 (N_n)^2, \mu_1 N_{1,n}^* + \mu_2 N_{2,n}^* + \mu_3 \langle N_n^*\rangle\right) \\ \rightsquigarrow \; & \left(\lambda_1 B_1 B_2 + \lambda_2 B_1 B_3 + \lambda_3 B_1^2, \mu_1 B_2^* + \mu_2 B_3^* + \mu_3 t\right) \end{aligned}$$

in $\mathbb{D}([0,1], \mathbb{R}^2)$. Furthermore for the process $\mu_1 N_{1,n}^* + \mu_2 N_{2,n}^* + \mu_3 \langle N_n^*\rangle$ we have that

$$\begin{aligned} & \mathbf{E}\sup_{r\leq s}\left|\mu_1 N_{1,n}^*(r) + \mu_2 N_{2,n}^*(r) + \mu_3 \langle N_n^*\rangle(r)\right| \\ \leq \; & |\mu_1|\,\mathbf{E}\sup_{r\leq s}\left|N_{1,n}^*(r)\right| + |\mu_2|\,\mathbf{E}\sup_{r\leq s}\left|N_{2,n}^*(r)\right| + |\mu_3|\,\mathbf{E}\sup_{r\leq s}\langle N_n^*\rangle(r). \end{aligned}$$

By noticing that due to the definition of $(z_t)_{t\in\mathbb{Z}}$, and assumption A.2 the processes $N_{1,n}^*$, $N_{2,n}^*$ and $N_n^*$ are square integrable martingales and via remark 3.4 of Ibragimov and Phillips [6] and Jensen's inequality there exist constants independent of $n$ that bound from above each of the two first terms in the sum of the right hand side of the previous display. For the last one we simply have

$$\mathbf{E}\sup_{r\leq s}\langle N_n^*\rangle(r) = \mathbf{E}\sup_{r\leq s} n^{-1}\sum_{t=0}^{[nr]} z_t^2 \leq \mathbf{E}\sup_{r\leq s} n^{-1}\sum_{t=0}^{n} z_t^2 \leq 2.$$

This implies that

$$\sup_n \mathbf{E}\left|\sup_{r\leq s}\Delta\left(\mu_1 N_{1,n}^*(r) + \mu_2 N_{2,n}^*(r) + \mu_3 \langle N_n^*\rangle(r)\right)\right| < +\infty$$

which in turn means that the uniform tightness condition in Proposition 3.2 (a) of Jakubowski, Memin and Pages [5] follows for the integrator semimartingale

[5] Remember that $\mathbb{D}([0,1], \mathbb{R}^m)$ denotes the set of $\mathrm{cadlag}$ functions $[0,1] \to \mathbb{R}^m$ equiped with the Skorokhod topology. When $m = 1$ it is abbreviated as $\mathbb{D}([0,1])$.

sequence $\left(\mu_1 N^*_{1,n} + \mu_2 N^*_{2,n} + \mu_3 \langle N^*_n \rangle\right)_{n\in\mathbb{N}}$. Then Theorem 2.6 of Jakubowski, Memin and Pages [5] implies that

$$\left(H_n, S_n, \int_0^{\cdot} H_n dS_n\right) \rightsquigarrow \left(H, S, \int_0^{\cdot} H dS\right)$$

in $\mathbb{D}\left([0,1], \mathbb{R}^3\right)$ where $H_n = \lambda' \mathbf{N}_n$ with $\mathbf{N}_n = \left(N_n N_{1,n}, N_n N_{2,n}, (N_n)^2\right)'$, $S_n = \mu' \mathbf{S}_n$ with $\mathbf{S}_n = \left(N^*_{1,n}, N^*_{2,n}, \langle N^*_n \rangle\right)$ and $H = \lambda' \mathbf{N}$ with $\mathbf{N} = \left(B_1 B_2, B_1 B_3, B_1^2\right)$, $S = \mu' \mathbf{S}$ with $\mathbf{S} = \left(B^*_2, B^*_3, \langle B^*_1 \rangle\right)$. Then due to the fact that

$$\int_0^{\cdot} H_n dS_n = \int_0^{\cdot} (\lambda \otimes \mu)' \operatorname{vec}\left(\mathbf{N}_n d\mathbf{S}'_n\right)$$

the Cramer-Wold theorem implies that $\int_0^{\cdot} \mathbf{N}_n d\mathbf{S}'_n \rightsquigarrow \int_0^{\cdot} \mathbf{N} d\mathbf{S}'$ and the result follows by the continuous mapping theorem. ■

# 3 Limit Theory for the Auxiliary Estimator under Local Alternatives

Let $(\eta_n)_{n\in\mathbb{N}}$ denote a real sequence for which $\eta_n \to +\infty$ and for any $\theta \in \mathbb{R}$ let $\theta_{\eta,c} = \theta + \frac{c}{\eta_n}$ for some $c \in \mathbb{R}$. Let $\underset{\theta_{\eta,c}}{\rightsquigarrow}$ denote weak convergence under the sequence of measures induced by the stochastic process defined when $\theta$ is replaced by $\theta_{\eta,c}$ in (1). Consider the first order sample autocorrelation statistic[6]

$$\beta_n = \frac{\sum_{t=1}^n y_t y_{t-1}}{\sum_{t=1}^n y_{t-1}^2}$$

Notice that in the scope of assumption A.1 or A.2 due to lemma 2.1 $\beta_n$ is well defined with $P$ probability that converges to $1$. $\beta_n$ represents the $\mathrm{OLS}$ estimator in the scope of the statistical model comprised of the $\mathrm{AR}(1)$ processes recursively built on the $(\varepsilon_t)_{t\in\mathbb{Z}}$ sequence. Hence it corresponds to the *auxiliary estimator* of any indirect inference procedure for the estimation of $\theta$, based on the $\mathrm{AR}(1)$ auxiliary model *and* the least squares criterion. In this paragraph we are interested in the asymptotic behavior of $\beta_n$ under sequences of local alternatives given our assumption framework. We readily obtain the following result establishing the *continuous* limit theory for the auxiliary estimator. This also establishes that

[6] Since $\beta_n$ is a function of $\theta$ we use the notation $\beta_n(\theta) \to \cdot$ or $\beta_n \underset{\theta}{\to} \cdot$ interchangeably where $\to$ denotes the underlying limiting mode.

under either assumption A.1 or assumption A.2 $\beta_n \underset{\theta_{\eta,c}}{\rightsquigarrow} b(\theta) = \frac{\theta}{1+\theta^2}$.[7] The latter is the binding function in this framework. Given our structure it corresponds to a parametric representation of an underlying (multi-) function defined on the set of the probability measures (*statistical model*) described by the $\mathrm{MA}(1)$ processes and to the analogous set (*auxiliary model*) described by the $\mathrm{AR}(1)$ processes.

**Lemma 3.1** *In both the considered cases*

$$\sqrt{n}\left(\beta_n - \frac{\theta_{\eta,c}}{1+\theta_{\eta,c}^2}\right) \underset{\theta_{\eta,c}}{\rightsquigarrow} z_\theta$$

*where:* **a**. *under assumption A.1*

$$z_\theta \backsim N(0, v_\theta)$$

*and* $v_\theta = \frac{\left(1-\theta^2+\theta^4\right)^2 \mathbf{E}\left(z_{-1}^2 v_{-1}^2 v_0^2\right) + \theta^2\left(1+\theta^2\right)^2 \mathbf{E}\left(z_{-2}^2 v_{-2}^2 v_0^2\right) + 2\theta\left(1-\theta^2+\theta^4\right)\left(1+\theta^2\right)\mathbf{E}\left(z_{-1} z_{-2} v_{-1} v_{-2} v_0^2\right)}{\left(1+\theta^2\right)^2 \sigma^2}$,

*and* **b**. *under assumption A.2*

$$z_\theta = \frac{\left(1-\theta^2+\theta^4\right)\int_0^1 B_1(s)\, B_2(s)\, dB_2^*(s) + \theta\left(1+\theta^2\right)\int_0^1 B_1(s)\, B_3(s)\, dB_3^*(s)}{\left(1+\theta^2\right)\int_0^1 B_1^2(s)\, ds}.$$

**Proof.** First notice that trivial algebra shows that

$$\beta_n - \frac{\theta_{\eta,c}}{1+\theta_{\eta,c}^2} = \frac{\left(1-\theta_{\eta,c}^2+\theta_{\eta,c}^4\right)\sum_{t=1}^n \varepsilon_t\varepsilon_{t-1} + \left(\theta_{\eta,c}+\theta_{\eta,c}^3\right)\sum_{t=1}^n \varepsilon_t\varepsilon_{t-2} + c_n}{\left(1+\theta_{\eta,c}^2\right)\sum_{t=1}^n \varepsilon_t^2 + 2\theta_{\eta,c}\sum_{t=1}^n \varepsilon_t\varepsilon_{t-1} + c_n^*}$$

where

$$c_n = \left(\theta_{\eta,c} - 2\theta_{\eta,c}^2 + \theta_{\eta,c}^3\right)\left(\varepsilon_0\varepsilon_{-1} - \varepsilon_n\varepsilon_{n-1}\right) - \theta_{\eta,c}^3\left(\varepsilon_{-1} - \varepsilon_{n-1}\right)$$

and

$$c_n^* = \theta_{\eta,c}^2\left(\varepsilon_{-1} + \varepsilon_0 - \varepsilon_{n-1} - \varepsilon_n\right) + \left(\varepsilon_0 - \varepsilon_n\right) + 2\theta_{\eta,c}\left(\varepsilon_0\varepsilon_{-1} - \varepsilon_n\varepsilon_{n-1}\right).$$

Due to lemma 2.1 and the definition of the $(z_t)_{t\in\mathbb{Z}}$ and $\theta_{\eta,c}$ it can be easily deduced that $k_n c_n$, $k_n^* c_n^* \underset{\theta_{\eta,c}}{\rightsquigarrow} 0$ for $k_n = n^{-1/2}$, $k_n^* = n^{-1}$ in case **a**. and $k_n =$

[7] Notice that this could not be the case under some deviations from our framework that could be justified by the structure of multiplicativity. For example if assumption A.2 holds yet $z_t = 1$ $P$ almost surely for all $t$ (i.e. we deviate from the restriction that $\mathbf{E}z_0 = 0$) then Corollary 29.19 of Davidson [3], the continuous mapping theorem and the proof of lemma 3.1 would imply that $\beta_n \underset{\theta_{\eta,c}}{\rightsquigarrow} b(\theta) + \frac{\left(1-\theta^2+\theta^4\right)\int_0^1 B_1(s)B_2(s)ds + \theta\left(1+\theta^2\right)\int_0^1 B_1(s)B_3(s)ds}{\left(1+\theta^2\right)\int_0^1 B_1^2(s)ds + 2\theta\int_0^1 B_1(s)B_2(s)ds}$.

$n^{-3/2}$, $k_n^* = n^{-2}$ in case **b**. The results follow then from lemma 2.1, the definition of $\theta_{\eta,c}$ and the continuous mapping theorem by noticing that $k_n^* \sum_{t=1}^{n} \varepsilon_t \varepsilon_{t-1} \rightsquigarrow 0$ in both cases. ■

Any indirect estimator based on this structure would be defined via some approximation of the binding function.[8] It can be easily seen that $b$ is injective if and only if it is restricted to (any non empty subset of) $\Theta_1 = [-1, 1]$ or to (any non empty subset of) $\Theta_2 = (-\infty, -1] \cup [1, +\infty)$. Hence indirect identification of $\theta$ is possible if and only if the underlying parameter space satisfies this restriction. In the final section we define the indirect estimator via $\beta_n$ and the binding function and derive its limit theory by keeping in mind that the points $-1$ and $1$ constitute the *boundary* of indirect identification.

# 4 Limit Theory for the Indirect Estimator under Local Alternatives

In what follows $\Theta$ denotes an arbitrary non empty connected subset of $\mathbb{R}$ under the restriction that if $\Theta \cap \Theta_1 \supset \{-1, 1\}$ then $\Theta \cap \Theta_2 \subseteq \{-1, 1\}$ and symmetrically if $\Theta \cap \Theta_2 \supset \{-1, 1\}$ then $\Theta \cap \Theta_1 \subseteq \{-1, 1\}$. This restriction obviously corresponds to the indirect identification condition implied by the properties of the binding function and enables the definition (avoiding the use of measurable selections) and the possibility of consistency for the indirect estimator that we immediately define. Given $\beta_n$ and $\Theta$ consider the so called $\mathrm{GMR1}$ indirect estimator generally defined by

$$\theta_n \in \arg\min_{\Theta} \|\beta_n - b(\theta)\|,$$

where $\|\cdot\|$ is any (possibly stochastic) norm on $\mathbb{R}$.[9] The form of $\Theta$ and the binding function implies that this optimization problem admits in any case a unique solution. For the sake of simplicity and without much loss of generality we will consider only the cases where $\Theta = \Theta_1$ or $\Theta = \Theta_2$ whence the estimator is $P$ almost surely given by

$$\theta_n = \begin{cases} \frac{1-\sqrt{1-4\beta_n^2}}{2\beta_n} \text{ if } \Theta = \Theta_1 \text{ and } \beta_n \in \left[-\frac{1}{2}, \frac{1}{2}\right] \\ \frac{1+\sqrt{1-4\beta_n^2}}{2\beta_n} \text{ if } \Theta = \Theta_2 \text{ and } \beta_n \in \left[-\frac{1}{2}, \frac{1}{2}\right] \\ 1 \text{ if } \beta_n > \frac{1}{2} \text{ and } -1 \text{ if } \beta_n < -\frac{1}{2} \end{cases}. \tag{2}$$

We are interested in the limit theory of $\theta_n$. In what follows $\theta_{\eta,c}$ will be as in the previous section. The fact that $c$ can be an arbitrary real number and $\theta$ can

[8] Essentially it would be defined as a (possibly approximate) minimizer of some (possibly stochastic) distance between the $\beta_n$ and the approximation to the binding function.

[9] Or a sequence of (possibly stochastic) norms that converge in probability to a norm on $\mathbb{R}$.

belong to the boundary of $\Theta$ enables the study of the asymptotic behavior of our estimator when the statistical model is only *asymptotically well specified*. When $\theta$ is not in the boundary of indirect identification then the limit theory of $\theta_n$ is easily established via the delta method using first order approximations. This is not possible in the cases that $\theta = 1$ or $-1$ since the differentiability of $b$ imply that its Jacobian at these points is singular and this is precisely the source of the resulting discontinuity of the limit theory of the $\mathrm{GMR1}$ estimator as exhibited in the following result. In any of the cases described, remember that $z_\theta$ denotes the random element in lemma 3.1 that either follows the normal distribution under assumption A.1 or is of the form of a rational function w.r.t. to the stochastic integrals appearing in lemma 2.1 under assumption A.2.

**Lemma 4.1** *In both the considered cases*

$$r_n\left(\theta_n - \theta_{\eta,c}\right) \underset{\theta_{\eta,c}}{\rightsquigarrow} x_\theta$$

*where* **i.** *if $\theta \in \operatorname{Int}\Theta$ and $\Theta$ is either $\Theta_1$ or $\Theta_2$ then for all $c \in \mathbb{R}$ and $(\eta_n)_{n\in\mathbb{N}}$ such that $\eta_n \to +\infty$*

$$r_n = \sqrt{n} \text{ and } x_\theta = \frac{\left(1+\theta^2\right)^2}{1+\theta^2} z_\theta,$$

**ii.** *if $\theta = -1$ and $\Theta{=}\Theta_1$ then for all $c \in \mathbb{R}$ and $\eta_n = n^{1/4}$*

$$r_n = n^{1/4} \text{ and } x_{-1} = \begin{cases} 2\sqrt{z_{-1} + \frac{1}{4}c^2} \text{ if } z_{-1} \geq -\frac{1}{4}c^2 \\ -c \text{ if } z_{-1} < -\frac{1}{4}c^2 \end{cases},$$

**iii.** *if $\theta = 1$ and $\Theta{=}\Theta_1$ then for all $c \in \mathbb{R}$ and $\eta_n = n^{1/4}$*

$$r_n = n^{1/4} \text{ and } x_1 = \begin{cases} -2\sqrt{-z_1 - \frac{1}{4}c^2} \text{ if } z_1 \leq \frac{1}{4}c^2 \\ -c \text{ if } z_1 > \frac{1}{4}c^2 \end{cases},$$

**iv.** *if $\theta = -1$ and $\Theta{=}\Theta_2$ then for all $c \in \mathbb{R}$ and $\eta_n = n^{1/4}$*

$$r_n = n^{1/4} \text{ and } x_{-1} = \begin{cases} -2\sqrt{-z_{-1} + \frac{1}{4}c^2} \text{ if } z_{-1} \leq -\frac{1}{4}c^2 \\ -c \text{ if } z_{-1} > \frac{1}{4}c^2 \end{cases},$$

**v.** *if $\theta = 1$ and $\Theta{=}\Theta_2$ then for all $c \in \mathbb{R}$ and $\eta_n = n^{1/4}$*

$$r_n = n^{1/4} \text{ and } x_1 = \begin{cases} 2\sqrt{z_1 - \frac{1}{4}c^2} \text{ if } z_1 \geq \frac{1}{4}c^2 \\ -c \text{ if } z_1 < \frac{1}{4}c^2 \end{cases},$$

*and $z_\theta$ is as in lemma 3.1*.**a** *under assumption A.1 or as in lemma 3.1*.**b** *under assumption A.2.*

**Proof.** **i.** Initially notice that under either assumption A.1 or assumption A.1 for all $\theta, c \in \mathbb{R}$ and for any admissible $(\eta_n)_{n\in\mathbb{N}}$ and due to the fact that the support of $z_\theta$ is the real line, $P\left(\left|\beta_n - \frac{\theta_{\eta,c}}{1+\theta^2_{\eta,c}}\right| > \varepsilon\right) \to 0$ for all $\varepsilon > 0$. Due to the previous we have that if $\theta \in \operatorname{Int}\Theta$ and $\Theta$ is either $\Theta_1$ or $\Theta_2$ and for all $c \in \mathbb{R}$ and $(\eta_n)_{n\in\mathbb{N}}$ due to the fact that the Jacobian of the binding function evaluated at $\theta_{\eta,c}$ is eventually non singular, $\theta_n$ satisfies $\beta_n = b(\theta_n)$ with $P$ probability converging to $1$. The result follows from the mean value theorem applied on the binding function w.r.t. $\theta_{\eta,c}$, the previous remark and the fact that the Jacobian is continuous. **ii.** Suppose that $\theta = -1$ and $\Theta = \Theta_1$ choose some $c \geq 0$ and let $\eta_n = n^{1/2}$ or $\eta_n$ such that $n^{1/2} = o(\eta_n)$. Let $=_d$ denote distributional equivalence and $\underset{a.s.}{\to}$ almost sure convergence under the relevant probability measure. The Skorokhod representation (see for example Theorem 26.25 in Davidson [3]) via the embedding of $(\Omega, \mathcal{F}, P)$ to a "larger" probability space $(\Omega, \mathcal{F}, P)$ ensures the existence of random variables $\beta^*_n$, $z^*_\theta$ defined on $\Omega^*$, such that $\beta^*_n =_d \beta_n(\theta_{\eta,c})$, $z^*_\theta =_d z_\theta$ in each of the cases of lemma 3.1 and for any $\theta$ such that the weak limit results of that lemma hold as almost sure limits w.r.t. $\beta^*_n$ and $z^*_\theta$. Define $\theta^*_n$ using (2) and $\beta^*_n$. Notice that $\theta^*_n =_d \theta_n(\theta_{\eta,c})$. Again the sequences defined by $n^{1/4}(\theta^*_n - \theta_{\eta,c})$ and $n^{1/4}(\theta_n(\theta_{\eta,c}) - \theta_{\eta,c})$ must have weak limits that are distributionaly equivalent if any one of them converges. By a first order Taylor expansion of $b(\theta_{\eta,c})$ around $-1$ using the Lagrange remainder we have that for $\widetilde{\theta}_{\eta,c}$ between $\theta_{\eta,c}$ and $-1$

$$\begin{aligned} b(\theta_{\eta,c}) &= -\frac{1}{2} - \frac{\widetilde{\theta}_{\eta,c}\left(1+\widetilde{\theta}^2_{\eta,c}\right)\left(3\widetilde{\theta}^2_{\eta,c}-1\right)}{\left(1+\widetilde{\theta}^2_{\eta,c}\right)^4}\frac{c^2}{\eta_n^2} \\ &= -\frac{1}{2} - A_n\frac{c^2}{\eta_n^2} \end{aligned}$$

Due to the form of (2) and the previous expansion we have that

$$n^{1/4}\left(\theta^*_n + 1 - \frac{c}{\eta_n}\right) = \begin{cases} K_n + A_n\frac{n^{1/4}c^2}{\eta_n^2\beta_n} - \frac{n^{1/4}c}{\eta_n} & \text{if } \beta^*_n \in \left[-\frac{1}{2}, \frac{1}{2}\right] \\ -\frac{cn^{1/4}}{\eta_n} & \text{if } \beta^*_n < -\frac{1}{2} \\ n^{1/4}\left(2 - \frac{c}{\eta_n}\right) & \text{if } \beta^*_n > \frac{1}{2} \end{cases}$$

where

$$K_n = \frac{n^{1/4}\left(\beta^*_n - b(\theta_{\eta,c})\right) - \sqrt{n^{1/2}\left(b(\theta_{\eta,c}) - (\beta^*_n)^2\right) - A_n^2\frac{n^{1/2}c^4}{\eta_n^4} - A_n\frac{n^{1/2}c^2}{\eta_n^2}}}{\beta_n}.$$

In both cases in lemma 3.1 the event $\left\{\beta^*_n > \frac{1}{2}\right\} = \left\{\sqrt{n}\left(\beta^*_n - b(\theta_{\eta,c})\right) > \sqrt{n}\left(1 + \frac{c^2}{\eta_n^2}\right)\right\}$ converges to an event of zero probability. Analogously due to the definition of

$\eta_n$ the event $\left\{\beta_n^* < -\frac{1}{2}\right\}$ converges to $\left\{z_{-1}^* < -\frac{c^2}{4}\right\}$ under either assumption A.1 or assumption A.2. In the same manner the

$$\left\{\beta_n^* \in \left[-\frac{1}{2}, \frac{1}{2}\right]\right\} = \left\{\sqrt{n}\left(\beta_n^* - b\left(\theta_{\eta,c}\right)\right) \in \left[A_n \frac{n^{1/2}c^2}{\eta_n^2}, \sqrt{n}\left(1 + \frac{c^2}{\eta_n^2}\right)\right]\right\}$$

converges to $\left\{z_{-1}^* \in \left[-\frac{c^2}{4}, +\infty\right)\right\}$. Hence due to lemma 3.1, the definition of $\eta_n$, and by noticing that by the Delta method $n^{1/2}\left(\left(\beta_n^*\right)^2 - \left(-\frac{1}{2} - \frac{c}{\eta_n}\right)^2\right) \underset{a.s.}{\rightarrow} -z_{-1}^*$ we finally obtain that

$$n^{1/4}\left(\theta_n^* - \theta_{\eta,c}\right) \underset{a.s.}{\rightarrow} x_1^* = \begin{cases} 2\sqrt{z_{-1}^* + \frac{1}{4}c^2} \text{ if } z_{-1}^* \geq -\frac{1}{4}c^2 \\ -c \text{ if } z_{-1}^* < -\frac{1}{4}c^2 \end{cases}.$$

The other cases follow analogously and are omitted to conserve on space. ■

Analogous "discontinuities" have been reported in the $\mathrm{MA}$ literature for direct estimators when $\theta_{\eta,c}$ converges to $-1$ or $1$. A characteristic example is that of the $\mathrm{MLE}$ when $v_t = \sigma > 0$ for all $t$ and $z_0$ follows a standard Normal distribution (see Shephard [9]). There the estimator is actually $n$-consistent something which is in contrast to the present cases where convergence is becoming *slower*. This is the result of the fact that when $\theta$ lies (or converges to an element of) the boundary then the properties of the score process change discontinuously. This is not the case in our framework. Here the results are attributed solely to the local properties of the binding function around the boundary points. This behavior and the corresponding discontinuity of the limit theory is also vaguely reminincent of similar behaviors studied in economic theory (see for example Roberts [8]) where appropriate sequences of economies may converge to non perfectly competitive equilibria due to the behavior of the excess demand function around "critical" points.

Notice now that the present choices of $\Theta$ correspond to the "largest" possible covering of $\mathbb{R}$ that renders possible the estimation of the $\mathrm{MA}(1)$ parameter via the use of any of each members, inside our framework. In this respect they are the natural ones to reveal the "discontinuities" of the limit theory (i.e. the dependence of the rate on $\theta$, the discontinuity of the weak limits and the non regularity) for our indirect estimator that essentially emerge on the boundary of indirect identification due to the relevant properties of the binding function. This fact would have been obscured if $\Theta$ would have been chosen for instance as an open interval. Furthermore there are choices of $\Theta$ that would imply further discontinuities that do not stem from this *natural* structure.[10] Consider for

[10]Obviously such choices could be justified by the presence of information for the $\mathrm{MA}(1)$ parameter that is *external* to the present (and more general) structure.

example the case where $\Theta = [\theta_1, \theta_2]$ and $\theta_1 < -1$ or $\theta_1 > 1$. For $\theta = \theta_1$, any $c$ and any admissible $(\eta_n)_{n\in\mathbb{N}}$ we could obtain via the use of a similar methodology to the one in Andrews [1] that

$$\sqrt{n}\left(\theta_n - \theta_{\eta,c}\right) \underset{\theta_{\eta,c}}{\rightsquigarrow} z_{\theta_1} = \begin{cases} \frac{\left(1+\theta_1^2\right)^2}{1+\theta_1^2} z_{\theta_1} \text{ if } z_{\theta_1} \geq 0 \\ 0 \text{ if } z_{\theta_1} < 0 \end{cases}$$

i.e. a limit defined as an appropriate projection of the limit appearing in lemma 4.1.**i.** This is obviously a different kind of discontinuity (see for example that the rate of convergence does not depend on this artificial boundary and that the limit distribution does not depend on $c$) that is essentially imposed by the *choice* of $\Theta$ and not the structure at hand.

Second, it is evident from the proof of the previous lemma that the *form* of the result does not depend on the results of lemma 3.1. That is if we had that $q_n\left(\beta_n - b\left(\theta_{\eta,c}\right)\right) \underset{\theta_{\eta,c}}{\rightsquigarrow} z_\theta$ by a different set of assumptions for $q_n \to +\infty$ not depending on $\theta$ and $z_\theta$ a random element not necessarily equal to the ones in lemma 3.1, then it is easy to see that the results of lemma 4.1 would have the same phrasing except for the replacement of $r_n$ with $q_n$ in case **i**. and with $\sqrt{q_n}$ in all the remaining cases. As noted in a previous section the results of lemma 3.1 correspond simply to some illustrative cases described by our initial pair of assumptions.

Third, a similar question about the limit theory of other indirect estimators given the aforementioned structure could be of potential interest. The following discussion explores two simple cases in which indirect estimators associated with asymptotic bias correction in other frameworks are identified as $\mathrm{GMR1}$ estimators in this one. In the premises of either assumption A.1 or assumption A.2 consider[11]

$$\beta_n^* = \begin{cases} \beta_n \text{ if } \beta_n \in \Theta_1 \\ -\frac{1}{2} \text{ if } \beta_n < -\frac{1}{2} \text{ or } \frac{1}{2} \text{ if } \beta_n > \frac{1}{2} \end{cases}$$

and

$$\theta_n^* = \arg\min_{\Theta} \left\| \beta_n^* - b(\theta) - \frac{1_{\beta_n \leq -1/2}}{\sqrt{n}} \mathbf{E}\left(1_{z_\theta \in [0,+\infty)} z_\theta\right) - \frac{1_{\beta_n \geq 1/2}}{\sqrt{n}} \mathbf{E}\left(1_{z_\theta \in (-\infty,0]} z_\theta\right) \right\|.$$

$\theta_n^*$ is the natural extension of the $\mathrm{GMR2}\,(a)$ estimator considered in Arvanitis and Demos [2] for $a = \frac{1}{2}$. Since $1_{\beta_n \leq -1/2} = 1$ is equivalent to $\beta_n^* = -\frac{1}{2}$ and $\mathbf{E}\left(1_{z_\theta \in [0,+\infty)} z_\theta\right) \geq 0$ and by the analogous reasoning for the $\frac{1_{\beta_n \geq 1/2}}{\sqrt{n}} \mathbf{E}\left(1_{z_\theta \in (-\infty,0]} z_\theta\right)$

[11] $\beta_n^*$ is obviously a restricted $\mathrm{OLSE}$ in the framework of the auxiliary model, corresponding to the optimization of the least squares criterion on $\left[-\frac{1}{2}, \frac{1}{2}\right]$, something that is in accordance with the present structure due to the fact that this interval equals $b(\mathbb{R}) = b(\Theta_1) = b(\Theta_2)$.

term we obtain that in for $\Theta$ equal to either $\Theta_1$ or $\Theta_2$ $\theta_n^* = \mathrm{GMR1}$, $P$ almost surely for any $n$, and thereby the previous analysis holds also for this estimator.[12]

For the second case a recursive version of the $\mathrm{GMR2}\left(\frac{1}{2}\right)$ indirect estimator can be defined as $\theta_n^{**} = \arg\min_{\Theta} \|\Lambda_n\|$ where

$$\Lambda_n = \begin{cases} \theta_n - \theta - 2\frac{1_{\beta_n \leq -1/2}}{n^{1/4}}\mathbf{E}\left(1_{z_\theta \in [0,+\infty)}\sqrt{z_\theta}\right) + 2\frac{1_{\beta_n \geq 1/2}}{n^{1/4}}\mathbf{E}\left(1_{z_\theta \in (-\infty,0]}\sqrt{-z_\theta}\right) & \text{if } \Theta = \Theta_1 \\ \theta_n - \theta + 2\frac{1_{\beta_n \leq -1/2}}{n^{1/4}}\mathbf{E}\left(1_{z_\theta \in (-\infty,0]}\sqrt{-z_\theta}\right) - 2\frac{1_{\beta_n \geq 1/2}}{n^{1/4}}\mathbf{E}\left(1_{z_\theta \in [0,+\infty)}\sqrt{z_\theta}\right) & \text{if } \Theta = \Theta_2 \end{cases}.$$

$\theta_n^{**}$ is actually derived in a three step procedure where in the second step the auxiliary statistical model concides with the $\mathrm{MA}(1)$ and the $\mathrm{GMR1}$ estimator derived in the second step is used as the auxiliary estimator for the final one. Obviously due to the results in lemma 4.1 the binding function connecting the last two steps of the procedure is the *identity*. Notice that when $\Theta = \Theta_1$ then for any $n$, $\theta_n^{**} = \mathrm{GMR1}$, $P$ almost surely whereas when $\Theta = \Theta_2$ then *eventually* for large enough $n$, $\theta_n^{**} = \mathrm{GMR1}$, $P$ almost surely. Hence in these cases lemma 4.1 also provides with the limit theory concerning $\theta_n^*$ and $\theta_n^{**}$.

# 5 Conclusions

In the context of estimation of the $\mathrm{MA}(1)$ parameter we have examined a simple example of an indirect estimator having a "discontinuous" limit theory even in cases where the auxiliary estimator does not. The "discontinuities" involve the dependence of the rates on the parameter, the non continuity of the limit distribution w.r.t. the parameter and the dependence of the limit on the choice of the sequence of local alternatives. These emerge at the boundary points of indirect identification and are entirely attributed to the local behavior of the binding function near those points. The dependence of the rates on the parameter and of the limit distribution on the sequence of local alternatives do not manifest in the limit theory when the parameter space is arbitrarily bounded outside the boundary of indirect identification. These constitute of a simple manifestation of the fact that the limit theory of indirect estimators can be quite interesting.

[12] As noted previously the definition of $\theta_n^*$ can in some cases facilitate bias reduction. Presently we are only considering it in a *formal* manner without being occupied with the issue of uniform integrability for the auxiliary estimator. The definition can be readily extended even when the mean of the asymptotic distribution of the auxiliary estimator is not well defined, by restricting the domains of integration appearing in the norm to arbitrary bounded intervals.

## References


[1] Andrews, D. W. K. (*1999*), "Estimation when a parameter is on boundary.", *Econometrica,* 67, pp. 1341-1383.

[2] Arvanitis, S. & A. Demos (2012), "A Class of Indirect Estimators: Higher Order Asymptotics and Approximate Bias Correction.", mimeo, AUEB.

[3] Davidson J. (1995), "*Stochastic Limit Theory.*", Advanced Texts in Econometrics, Oxford University Press.

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

[5] Jakubowski, A., Memin, J. and Pages, G. (1989). "Convergence en loi des suites d'integrales stochastiques sur l'espace $\mathbb{D}^1$ de Skorokhod.", *Probability Theory and Related Fields* 81, pp. 111-137.

[6] Ibragimov R. and P.C.B. Phillips (2008), "Regression Asymptotics Under Martingale Convergence Methods.", *Econometric Theory* 24(4), pp. 888-947.

[7] Protter Ph. (1990), "*Stochastic Integration and Differential Equations. A New Approach.*", Springer-Verlag.

[8] Roberts, K. (1980), "The Limit Points of Monopolistic Competition.", *Journal of Economic Theory,* Vol. 22, pp. 256-278."

[9] Shephard Neil (1993), "Distribution of the ML Estimator of an MA(1) and a Local Level Model.", *Econometric Theory,* Vol. 9, pp. 377-401.

[10] Straumann, D. (2004), "*Estimation in Conditionally Heteroscedastic Time Series Models.*", Springer.

[11] van der Vaart, A.W. (1997), "*Asymptotic Statistics*", Cambridge University Press.