# Calibration by Simulation for Small Sample Bias Correction

Christian GOURIEROUX[1,2] Eric RENAULT[3] Nizar TOUZI[1,4]

July 1994
revised July 1995

[1]CREST, 15 bd Gabriel Péri 92245 Malakoff cedex, France.
[2]CEPREMAP, 142 rue du Chevaleret, 75013 Paris, France.
[3]Institut Universitaire de France, GREMAQ and IDEI, Université de Toulouse I, Place Anatole France 31042 Toulouse cedex, France.
[4]CEREMADE, Université Paris Dauphine.

## Abstract

This paper is interested in the small sample properties of the indirect inference procedure which has been previously studied only from an asymptotic point of view. First, we highlight the fact that the Andrews (1993) median-bias correction procedure for the autoregressive parameter of an AR(1) process is closely related to indirect inference; we prove that the counterpart of the median-bias correction for indirect inference estimator is an exact bias correction in the sense of a generalized mean. Next, assuming that the auxiliary estimator admits an Edgeworth expansion, we prove that indirect inference operates automatically a second order bias correction. The latter is a well known property of the Bootstrap estimator; we therefore provide a precise comparison between these two simulation based estimators.



## Résumé

Cet article s'intéresse aux propriétés de petits échantillons de la méthode par inférence indirecte, qui a été essentiellement étudiée d'un point de vue asymptotique. Nous commençons par noter que la procédure de correction du biais médian proposée par Andrews (1993) pour le paramètre autorégressif d'un processus AR(1) est étroitement reliée à l'approche par inférence indirecte. Nous montrons que la démarche équivalente pour l'estimateur d'inférence indirecte conduit à une correction parfaite du biais au sens d'une moyenne généralisée. Puis, supposant l'existence d'un développement d'Edgeworth pour le paramètre auxiliaire, nous établissons que l'inférence indirecte induit automatiquement une correction de biais jusqu'au second l'ordre. Cette propriété est également satisfaite pour l'estimateur Bootstrap, ce qui nous conduit à comparer ces deux estimateurs corrigés par simulation.



**Keywords** : Bias correction, indirect inference, Bootstrap, Edgeworth correction.

**Mots clefs** : Correction de biais, inférence indirecte, Bootstrap, correction d'Edgeworth.

**JEL classification** : C13.

# Introduction

In this paper we study the small sample properties of the indirect inference procedure, introduced by Smith (1993 [23]) and generalized independently by Gallant-Tauchen (1994 [7]) and Gouriéroux-Monfort-Renault (1993 [11]). This statistical procedure can be seen as an extension of the simulated method of moments in the sense that the information contained in the data is summarized by a general auxiliary criterion rather than a given number of empirical moments. Under usual regularity conditions, the indirect inference estimator has been shown to be consistent and asymptotically normal. In practice, this method has been implemented on simulated or real data, and appeared to perform well (see Pastorello-Renault-Touzi (1993 [18]), Pastorello (1994 [17]), Broze-Scaillet-Zakoian (1993 [4])). In this paper, we provide some additional properties of the indirect inference for small samples.

First, we relate the median bias correction procedure, suggested by Andrews (1993 [1]) for first order autoregressive models (AR(1)), to the general indirect inference procedure. Andrews' [1] procedure is an exact bias correction for the LS estimator in the sense of the median indicator, and can be described as follows : if the least squares (LS) estimator of the autoregressive parameter $\alpha$ for a sample size $T$ is $\widehat{\beta}_T$ then, the estimator $\widehat{\alpha}_T^U$, defined as the value of $\alpha$ that yields the distribution of the LS estimator to have a median of $\widehat{\beta}_T$, is exactly median unbiased. The intuition behind the choice of the median unbiasedness criterion for small sample accuracy seems to be the important skewness of the distribution of the LS estimator, especially when the AR parameter is close to 1, which makes the median a better measure of central tendency than the mean.

Andrews' [1] procedure is shown to be closely related to the indirect inference approach. Therefore, we generalize such a bias correction procedure to a general class of dynamic models. A comparison by simulations between the median bias correction procedure and the indirect inference for AR(p) models is provided in section 3.

However, the most popular bias correction procedures relie on the computation of the bias. In some simple cases, an explicit formula for the small sample bias is available, as for the maximum likelihood estimator of the variance parameter in a sample of independent variables distributed as a normal $\mathcal{N}(m, \sigma^2)$. Such a characterization of the bias can be exploited to define an unbiased estimator from the initial biased one.

In general, an explicit formula for the small sample bias is not available. If the first terms of the bias expansion in powers of $\frac{1}{T}$ can be computed, then a new estimator can be defined such that the bias is reduced up to some order $\frac{1}{T^{\alpha}}$. For instance, Orcutt-Winokur (1969 [16]) showed that the first term in the expansion of the bias of the LS estimator of the AR parameter in an AR(1) model is of order $\frac{1}{T}$, and thus a second order unbiased estimator can be computed (see e.g. Rudebusch (1993 [20])). A generalization of the results of Orcutt-Winokur to the AR(p) case is provided by Shaman-Stine (1989 [22]).

In most cases of interest, even the first terms of the expansion of the bias are difficult to compute explicitly. The Bootstrap estimator, introduced by Efron (1979 [6]), presents the valuable advantage of operating a second order correction of the bias automatically, thanks to simulations. For an infinite number of simulations, we show that the indirect inference also operates a second order bias correction. However, in contrast with the Bootstrap methods, this result does not hold for a finite number of replications. A precise comparison between both estimators up to the third order of an Edgeworth expansion is provided in the case of an infinite number of simulations; we find no evidence for the dominance of one of these methods.

The paper is organized as follows. In *section* 1, we recall briefly the indirect inference procedure and the Andrews [1] bias correction procedure; then we study the exact small sample properties of the indirect inference estimator. In *section* 2, we use Edgeworth expansions in order to examine the second order bias of the indirect inference estimator, and we provide a precise comparison with the Bootstrap estimator. *Section 3* presents some simulation results for AR(1) and AR(2) models, and compares the indirect inference estimator to the median-bias correction procedure of Andrews [1].

# 1 Small sample properties of indirect inference

## 1.1 The indirect inference principle

In this paragraph, we provide a quick review of the indirect inference procedure introduced by Smith [23] and generalized independently by Gallant-Tauchen ([7], GT hereafter) and Gouriéroux-Monfort-Renault ([11], GMR hereafter). It is well known that the estimators of GT and GMR are asymptotically equivalent (see GMR), and that they coincide in the special case where the auxiliary model and the true one have the same number of parameters ($p = d$ in the following notations). The results derived in this paper concern essentially the latter case, and therefore we only present GMR's approach. Consider the general model :

$$\begin{aligned} Z_t &= \varphi(Z_{t-1}, u_t; \theta), & (1.1) \\ Y_t &= r(Y_{t-1}, Z_t; \theta), & (1.2) \end{aligned}$$

where $\{u_t, t = 1 \dots T\}$ is a white noise process with known distribution $G_0$, $\{Z_t, t = 0 \dots T\}$ is an unobservable stationary state variable whose dynamics is characterized by the transition equation (1.1), for a given unknown value $\theta^0$ of the parameter $\theta$, lying in an open bounded subset $\Theta \subset I\!R^p$ and a given function $\varphi$, and $\{Y_t, t = 0 \dots T\}$ is a stationary process whose dynamics is defined by the measurement equation (1.2), for the value $\theta^0$ of the parameter and a given function $r$.

The important feature that the dynamic model (1.1)-(1.2) has to satisfy is that one can draw simulated paths according to it, given a value $\theta$ of the parameter and an initial condition $(\tilde{Y}_0, \tilde{Z}_0)$. This is achieved by drawing independent simulated disturbance paths $\{u_t^h, t = 1 \dots T\}, h = 1 \dots H$, in the distribution $G_0$, and computing simulated paths $\{Y_t^h(\theta),$ $t = 0 \dots T\}$ according to the recursive system :

$$\begin{aligned} Z_t^h(\theta) &= \varphi(Z_{t-1}^h(\theta), u_t^h; \theta), \\ Y_t^h(\theta) &= r(Y_{t-1}^h(\theta), Z_t^h(\theta); \theta), \end{aligned}$$

with initial values $Y_0^h(\theta)$ and $Z_0^h(\theta)$ drawn for instance in the stationary distribution of $(Y, Z)$ with the value $\theta$ of the parameter, or taken as initial fixed values $\tilde{Y}_0$, $\tilde{Z}_0$. The main idea of indirect inference is to match simulated data with observed ones in order to estimate the parameters of the model. Let $Q_T$ be a given function mapping $\mathbb{R}^T \times B$ into $\mathbb{R}$ for some open bounded subset $B$ of $\mathbb{R}^d$ with $d \geq p$, and define :

$$\begin{aligned} \widehat{\beta}_T &= \arg\max_{\beta \in B} Q_T\left(\underline{Y}_T, \beta\right), \\ \widetilde{\beta}^h_T(\theta) &= \arg\max_{\beta \in B} Q_T\left(\underline{Y}^h_T(\theta), \beta\right), \quad h = 1, \ldots, H, \end{aligned} \tag{1.3}$$

where $\underline{Y}_T = \{Y_t,\ t = 1, \ldots, T\}$ and $\underline{Y}^h_T(\theta) = \{Y^h_t(\theta),\ t = 1, \ldots, T\}$. $Q_T$ can be interpreted as the estimation criterion corresponding to an auxiliary model which is a good approximation of the true model, and which allows for classical estimation procedures. The pseudo-maximum likelihood of Gouriéroux-Monfort-Trognon (1989 [10]) is an example of such an auxiliary criterion. As defined, $\widehat{\beta}_T$ summarizes the information given by the sample path $\underline{Y}_T$. For instance, if we choose as auxiliary criterion $Q_T = -\|\overline{k}_T - \beta\|^2$, where $\overline{k}_T$ is a vector of $d$ empirical moments of $Y$, then the sample path $\underline{Y}_T$ is summarized by these $d$ empirical moments.

The indirect inference estimator in the sense of GMR is defined by :

$$\begin{aligned} \widehat{\theta}^H_T &= \arg\min_{\theta \in \Theta} \left\| \widehat{\beta}_T - \frac{1}{H} \sum_{h=1}^{H} \widetilde{\beta}^h_T(\theta) \right\|^2_{\Omega_T} \\ &= \arg\min_{\theta \in \Theta} \left[ \widehat{\beta}_T - \frac{1}{H} \sum_{h=1}^{H} \widetilde{\beta}^h_T(\theta) \right]' \Omega_T^{-1} \left[ \widehat{\beta}_T - \frac{1}{H} \sum_{h=1}^{H} \widetilde{\beta}^h_T(\theta) \right], \end{aligned} \tag{1.4}$$

where $\Omega_T$ is a symmetric positive definite matrix which converges almost surely to a symmetric positive definite matrix $\Omega$. For instance, for $Q_T = -\|\overline{k}_T - \beta\|^2$, $\widehat{\theta}^H_T$ defined in (1.4) is the MSM (Method of Simulated Moments) estimator of $\theta$ (Duffie-Singleton (1993 [5])), and the indirect inference procedure appears as a natural generalization of the MSM.

As the number $H$ of simulated sample paths goes to infinity, the limit indirect inference estimator is :

$$\widehat{\theta}_T = \arg\min_{\theta \in \Theta} \left\| \widehat{\beta}_T - E\left[\widetilde{\beta}_T(\theta)\right] \right\|^2_{\Omega_T}, \tag{1.5}$$

where the expectation is with respect to the distribution $G_0$ of the error term. While the indirect inference procedure is presented as an asymptotic estimation methodology, we focus in this paper on its small sample properties ($T$ small) in the case where the auxiliary and the true models have the same number of parameters i.e. $p = d$; under this condition, the estimator is independent of the weighting matrix $\Omega_T$. Let us define the function $b_T$ mapping $\Theta$ into $b_T(\Theta)$ by :

$$b_T(\theta) = E\left[\widetilde{\beta}_T(\theta)\right], \tag{1.6}$$

which is the binding function in the finite sample context, and assume the usual identifiability condition :

**Assumption 1.1** *The finite sample binding function $b_T$, mapping $\Theta$ into $b_T(\Theta)$, is uniformly continuous and one-to-one.*

In contrast with the usual asymptotic analysis, the distribution of the auxiliary estimator $\widehat{\beta}_T$ may recover some values of $I\!R^p$ which are not attained by the function $b_T$, and the minimum in (1.5) may be positive. In order to simplify the presentation, we therefore make the following additional assumption :

**Assumption 1.2** *The support of the distribution of $\widehat{\beta}_T$ is included in $b_T(\Theta)$.*

Under the last assumption, for an infinite number of replications, the indirect inference estimator is simply given by :

$$\widehat{\theta}_T \;=\; b_T^{-1}(\widehat{\beta}_T). \tag{1.7}$$

In the general case, an expression of the indirect inference estimator in the form (1.7) can always be obtained by considering an (asymptotically equivalent) modification of the estimator (1.5). Thanks to the uniform continuity of $b_T$ on the open bounded set $\Theta$, a continuous extension of $b_T$ to the closure of $\Theta$ exists. This allows to construct an extension $\overline{b}_T$ of $b_T$, which is one-to-one on the whole space $I\!R^p$. We can therefore define the slightly modified indirect inference estimator $\widehat{\widehat{\theta}}_T = \overline{b}_T^{-1}(\widehat{\beta}_T)$ for which the results of subsection 1.3 can be stated in terms of the extension $\overline{b}_T$.

Before studying the small sample properties of the indirect estimator (1.7), let us recall some related bias correction procedures appeared in the literature on autoregressive models.

## 1.2 Median bias correction in autoregressive models

Andrews [1] suggested an exact median unbiased estimator for the autoregressive coefficient of an AR(1) model. Extensions of this methodology to the AR($p$) case have been proposed by Andrews-Chen [2] and Rudebusch [19]; in the case $p > 1$, the estimators are only approximately median unbiased since the median is not suitable for vector variables. The purpose of this section is to provide a presentation of these procedures which highlights the analogy with the indirect inference methodology.

Consider the following latent AR($p$) time series $\{Y_t^*,\ t = 0,\ldots,T\}$ :

$$\phi(L)Y_t^* \;=\; u_t, \;\; \text{for } t = p,\ldots,T, \tag{1.8}$$

where $L$ is the lag operator, $\phi(L) = 1 - \sum_{j=1}^{p} \alpha_j L^j$ is the lag polynomial whose roots are assumed to lie on or outside the unit circle and $\{u_t,\ t = 1\ldots T\}$ is a gaussian white noise with variance $\sigma^2$; $Y_0^*, \ldots Y_{p-1}^*$ are drawn from the stationary distribution of the process $Y^*$, if all the roots of $\phi(L)$ lie outside the unit circle, and are arbitrary constant otherwise. We denote by $\Theta^p$ the set of vectors $\alpha \in I\!R^p$ such that the roots of the polynomial $1 - \sum_{j=1}^{p} \alpha_j x^j$ lie outside the unit circle. In the AR(1) context, it is known that $\Theta^1 = (-1, 1)$. More generally, it is shown in the appendix that $\Theta^p$ is an open bounded subset of $I\!R^p$.

Next, we consider the following models for the observed process $\{Y_t,\ t = 0,\ldots,T\}$ :

$$\begin{aligned}
\text{model 1 :}\;\; Y_t &= Y_t^*,\;\; t = 0,\ldots,T,\\
\text{model 2 :}\;\; Y_t &= \mu + Y_t^*,\;\; t = 0,\ldots,T,\\
\text{model 3 :}\;\; Y_t &= \mu + \gamma t + Y_t^*,\;\; t = 0,\ldots,T,
\end{aligned} \tag{1.9}$$

where $\mu$ and $\gamma$ are two unknown parameters. The nonstationary case ($\alpha$ is on the frontier of $\Theta^p$) can only be handled within models 2 and 3, and the following restriction appears to be necessary :

$$\alpha \in \Theta^p \text{ in model 1 of (1.9).} \tag{1.10}$$

Now, let $\widehat{\beta}_T$ be the (unconstrained) LS estimator of the AR parameters $\alpha = (\alpha_1, \ldots, \alpha_p)'$.[5] Notice that $\widehat{\beta}_T$ is also the maximum likelihood estimator of the AR parameters and therefore inherits the asymptotic efficiency property. However it is biased in finite samples because of the presence of lagged dependent variables which violates the assumption of nonstochastic regressors in the classical linear regression model. In particular, for estimating the sum of the AR coefficients $\sum_{j=1}^{p} \alpha_j$, which is useful in the study of the long run persistence properties[6], the bias tends to be downward and quite large. For estimating the time trend coefficient $\gamma$, the bias is upward and quite large. For an AR(1) model, we refer to table 2 of Rudebusch [20] which shows that the probability to underestimate the AR parameter when the latter is 0.9 equals 0.89 and still increases for values of the AR parameter closer to 1. In practice, the latter case appears very frequently; in financial applications for instance, interest rates or asset prices volatilities are usually modelled by a latent continuous time Ornstein-Uhlenbeck process (see e.g. Vasicek 1977 [25] and Scott 1987 [21]), which time discretization yields to an AR(1) process with AR parameter converging to 1 as the time space between observations goes to zero.

The bias correction procedure, suggested by Andrews in the AR(1) framework, and generalized to the AR($p$) one by Rudebusch [19] and Andrews-Chen [2], relies on the independence of the LS estimator of the AR parameter $\alpha$ on the other parameters of the model. Therefore, given a value of the AR parameter $\alpha$, one can define a unique random variable $\widetilde{\beta}_T(\alpha)$ which is the LS estimator induced by a sample of length $T$, when the true value of the AR parameter is $\alpha$.

In the AR(1) framework, one can define the function $m_T(\alpha)$, as the median of the random variable $\widetilde{\beta}_T(\alpha)$, and the estimator $\widehat{\alpha}_T^U$ by :

$$\widehat{\alpha}_T^U = \arg \min_{\alpha \in [-1,1]} \left| \widehat{\beta}_T - m_T(\alpha) \right|. \tag{1.11}$$

Assuming that the function $m_T(.)$ is increasing (which should be the case from the simulations of Andrews [1]), the last estimator can be written :

$$\widehat{\alpha}_T^U = \begin{cases} 1 & \text{if } \widehat{\beta}_T > m_T(1), \\ m_T^{-1}(\widehat{\beta}_T) & \text{if } m_T(-1) < \widehat{\alpha}_T \leq m_T(1), \\ -1 & \text{if } \widehat{\beta}_T \leq m_T(-1), \end{cases} \tag{1.12}$$

where $m_T(\pm 1) = \lim_{\alpha \to \pm 1} m_T(\alpha)$. The estimator defined in (1.12) is median unbiased since, by the increasing property of the function $m_T(.)$, we have $\widehat{\alpha}_T^U \geq \alpha$ iff $m_T(\widehat{\alpha}_T^U) \geq m_T(\alpha)$,

[5] There is no specification error in the sense that if model $i \in \{1, 2, 3\}$ is the true model, then the regression is performed according to the same model $i$.

[6] Andrews-Chen [2] suggested to use the cumulative impulse response (CIR) as a measure of persistence that summarizes the information contained in the impulse response function (IRF); in the context of an AR(p) model, this measure turns to be a very simple function of the sum of the AR coefficients : CIR $= 1/(1 - \sum_{j=1}^{p} \alpha_j)$.

and from the definition of $\widehat{\alpha}_T^U$, this is equivalent to $\widehat{\beta}_T \geq m_T(\alpha)$. Note that the median unbiasedness property of $\widehat{\alpha}_T^U$ does not depend on the values assigned on the bounds $m_T(-1)$ and $m_T(1)$ since the median of a distribution does not depend on the values taken from both sides of the median; these values have just to be larger than $\alpha$ if $\widehat{\beta}_T > m_T(1)$ and vice versa, and, since $\alpha$ lies in $(-1, 1)$, the values on the bounds in (1.12) are well suited.

The practical implementation of this procedure requires the computation of the median function $m_T$. By drawing simulated paths $\{\widetilde{y}_t^h(\alpha), t = 0, \ldots, T\}$, $h = 1, \ldots, H$, and computing the LS estimator for each path, we get $H$ independent and identically distributed realizations $\widetilde{\beta}_T^h(\alpha)$, $h = 1, \ldots, H$. An approximation of $m_T(\alpha)$ can thus be obtained as the median of the $\widetilde{\beta}_T^h(\alpha)$'s. For an infinite number of simulated paths, such an approximation converges towards the required limit $m_T(\alpha)$. Therefore, the median unbiased estimator suggested by Andrews [1] is nothing but an application of the indirect inference where the binding function defined in (1.6) is replaced by the median of the auxiliary estimator. An important feature of this application of indirect inference is that the auxiliary model coincides with the true model.

The problem in generalizing the median bias correction procedure to the AR($p$) case is that the median indicator is not suited for vector variables. However, defining the median of the vector variable $\widetilde{\beta}_T(\alpha)$ as the median of each individual variable, Rudebusch [20] and Andrews-Chen [2] suggested a direct generalization of the last procedure for the AR($p$) framework. Unfortunately, such estimators are only approximately median unbiased because of the inadequacy of the median indicator within this context.

In the sequel, we study the small sample properties of the indirect inference estimator which handles with any vector variable since it is based on the mean indicator.

## 1.3 Mean bias correction by indirect inference

In this section, we provide analogous sample properties for the indirect inference estimator (1.7).

**Proposition 1.1** *(i) Suppose that the true model and the auxiliary one have the same number of parameters ($p = d$). Then, under assumptions 1.1 and 1.2, the indirect inference estimator $\widehat{\theta}_T$ defined in* (1.5) *is $b_T$-mean unbiased i.e.*

$$b_T^{-1}\left\{E\left[b_T\left(\widehat{\theta}_T\right)\right]\right\} \;=\; \theta^0,$$

*where $\theta^0$ is the true value of the parameters.*
*(ii) Suppose that the auxiliary model coincides with the true one, and that the first step estimator $\widehat{\beta}_T$ is mean unbiased i.e. $E(\widehat{\beta}_T) = \theta^0$. Then the indirect inference estimator $\widehat{\theta}_T$ coincides with the first step estimator i.e. $\widehat{\theta}_T = \widehat{\beta}_T$.*

**Proof.** $(i)$ From the expression (1.7) of the indirect inference estimator, $E[b_T(\widehat{\theta}_T)] = E[\widehat{\beta}_T]$ $= E[\widetilde{\beta}_T(\theta^0)] = b_T(\theta^0)$, and the result follows from the one-to-one property of the function $b_T(.)$.
$(ii)$ The result is obvious since $b_T$ is the identity function under the unbiasedness condition. □

The second part of the proposition says that if the first step estimator is mean unbiased, then the indirect inference procedure does not make it worse. Part $(i)$ is the counterpart of the median unbiasedness property in the Andrews bias correction procedure. In particular, if the bias of the estimator is an affine function of the unknown parameter i.e. $b_T$ is affine, then the indirect inference estimator is exactly mean-unbiased. But, since $b_T$ is unknown in general, we can not conclude from this property that indirect inference will reduce the bias of the first step estimator.

In order to understand the result of the first part of the last proposition consider the following iterative procedure. Define the function $b_T^{(1)}$ from the estimator $\widehat{\theta}_T$ as $b_T$ has been defined from $\widehat{\beta}_T$ i.e. $b_T^{(1)}(\theta) = E_\theta(\widehat{\theta}_T)$, for $\theta \in \Theta$. We can therefore define the estimator $\widehat{\theta}_T^{(1)}$ $= b_T^{(1)^{-1}}(\hat{\theta}_T)$. More generally, we define the sequence of estimators :

$$\widehat{\theta}_T^{(k)} \;=\; b_T^{(k)^{-1}}(\widehat{\theta}_T^{(k-1)}) \quad \text{with} \quad b_T^{(k)}(\theta) \;=\; E_\theta\left(\widehat{\theta}_T^{(k-1)}\right), \quad \theta \in \Theta, \tag{1.13}$$

assuming that the functions $b_T^{(k)}$ and the estimators $\widehat{\theta}_T^{(k-1)}$ satisfy assumptions 1.1 and 1.2. Then, if this procedure converges i.e. if the limits $b_T^{(\infty)} = \lim_{k\to\infty} b_T^{(k)}$ and $\widehat{\theta}_T^{(\infty)} = \lim_{k\to\infty} \widehat{\theta}_T^{(k)}$ exist, then the limit estimator $\widehat{\theta}_T^{(\infty)}$ is mean-unbiased. To see this notice that for such a limit point, we have $b_T^{(\infty)}(\widehat{\theta}_T^{(\infty)}) = \widehat{\theta}_T^{(\infty)}$, which means that $b_T^{(\infty)}$ equals the identity function on $\widehat{\theta}_T^{(\infty)}(\Theta)$, the set of values which might be taken by the estimator $\widehat{\theta}_T^{(\infty)}$ when the true value of the parameter $\theta$ varies in $\Theta$. It then follows from the definition of $b_T^{(\infty)}$ that $E_{\theta_0}[\widehat{\theta}_T^{(\infty)}] = b_T^{(\infty)}(\theta^0) = \theta^0$, where the last equality follows from the fact that $\theta^0 \in \widehat{\theta}_T^{(\infty)}(\Theta)$.

## 1.4 Examples

**Example 1.** Consider independent and identically $\mathcal{N}(m, \sigma^2)$ distributed observations $Y_1$, ..., $Y_T$, and take as a first step (auxiliary) estimator the maximum likelihood one. Then it is well known that the variance estimator :

$$\widehat{s}_T^2 \;=\; \frac{1}{T}\sum_{t=1}^{T}(Y_t - \overline{Y}_T)^2, \text{ with } \overline{Y}_T \;=\; \frac{1}{T}\sum_{t=1}^{T} Y_t, \tag{1.14}$$

is biased in finite samples. The expectation of this first step estimator can be computed explicitly in this simple example :

$$E(\widehat{s}_T^2) \;=\; \sigma^2 - \frac{1}{T}\sigma^2,$$

and the finite sample binding function $E[\widehat{s}_T^2(.)]$ is thus linear in the variance parameter $\sigma^2$. Therefore the indirect inference estimator (1.7) (corresponding to an infinite number of replications) is unbiased, and is equal to :

$$\widehat{\sigma}_T^2 \;=\; \frac{T}{T-1}\widehat{s}_T^2. \tag{1.15}$$

**Example 2.** In the previous example, we pointed out the fact that the indirect inference estimator is mean-unbiased if the bias of the auxiliary estimator is an affine function of the true value of the parameter. We now give an example where the (finite sample) binding function is a power of the true value of the parameter. Consider the model :

$$Y_t = \theta u_t,$$

and the estimator :

$$\widehat{\beta}_T = \overline{Y}_T^k = \left(\frac{1}{T}\sum_{t=1}^{T} Y_t\right)^k,$$

for a given integer $k$. The expectation of this estimator is :

$$b_T(\theta) = \mu(T,k)\theta^k \quad \text{where} \quad \mu(T,k) = E\left[\left(\frac{1}{T}\sum_{t=1}^{T} u_t\right)^k\right],$$

so that the indirect inference estimator is given by :

$$\widehat{\theta}_T = b_T^{-1}(\widehat{\beta}_T) = [\mu(T,k)]^{-1/k}\overline{Y}_T.$$

The expectation of this estimator is :

$$b_T^{(1)}(\theta) = [\mu(T,k)]^{-1/k}\mu(T,1)\,\theta.$$

Since $b_T^{(1)}$ is a linear function of the parameter $\theta$, the next step estimator $\widehat{\theta}_T^{(1)} = b_T^{(1)^{-1}}(\widehat{\theta}_T)$ is unbiased and for any $p \geq 2$, $\widehat{\theta}_T^{(p)} = \widehat{\theta}_T^{(1)}$.

## 2 Edgeworth expansions

When the bias of a given estimator can be computed, as in the case of the variance parameter of a linear regression, a mean-unbiased estimator can be defined from the initial estimator. However, the bias can not be computed explicitly in general. Another approach consists in computing explicitly the first terms of the expansion of the bias in $\frac{1}{T}$, so as to define a new estimator with reduced bias. This is a usual practice in autoregressive models, where the expansion up to the first order has been provided by Orcutt-Winokur (1969 [16]) for AR(1) models and by Shaman-Stine (1989 [22]) for general AR(p) models. However, even this methodology requires the explicit computation of the expansion up to some order, which is very difficult in general.

Bootstrap methods introduced by Efron (1979 [6]), which are based on simulations, has been shown to operate the latter correction automatically : the bias of order $\frac{1}{T}$ disappears in the Bootstrap estimator. In the following sections, we show that the indirect inference estimator presents the same property and we compare both estimation methodologies by focusing on the next term of the expansion.

## 2.1 Second order bias correction by indirect inference

We suppose again that the auxiliary model and the true one have the same number of parameters, and that the auxiliary estimator $\widehat{\beta}_T$ is a consistent estimator of the parameter $\theta^0$. The following analysis relies on the assumption that the auxiliary estimator admits an Edgeworth expansion :

$$\widehat{\beta}_T = \theta^0 + \frac{A(v,\theta^0)}{\sqrt{T}} + \frac{B(v,\theta^0)}{T} + \frac{C(v,\theta^0)}{T^\alpha} + o(\frac{1}{T^\alpha}), \tag{2.1}$$

where $\alpha \in \{\frac{3}{2}, 2\}$, $A(v,\theta^0)$, $B(v,\theta^0)$ and $C(v,\theta^0)$ are random vectors depending on some asymptotic random term $v$, and the expansion is to be understood in the probability sense (see e.g. Hall 1992 [12], chapter 2). The next order after the $\frac{1}{T}$ one is $\frac{1}{T^2}$ in most situations; in order to deal with the general case we introduce the order $\frac{1}{T^\alpha}$ where $\alpha$ could be $\frac{3}{2}$ or 2. Such an Edgeworth expansion exits in many cases of interest where the statistic under consideration has a limiting standard normal distribution (see Hall [12], paragraph 2.3, p. 46).

Under some regularity conditions on the random coefficients $A$, $B$ and $C$ of the Edgeworth expansion (2.1), we can show that the indirect inference estimator also has an Edgeworth expansion which can be fully characterized :

**Proposition 2.1** *Under some regularity conditions, the indirect inference estimator $\widehat{\theta}_T^H$, given in* (1.4), *has the following Edgeworth expansion :*

$$\widehat{\theta}_T^H = \theta^0 + \frac{A_H^*}{\sqrt{T}} + \frac{B_H^*}{T} + \frac{C_H^*}{T^{3/2}} + o(T^{-3/2}), \tag{2.2}$$

*where the coefficients $A_H^*$, $B_H^*$ and $C_H^*$ are deduced from $A$, $B$ and $C$ by :*

$$A_H^* = A(v,\theta^0) - \frac{1}{H}\sum_{h=1}^{H} A(v^h,\theta^0), \tag{2.3}$$

$$B_H^* = B(v,\theta^0) - \frac{1}{H}\sum_{h=1}^{H} B(v^h,\theta^0) - \left[\frac{1}{H}\sum_{h=1}^{H}\frac{\partial A}{\partial\theta'}(v^h,\theta^0)\right] A_H^*, \tag{2.4}$$

$$\begin{aligned} C_H^* = {} & \left[C(v,\theta^0) - \frac{1}{H}\sum_{h=1}^{H} C(v^h,\theta^0)\right] 1_{\{\alpha=3/2\}} \\ & - \left[\frac{1}{H}\sum_{h=1}^{H}\frac{\partial B}{\partial\theta'}(v^h,\theta^0)\right] A_H^* - \left[\frac{1}{H}\sum_{h=1}^{H}\frac{\partial A}{\partial\theta'}(v^h,\theta^0)\right] B_H^* - \frac{1}{2}A_H^{*\prime}\left[\frac{1}{H}\sum_{h=1}^{H}\frac{\partial^2 A}{\partial\theta\partial\theta'}(v^h,\theta^0)\right] A_H^*; \end{aligned} \tag{2.5}$$

*the random variables $v$, $v^h$, $h = 1\ldots H$ are independent and identically distributed.*

**Proof.** See appendix 2

□

Let us first consider the limit case of an infinite number of replications. We get :

$$\begin{aligned} A^*_\infty &= \lim_{H\to\infty} A^*_H = A(v,\theta^0) - E[A(v,\theta^0)], \\ B^*_\infty &= \lim_{H\to\infty} B^*_H = B(v,\theta^0) - E[B(v,\theta^0)] - E\left[\frac{\partial A}{\partial \theta'}(v,\theta^0)\right]\left\{A(v,\theta^0) - E[A(v,\theta^0)]\right\}, \end{aligned}$$

and we can deduce the following result :

**Corollary 2.1** *The indirect inference estimator $\widehat{\theta}_T$, corresponding to an infinite number of simulation* (1.5), *is unbiased up to the second order i.e. the terms of order $\frac{1}{\sqrt{T}}$ and $\frac{1}{T}$ in the Edgeworth expansion satisfy : $E(A^*_\infty) = E(B^*_\infty) = 0$.*

For a fixed number of replications $H$, the last property is not satisfied. We still have $E(A^*_H) = 0$ and the first order bias vanishes. In contrast with the auxiliary estimator, the second order bias of the indirect inference estimator does not depend on the coefficient $B$, and is determined by :

$$\begin{aligned} E(B^*_H) &= -E\left\{\frac{1}{H}\sum_{h=1}^{H}\frac{\partial A}{\partial \theta'}(v^h,\theta^0)\left[A(v,\theta^0) - \frac{1}{H}\sum_{h=1}^{H}A(v^h,\theta^0)\right]\right\} \\ &= -\sum_{j=1}^{p} E\left\{\frac{1}{H}\sum_{h=1}^{H}\frac{\partial A}{\partial \theta_j}(v^h,\theta^0)\left[A_j(v,\theta^0) - \frac{1}{H}\sum_{h=1}^{H}A_j(v^h,\theta^0)\right]\right\} \\ &= -\sum_{j=1}^{p} Cov\left\{\frac{1}{H}\sum_{h=1}^{H}\frac{\partial A}{\partial \theta_j}(v^h,\theta^0)\ ;\ A_j(v,\theta^0) - \frac{1}{H}\sum_{h=1}^{H}A_j(v^h,\theta^0)\right\} \\ &= \frac{1}{H^2}\sum_{j=1}^{p}\sum_{h=1}^{H} Cov\left\{\frac{\partial A}{\partial \theta_j}(v^h,\theta^0)\ ;\ A_j(v^h,\theta^0)\right\} \\ &= \frac{1}{H}\sum_{j=1}^{p} Cov\left\{\frac{\partial A}{\partial \theta_j}(v,\theta^0)\ ;\ A_j(v,\theta^0)\right\}, \end{aligned} \tag{2.6}$$

where the equalities follow from the independence of the random variables $v$ and $v^h$, $h = 1,\ldots,H$. Therefore the second order bias of the indirect inference estimator is smaller than the auxiliary estimator one as soon as :

$$\frac{1}{H}\left|\sum_{j=1}^{p}\text{Cov}\left\{\frac{\partial A}{\partial \theta_j}(v,\theta^0); A_j(v,\theta^0)\right\}\right| \leq \left| E[B(v,\theta^0)]\right|, \tag{2.7}$$

which provides the minimum number of replications in order to improve the second order bias of the estimator.

## 2.2 Comparison with the Bootstrap bias correction

The important result of corollary 2.1 is a well known property of the Bootstrap etimators, which are also based on simulations. We first recall briefly the expansion of the Bootstrap estimator in our context before comparing it with the indirect inference one.

Suppose that the auxiliary estimator $\widehat{\beta}_T$ is a consistent estimator of the parameter $\theta^0$, so that $E[\widetilde{\beta}_T(\widehat{\beta}_T)] - \widehat{\beta}_T$ is a "good" approximation of the bias of $\widehat{\beta}_T$. Therefore, by drawing $H$ replications in the distribution induced by the parameter $\theta = \widehat{\beta}_T$, the Bootstrap estimator is defined by :

$$\overline{\theta}_T^H = \widehat{\beta}_T - \left[\frac{1}{H}\sum_{h=1}^{H}\widetilde{\beta}_T^h(\widehat{\beta}_T) - \widehat{\beta}_T\right]. \tag{2.8}$$

As in the previous section, if we assume that the first step (auxiliary) estimator has an Edgeworth expansion, the Bootstrap estimator defined in (2.8) also has an Edgeworth expansion (under some regularity conditions) :

$$\overline{\theta}_T^H = \theta^0 + \frac{A_H^b}{\sqrt{T}} + \frac{B_H^b}{T} + \frac{C_H^b}{T^{3/2}} + o(T^{-3/2}), \tag{2.9}$$

where the coefficients $A_H^b$, $B_H^b$ and $C_H^b$ are deduced from $A$, $B$ and $C$ by :

$$A_H^b = A(v,\theta^0) - \frac{1}{H}\sum_{h=1}^{H} A(v^h,\theta^0) \tag{2.10}$$

$$B_H^b = B(v,\theta^0) - \frac{1}{H}\sum_{h=1}^{H} B(v^h,\theta^0) - \left[\frac{1}{H}\sum_{h=1}^{H}\frac{\partial A}{\partial\theta'}(v^h,\theta^0)\right] A(v,\theta) \tag{2.11}$$

$$\begin{aligned} C_H^b = {} & \left[C(v,\theta^0) - \frac{1}{H}\sum_{h=1}^{H} C(v^h,\theta^0)\right] 1_{\{\alpha=3/2\}} \\ & - \left[\frac{1}{H}\sum_{h=1}^{H}\frac{\partial B}{\partial\theta'}(v^h,\theta^0)\right] A(v,\theta^0) - \left[\frac{1}{H}\sum_{h=1}^{H}\frac{\partial A}{\partial\theta'}(v^h,\theta^0)\right] B(v,\theta^0) \\ & -\frac{1}{2}A'(v,\theta^0)\left[\frac{1}{H}\sum_{h=1}^{H}\frac{\partial^2 A}{\partial\theta\partial\theta'}(v^h,\theta^0)\right] A(v,\theta^0). \end{aligned} \tag{2.12}$$

As for the indirect inference estimator, the first order bias is zero, i.e. $E(A_H^b) = 0$. The second order bias is given by :

$$\begin{aligned} E(B_H^b) &= -\frac{1}{H}\sum_{h=1}^{H} E\left[\frac{\partial A}{\partial\theta'}(v^h,\theta)A(v,\theta)\right] \\ &= E[A(v,\theta^0)]E\left[\frac{\partial A}{\partial\theta'}(v,\theta^0)\right], \end{aligned} \tag{2.13}$$

where the last equality follows from the independence of the random variables $v$ and $v^h$, $h = 1,\ldots,H$. Equation (2.13) shows that :

- If $E[A(v,\theta^0)] \neq 0$, then, even for an infinite number of simulations, the Bootstrap estimator presents a second order bias; from this viewpoint, the indirect inference estimator is preferred since its second order bias vanishes for an infinite $H$.

- If $E[A(v,\theta^0)] = 0$, then the second order bias of the Bootstrap estimator vanishes for a finite number of replications $H$, i.e. $E(B_H^b) = 0$; from this viewpoint, the Bootstrap estimator dominates the indirect inference one.

In the case $H = \infty$ and $E[A(v,\theta^0)] = 0$, both estimators correct from the second order bias. We therefore examine the third order bias. We denote by $C_\infty^* = \lim_{H\to\infty} C_H^*$ and $C_\infty^b = \lim_{H\to\infty} C_H^b$; using again the independence of $v$ and the $v^h$, $h = 1,\ldots,H$, and the fact that $E[A(v,\theta^0)] = 0$, we obtain the third order bias of the indirect inference estimator :

$$E(C_\infty^*) \;=\; -\frac{1}{2}E\left[A(v,\theta^0)'E\left(\frac{\partial^2 A}{\partial\theta\partial\theta'}(v,\theta^0)\right)A(v,\theta^0)\right], \tag{2.14}$$

and that of the Bootstrap estimator :

$$E(C_\infty^b) \;=\; -E\left[\frac{\partial A}{\partial\theta'}(v,\theta^0)\right]E[B(v,\theta^0)] - \frac{1}{2}E\left[A(v,\theta^0)'E\left(\frac{\partial^2 A}{\partial\theta\partial\theta'}(v,\theta^0)\right)A(v,\theta^0)\right] \tag{2.15}$$

Clearly, the two expressions (2.14) and (2.15) can not be compared in general, and the indirect inference estimator and the Bootstrap one are competitors for an infinite number of simulations.

## 2.3 Examples

To illustrate these results, we consider again the first example of paragraph 1.4 : $Y_1$, ..., $Y_T$ are independent and identically distributed observations from the normal $\mathcal{N}(m,\sigma^2)$, and the first step (auxiliary) estimator is the maximum likelihood one :

$$\widehat{m}_T \;=\; \frac{1}{T}\sum_{t=1}^{T} Y_t \quad \text{and} \quad \widehat{s}_T^2 \;=\; \frac{1}{T}\sum_{t=1}^{T}(Y_t - \widehat{m}_T)^2.$$

By drawing $H$ replications, we can construct :

$$\tilde{s}_T^{2^h}(m,\sigma^2) \;=\; \frac{1}{T}\sum_{t=1}^{T}\left[Y_t^h(m,\sigma^2) - \widetilde{m}_T^h(m,\sigma^2)\right]^2.$$

Recalling that $Y_t^h(m,\sigma^2) = m + \sigma u_t^h$, where $u_t^h$, $h = 1,\ldots,T$ are drawn independently in the standard normal distribution, the last expression can be written as :

$$\tilde{s}_T^{h^2}(m,\sigma^2) \;=\; \sigma^2\frac{1}{T}\sum_{t=1}^{T}(u_t^h - \overline{u}_T^h)^2, \quad \text{with} \quad \overline{u}_T^h \;=\; \frac{1}{T}\sum_{t=1}^{T}u_t^h.$$

Therefore, by equating $\widehat{s}_T^2$ with $\frac{1}{H}\sum_{h=1}^{H}\tilde{s}_T^{2^h}(m,\sigma^2)$, we obtain the indirect inference estimator of $\sigma^2$ :

$$\widehat{\sigma}_T^{2^H} \;=\; \sigma^{0^2}\frac{\sum_{t=1}^{T}(u_t - \overline{u}_T)^2}{\frac{1}{H}\sum_{h=1}^{H}\sum_{t=1}^{T}(u_t^h - \overline{u}_T^h)^2}.$$

The finite sample distribution of the indirect inference estimator is such that :

$$\frac{\hat{\sigma}_T^{2^H}}{\sigma^{0^2}} \rightsquigarrow F[T-1, H(T-1)],$$

where $F(p, q)$ stands for the Fisher distribution, and in the limit case $H = \infty$, we have :

$$\frac{\hat{\sigma}_T^2}{\sigma^{0^2}} \rightsquigarrow \chi^2(T-1).$$

For fixed $H$, the bias of the indirect inference estimator is given by :

$$E(\hat{\sigma}_T^2) - \sigma^{0^2} = \frac{2\sigma^{0^2}}{H(T-1)-2}, \tag{2.16}$$

while the bias of the first step estimator is :

$$E(\hat{s}_T^2) - \sigma^{0^2} = -\frac{1}{T}\sigma^{0^2}.$$

We can thus conclude that the indirect inference estimator bias is smaller than that of the first order estimator as soon as :

$$H > 2\,\frac{T+1}{T-1}.$$

# 3 Simulations results

In this section, we examine the empirical content for AR(p) models of the theoretical results of sections 1 and 2. Indeed, unbiasedness of the indirect inference estimators does not give direct intuition on their performance. Section 3.1 gives simulations results for the AR(1) case and compares the median unbiased estimator of Andrews [1] to the indirect inference estimator introduced in (1.5). Then, section 3.2 presents an application to the AR(2) context and compares the indirect inference methodology to the approximately median bias correction procedures suggested by Rudebusch [19] and Andrews-Chen [2].

As stated previously, we do not present any application for the general model (1.1)-(1.2), and we refer to Pastorello-Renault-Touzi [18] for an application of indirect inference to the estimation of the volatility process parameters, from option prices data, in stochastic volatility models which are popular in option pricing literature. More precisely, these authors compare the estimators obtained by an E.M. algorithm combined, with the Andrews' bias correction methodology, to the indirect inference estimators. Their results show that the latter estimators perform as well as the former ones even though no (appearent) bias correction is performed.

## 3.1 Application to AR(1) models

As in Andrews [1], our assumption 1.1 is not justified by an analytic proof and we use simulations to chek its validity for different sample sizes $T$. Simulations are performed as described in section 2 for model 2 of (1.9) without time trend. Values of $b_T(\alpha)$ and $m_T(\alpha)$ for $\alpha = -0.99$, $-0.995 + 0.005k$, $k = 0, \ldots, 399$, are computed by Monte Carlo simulations without using any numerical trick to improve the efficiency of the algorithm : $b_T(\alpha)$ and $m_T(\alpha)$ are simply approximated by their finite sample counterparts with finite large enough $H$.

In order to agree with Andrews' numerical results presented in table 2 of [1] up to the third decimal, we had to use a very large number of simulations ($H \geq 25,000$). Table 1.1 and figure 1.1 present our simulation results with $H = 15,000$, which provides very close values for the function $m_T$ to those of Andrews' table 2, and show clearly the increasing feature of $m_T$ and $b_T$.

TABLE 1.1. MEAN AND MEDIAN OF THE LS ESTIMATOR OF THE AR PARAMETER IN AN AR(1) MODEL WITHOUT TIME TREND. (H=15,000)

| | $T=40$ | | $T=50$ | | $T=80$ | |
|---|---|---|---|---|---|---|
| $\alpha$ | Mean | Median | Mean | Median | Mean | Median |
| −.999 | −.937 | −.950 | −.989 | −.997 | −.964 | −.973 |
| −.80 | −.727 | −.750 | −.780 | −.789 | −.761 | −.771 |
| −.60 | −.545 | −.560 | −.586 | −.596 | −.570 | −.578 |
| −.40 | −.363 | −.371 | −.397 | −.404 | −.381 | −.385 |
| −.20 | −.181 | −.184 | −.208 | −.212 | −.191 | −.191 |
| .00 | −.000 | −.002 | −.022 | −.023 | −.001 | −.000 |
| .10 | .091 | .095 | .072 | .073 | .094 | .095 |
| .20 | .181 | .187 | .166 | .169 | .189 | .191 |
| .30 | .272 | .280 | .260 | .265 | .284 | .287 |
| .40 | .362 | .372 | .354 | .361 | .379 | .384 |
| .50 | .452 | .465 | .452 | .461 | .474 | .480 |
| .60 | .542 | .556 | .545 | .556 | .569 | .576 |
| .70 | .631 | .647 | .638 | .651 | .664 | .673 |
| .80 | .719 | .736 | .730 | .745 | .759 | .769 |
| .85 | .762 | .782 | .775 | .792 | .805 | .816 |
| .90 | .805 | .824 | .819 | .836 | .852 | .863 |
| .93 | .829 | .849 | .845 | .862 | .879 | .890 |
| .97 | .860 | .880 | .877 | .895 | .913 | .925 |
| .99 | .874 | .893 | .892 | .910 | .929 | .941 |
| 1.00 | .880 | .899 | .899 | .916 | .936 | .947 |

Comparing the functions $b_T$ and $m_T$, we see that there exists some $\alpha_T^*$ ($\approx 0.06$, for $T = 50$) such that the median of the LS estimator is larger than the its mean if and only if the true value of the AR parameter is larger than $\alpha_T^*$, i.e. the distribution of the LS estimator for the AR parameter is skewed to the right for values of $\alpha$ larger than $\alpha_T^*$, and to the left for values of $\alpha$ smaller than $\alpha_T^*$. A direct consequence is that, when the LS estimator of the AR parameter $\widehat{\beta}_T$ lies in $m_T([-1,1]) \cap \varepsilon_T([-1,1])$ :

$$\text{if } \widehat{\beta}_T > \alpha_T^*, \text{ then } \hat{\alpha}_T > \hat{\alpha}_T^U,$$
$$\text{if } \widehat{\beta}_T < \alpha_T^*, \text{ then } \hat{\alpha}_T < \hat{\alpha}_T^U,$$
$$\text{if } \widehat{\beta}_T = \alpha_T^*, \text{ then } \hat{\alpha}_T = \hat{\alpha}_T^U.$$

As suggested by Andrews for computing the median unbiased estimator, the indirect inference estimator can be determined by linear interpolation in table 1.1. Figure 1.1 justifies the linear approximation of the functions $m_T$ and $b_T$. Table 1.2 provides some comparative results for indirect inference and median unbiased estimators, and shows that there is no important difference between them in practice. Hence, we can conclude that the indirect inference estimator suggested in this paper perform as well as the median unbiased one suggested by Andrews [1]. Finally, figure 1.2 plots the kernel estimates of the densities of the LS estimator $\widehat{\beta}_T$ and the indirect inference one $\widehat{\alpha}_T$ for a sample size $T = 50$ and values of the AR parameter $\alpha = 0.75, 0.85$.

Table 1.2. Some comparative results between
Median Unbiased and Indirect Inference estimators
($T = 50$)

| $\alpha = -0.8$ | | | $\alpha = 0.2$ | | | $\alpha = 0.9$ | | |
|---|---|---|---|---|---|---|---|---|
| LS $\widehat{\beta}_T$ | MU $\widehat{\alpha}_T^U$ | II $\widehat{\alpha}_T$ | LS $\widehat{\beta}_T$ | MU $\widehat{\alpha}_T^U$ | II $\widehat{\alpha}_T$ | LS $\widehat{\beta}_T$ | MU $\widehat{\alpha}_T^U$ | II $\widehat{\alpha}_T$ |
| −.746 | −.755 | −.765 | .222 | .255 | .260 | .817 | .878 | .897 |
| −.859 | −.871 | −.888 | .109 | .138 | .139 | .843 | .908 | .928 |
| −.592 | −.596 | −.607 | .322 | .359 | .366 | .681 | .732 | .747 |
| −.768 | −.779 | −.793 | .034 | .059 | .064 | .814 | .875 | .894 |
| −.779 | −.790 | −.805 | .324 | .361 | .368 | .807 | .867 | .885 |
| −.823 | −.834 | −.850 | .184 | .216 | .219 | .811 | .871 | .890 |
| −.756 | −.766 | −.778 | .087 | .115 | .116 | .647 | .696 | .710 |
| −.710 | −.718 | −.732 | .110 | .139 | .141 | .589 | .635 | .648 |
| −.830 | −.842 | −.857 | −.036 | −.017 | −.022 | .816 | .877 | .896 |
| −.879 | −.892 | −.909 | .246 | .280 | .285 | .910 | .990 | 1.000 |

## 3.2 Application to AR(2) models

First, we characterize the set $\Theta^2$ of autoregressive coefficients $\alpha = (\alpha_1, \alpha_2) \in I\!R^2$ such that the latent process $Y^*$, with convenient initial values, is stationary i.e. the roots of the lag polynomial operator $\phi(x) = 1 - \alpha_1 x - \alpha_2 x^2$ are outside the unit circle. Let $\Delta = \alpha_1^2 + 4\alpha_2$ be the discriminant of the lag polynomial operator.

- *Case A* : $\Delta \geq 0$, then, since $\phi(0) = 1 > 0$, the roots of $\phi(.)$ are outside the unit circle iff $\phi(-1) > 0$ and $\phi(1) > 0$, i.e. $1 - \alpha_1 - \alpha_2 > 0$ and $1 + \alpha_1 - \alpha_2 > 0$.
- *Case B* : $\Delta < 0$, then $\phi$ has two conjugate complex roots which are outside the unit circle iff $\alpha_2 \in (-1, 0)$.

We thus conclude that :

$$\begin{aligned} \Theta^2 &= \Big\{(\alpha_1,\alpha_2) \in I\!R^2 \mid \alpha_1^2 + 4\alpha_2 \geq 0 \ , \ 1 - \alpha_1 - \alpha_2 > 0 \text{ and } 1 - \alpha_1 - \alpha_2 > 0 \Big\} \\ &\quad \cup \Big\{(\alpha_1,\alpha_2) \in I\!R^2 \mid \alpha_1^2 + 4\alpha_2 \geq 0 \text{ and } -1 < \alpha_2 < 0 \Big\}, \end{aligned}$$

which can be written in :

$$\Theta^2 = \Big\{(\alpha_1,\alpha_2) \in I\!R^2 \mid \alpha_1 + \alpha_2 < 1 \ , \ \alpha_1 - \alpha_2 > -1 \text{ and } \alpha_2 > -1 \Big\}. \qquad (3.1)$$

In contrast with the application 3.1, the finite sample binding function $b_T$ is now charging a subset of $\mathbb{R}^2$, and verification of assumption 1.1 through Monte-Carlo simulations is much more time consuming. By definition, given the LS estimator $\widehat{\beta}_T$, the indirect inference estimator $\widehat{\alpha}_T$ is the unique solution to the equation $\widehat{\beta}_T = \overline{b}_T(\alpha)$, where $\overline{b}_T$ is an extension of $b_T$, as described in section 1. Since no explicit expression of the function $\overline{b}_T^{-1}$ is available, we have to use numerical methods in order to solve for $\widehat{\alpha}_T$, which usually require numerical evaluation of the gradient of the finite sample binding function.

For the present AR(2) model, we use an algorithm which is likely to avoid such time consuming numerical procedures. The basic idea behind our procedure, is that the finite sample binding function in the AR(1) framework is close to the identity function (up to a constant), according to figure 1.1. Therefore, we can hope that such a property is still valid for the AR(2) case so that the function :

$$g_T^1(\alpha) \;=\; \alpha + \widehat{\beta}_T - \overline{b}_T(\alpha), \tag{3.2}$$

is a strong contraction, and the indirect inference estimator $\widehat{\alpha}_T$ is its unique fixed point. Thus, for a given $\widehat{\beta}_T$ we construct the sequence $(\widehat{\alpha}_T^{(n)})_{n\geq 0}$ by :

$$\widehat{\alpha}_T^{(0)} \;=\; \widehat{\beta}_T \quad \text{and} \quad \widehat{\alpha}_T^{(n+1)} \;=\; g_T^1\left(\widehat{\alpha}_T^{(n)}\right). \tag{3.3}$$

If $g_T^1$ is a strong contraction, this sequence converges towards the unique fixed point $\widehat{\alpha}_T$.

For our application, we consider $\alpha_1 = 1.2$ and $\alpha_2 = -0.4$ as true values of the AR parameters, and we fix $\sigma = 0.5$ and $\mu = 1$. The sample size is set to $T = 40$ and the number of simulations in the indirect inference procedure is fixed to $H = 5000$, i.e. $b_T(\alpha)$ is approximated by its sample moment counterpart with 5000 observations. We perform 1000 experiments by simulating the AR(2) process, computing the corresponding LS and indirect inference estimators, and we construct (gaussian) kernel estimates of the density of each estimator.

The algorithm described above appears to perform well since convergence of the procedure, up to an error of $10^{-4}$, is achieved for a maximum of 6 iterations[7]. However, for some simulated paths, the LS estimator happens to be close to the frontier of the set $b_T(\Theta^2)$ and the algorithm fails to be contracting. In such cases, we define the sequence :

$$\widehat{\alpha}_T^{(0)} = \widehat{\beta}_T \quad \text{and} \quad \widehat{\alpha}_T^{(n+1)} = g_T^{\lambda}\left(\widehat{\alpha}_T^{(n)}\right),$$

where :

$$g_T^{\lambda}(\alpha) \;=\; \alpha + \lambda\left(\widehat{\beta}_T - \overline{b}_T(\alpha)\right),$$

and $\lambda$ is chosen so as $g_T^{\lambda}$ is a strong contraction. In our application, we obtain convergence in all cases with $\lambda = 0.2$.

We also wish to compare the performance of indirect inference to the approximately median unbiased procedure suggested by Rudebusch [19]. We therefore compute for the same

[7] More precisely, let $\overline{b}_T(\alpha) = (\overline{b}_{1,T}(\alpha), \overline{b}_{2,T}(\alpha))'$ and $\widehat{\beta}_T = (\widehat{\beta}_{1,T}, \widehat{\beta}_{2,T})'$; by convention, convergence of the algorithm occurs when $\sum_{j=1}^{2} |\overline{b}_{j,T}(\widehat{\alpha}_T) - \widehat{\beta}_{j,T}| \leq 0.0001$.

experiments the associated approximately median unbiased estimators and we construct kernel estimates of their density functions. Figure 2.1 presents plots of kernel estimates of the density functions of the different estimators and shows clearly the bias correcting feature of both indirect inference and approximately median unbiased procedures; the mean and the median of the different estimators are reported in table 2.1. Another important conclusion that we can draw from figure 2.1 is that the indirect inference and the approximately median unbiased estimators are very close, as already noticed in the AR(1) context of section 4.1, and there is no significant difference between their mean and median. Therefore, we conclude that the $\bar{b}_T$-mean indicator of central tendency is a good measure which takes into account the asymmetry of the LS estimator distribution.

TABLE 2.1. ESTIMATORS OF THE AR COEFFICIENTS ($\alpha_1 = 1.2$ AND $\alpha_2 = -0.4$) $\mu = 1$, $\sigma = 0.5$, $T = 40$, $H = 5{,}000$, $1{,}000$ EXPERIMENTS.

| | $\alpha_1$ | | $\alpha_2$ | |
|---|---|---|---|---|
| Procedure | Mean | Median | Mean | Median |
| LS $\beta_T$ | 0.936 | 0.947 | −0.191 | −0.215 |
| MU $\widehat{\alpha}_T^U$ | 1.191 | 1.213 | −0.396 | −0.430 |
| II $\widehat{\alpha}_T$ | 1.202 | 1.229 | −0.406 | −0.444 |

Next, we compare the indirect inference procedure to the approximately median unbiased procedure of Andrews-Chen [2]. These authors suggested a generalization of Andrews' [1] methodology to the AR(p) case in the same way as Rudebusch [19], but using a "Dickey-Fuller" regression form for the AR(p) model :

$$Y_t^* = \gamma_1 Y_{t-1}^* + \gamma_2 \Delta Y_{t-1}^* + \ldots + \gamma_p \Delta Y_{t-p+1}^* + u_t \,, \quad t = p, \ldots, T, \qquad (3.4)$$

where : $\Delta Y_{t-i}^* = Y_{t-i}^* - Y_{t-i-1}^*$, $\gamma_1 = \sum_{i=1}^p \alpha_i$ and $\gamma_i = -\sum_{j=i}^p \alpha_j$ for $i = 2 \ldots p$. In our AR(2) context, we have $\gamma_1 = \alpha_1 + \alpha_2$ and $\gamma_2 = -\alpha_2$, and the set $\Theta^2$ in terms of this new parameterization can be deduced from (3.1) :

$$\Theta^2 = \left\{ (\gamma_1, \gamma_2) \in I\!R^2 \mid \gamma_1 < 1 \,,\ \gamma_2 < 1 \,,\ \text{and } \gamma_1 + 2\gamma_2 > -1 \right\}. \qquad (3.5)$$

Indirect inference and approximately median unbiased estimators are simultaneously computed according to the same numerical procedure as above. Figure 2.2 contains plots of kernel estimates of the density function for the different estimators of the AR coefficients, and shows clearly the bias correcting feature of both indirect inference and approximately median unbiased procedures; the mean and the median of the different estimators are reported in table 2.2. As noticed before, the two procedures produce very close estimators and the difference between their mean and median is very small.

TABLE 2.2. DICKEY-FULLER REGRESSION FORM ($\gamma_1 = 0.8$ AND $\gamma_2 = 0.4$)
$\mu = 1$, $\sigma = 0.5$, $T = 40$, $H = 5,000$, 1,000 EXPERIMENTS.

| | $\gamma_1$ | | $\gamma_2$ | |
|---|---|---|---|---|
| Procedure | Mean | Median | Mean | Median |
| LS $\hat{\beta}_T$ | 0.743 | 0.752 | 0.194 | 0.210 |
| MU $\hat{\alpha}_T^U$ | 0.783 | 0.796 | 0.388 | 0.416 |
| II $\hat{\alpha}_T$ | 0.794 | 0.807 | 0.410 | 0.446 |

Finally, we compare estimators of the sum of the AR coefficients obtained in the regular form and the "Dickey-Fuller" regression one of the AR(2) model. Figure 2.3 presents kernel estimates of the density function of each estimator and shows that these estimators are very close, as noticed by Andrews-Chen [2].

## APPENDIX 1

In this appendix, we prove that the set $\Theta^p$ of parameters $\alpha \in \mathbb{R}^p$ which induce a stationary AR(p) model is an open bounded subset of $\mathbb{R}^p$.

Let $(\alpha_j^{(n)})$, $j = 1, \ldots, p$, be $p$ real valued sequences converging towards $\alpha_j$, $j = 1, \ldots, p$ as $n$ goes to infinity, and consider the polynomial operators $\phi_n(z) = 1 - \sum_{j=1}^p \alpha_j^{(n)} z^j$ and $\phi(z) = 1 - \sum_{j=1}^p \alpha_j z^j$. We denote by $\mathcal{Z}_n$ and $\mathcal{Z}$ the set of roots of the polynomials $\phi_n$ and $\phi$. Now suppose that all the elements of $\mathcal{Z}_n$ are inside or on the unit circle i.e. $\forall n \in \mathbb{N}$, $\alpha^{(n)} = (\alpha_1^{(n)}, \ldots, \alpha_p^{(n)}) \in \mathbb{R}^p \setminus \Theta^p$, and consider a sequence $(z_n)$ such that $\forall n$, $z_n \in \mathcal{Z}_n$. Then it is easily seen that $\phi(z_n) = \sum_{j=1}^p (\alpha_j^{(n)} - \alpha_j) z_n^j$ and therefore $|\phi(z_n)| \leq \sum_{j=1}^p |\alpha_j^{(n)} - \alpha_j|$. Thus, from the convergence of $(\alpha_j^{(n)})_n$ towards $\alpha_j$, for $j = 1, \ldots, p$, we have $\lim_{n \to \infty} \phi(z_n) = 0$, which proves that the elements of $\mathcal{Z}$ lie in or on the unit circle as limit of elements of $\mathcal{Z}_n$ i.e. $\alpha \in \mathbb{R}^p \setminus \Theta^p$. Hence, $\mathbb{R}^p - \Theta^p$ is a closed subset of $\mathbb{R}^p$.
To see that $\Theta^p$ is bounded, recall that the lag polynomial can be written in terms of its roots as $\phi(x) = \prod_{z \in \mathcal{Z}} (1 - \frac{x}{z})$, so that the coefficents $\alpha_j$, $j = 1, \ldots, p$ are linear combinations of products of the $\frac{1}{z}$'s, $z \in \mathcal{Z}$. Since $|z| > 1$ for all $z \in \mathcal{Z}$, it is clear that the coefficients $\alpha_j$, $j = 1, \ldots, p$, are bounded, which proves that $\Theta^p$ is bounded.

## APPENDIX 2
Edgeworth expansion of the indirect inference estimator

The Edgeworth expansion (2.1) may be applied both to the auxiliary estimator :

$$\hat{\beta}_T \;=\; \theta^0 + \frac{A(v, \theta^0)}{\sqrt{T}} + \frac{B(v, \theta^0)}{T} + \frac{C(v, \theta^0)}{T^\alpha} + o(\frac{1}{T^\alpha}),$$

and to the estimators based on the simulated values :

$$\tilde{\beta}_T^h(\theta) = \theta + \frac{A(v^h,\theta)}{\sqrt{T}} + \frac{B(v^h,\theta)}{T} + \frac{C(v^h,\theta)}{T^\alpha} + o(\frac{1}{T^\alpha}), \quad h = 1,\ldots,H,$$

where $v$ and $v^h$, $h = 1\ldots H$ are independent with identical distribution. Now the indirect inference estimator is the solution $\hat{\theta}_T^H$ of :

$$\hat{\beta}_T^H = \frac{1}{H}\sum_{h=1}^{H} \tilde{\beta}_T^h(\hat{\theta}_T^H).$$

By plugging the Edgeworth expansions of $\hat{\beta}_T$ and $\tilde{\beta}_T^h$ we get :

$$\begin{aligned} &\theta^0 + \frac{A(v,\theta^0)}{\sqrt{T}} + \frac{B(v,\theta^0)}{T} + \frac{C(v,\theta^0)}{T^\alpha} + o(T^{-\alpha}) \\ = &\frac{1}{H}\sum_{h=1}^{H}\left[\hat{\theta}_T^H + \frac{A(v^h,\hat{\theta}_T^H)}{\sqrt{T}} + \frac{B(v^h,\hat{\theta}_T^H)}{T} + \frac{C(v^h,\hat{\theta}_T^H)}{T^\alpha} + o(T^{-\alpha})\right], \end{aligned}$$

which provides the form of the Edgeworth expansion for $\hat{\theta}_T^H$ as follows. Let :

$$\hat{\theta}_T^H = \theta^0 + \frac{A_H^*}{\sqrt{T}} + \frac{B_H^*}{T} + \frac{C_H^*}{T^{3/2}} + o(T^{-3/2})$$

be the Edgeworth expansion of $\hat{\theta}_T^H$. Then by Taylor expansion of $A(v^h,\hat{\theta}_T^H)$, $B(v^h,\hat{\theta}_T^H)$ and $C(v^h,\hat{\theta}_T^H)$ around $\theta^0$ and keeping only terms of order lower than $T^{-\alpha}$ we get :

$$\begin{aligned} &\theta^0 + \frac{A(v,\theta^0)}{\sqrt{T}} + \frac{B(v,\theta^0)}{T} + \frac{C(v,\theta^0)}{T^\alpha} + o(T^{-\alpha}) \\ = \;&\theta^0 + \frac{A_H^*}{\sqrt{T}} + \frac{B_H^*}{T} + \frac{C_H^*}{T^{3/2}} + o(T^{-3/2}) \\ &+\left\{\frac{1}{\sqrt{T}}\frac{1}{H}\sum_{h=1}^{H} A(v^h,\theta^0) + \frac{1}{T}\frac{1}{H}\sum_{h=1}^{H}\frac{\partial A}{\partial\theta'}(v^h,\theta^0)A_H^* \right. \\ &\left. +\frac{1}{T^{3/2}}\frac{1}{H}\sum_{h=1}^{H}\frac{\partial A}{\partial\theta'}(v^h,\theta^0)B_H^* + \frac{1}{2}\frac{1}{T^{3/2}}\frac{1}{H}\sum_{h=1}^{H} A_H^{*\prime}\frac{\partial^2 A}{\partial\theta\partial\theta'}(v^h,\theta^0)A_H^* + o\left(T^{-3/2}\right)\right\} \\ &+\left\{\frac{1}{T}\frac{1}{H}\sum_{h=1}^{H} B(v^h,\theta^0) + \frac{1}{T^{3/2}}\frac{1}{H}\sum_{h=1}^{H}\frac{\partial B}{\partial\theta'}(v^h,\theta^0)A_H^* + o\left(T^{-3/2}\right)\right\} \\ &+\frac{1}{T^\alpha}\frac{1}{H}\sum_{h=1}^{H} C(v^h,\theta^0) + o(T^{-\alpha}). \end{aligned}$$

Identifying both sides of the equality provides the result announced in proposition 2.1.

FIGURE 1.1. MEDIAN AND MEAN OF THE LS ESTIMATOR OF THE AR PARAMETER IN AR(1) MODELS.

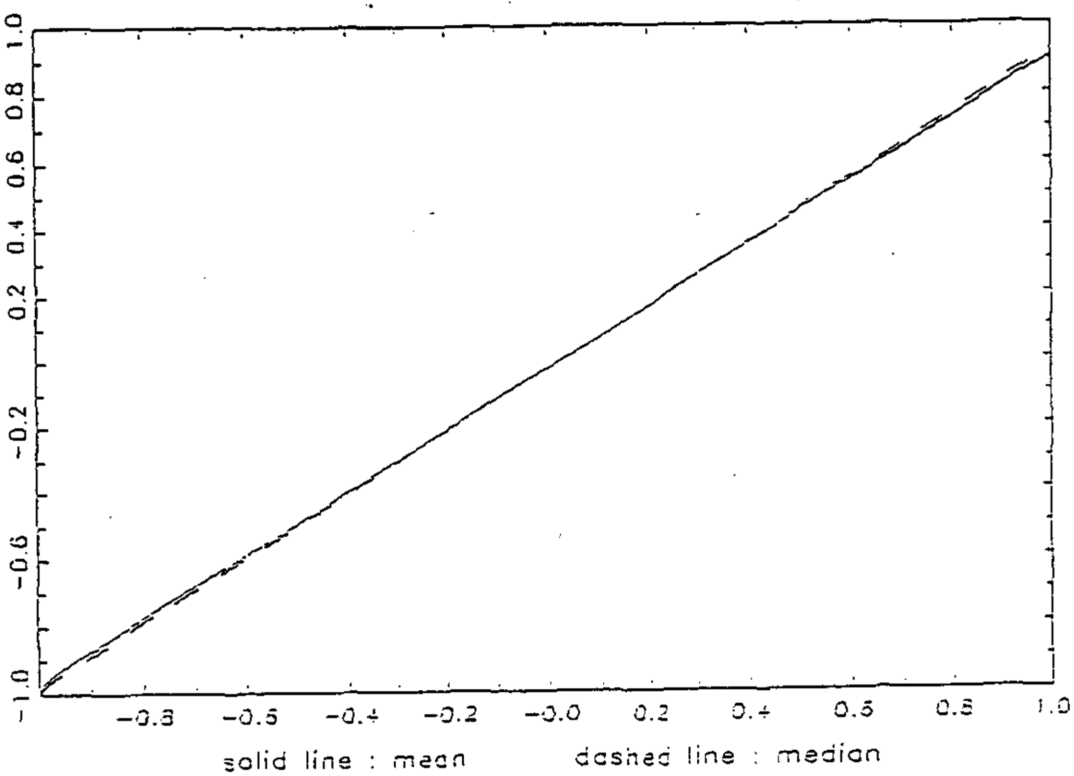
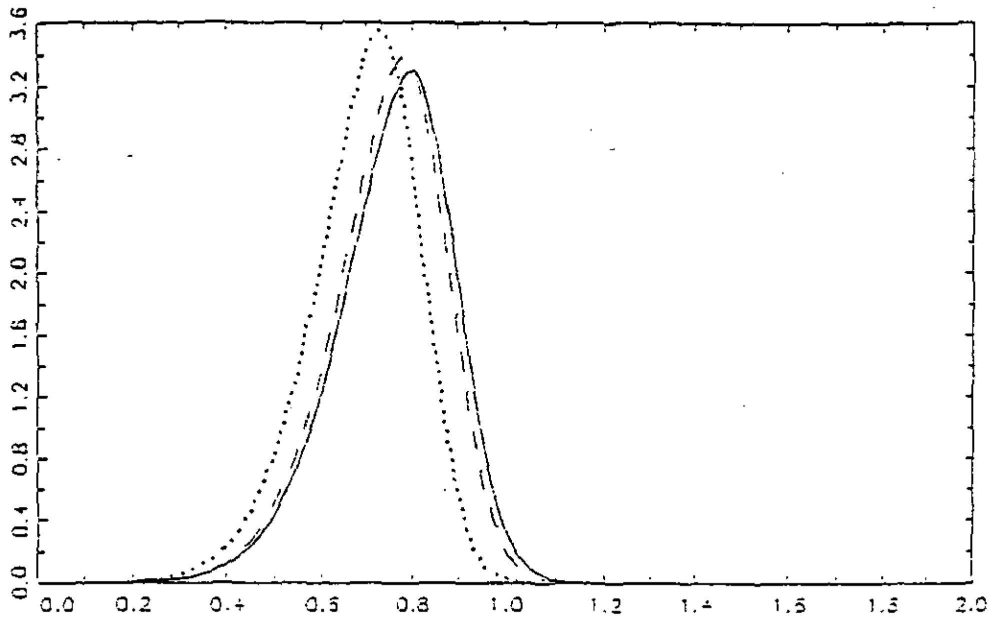
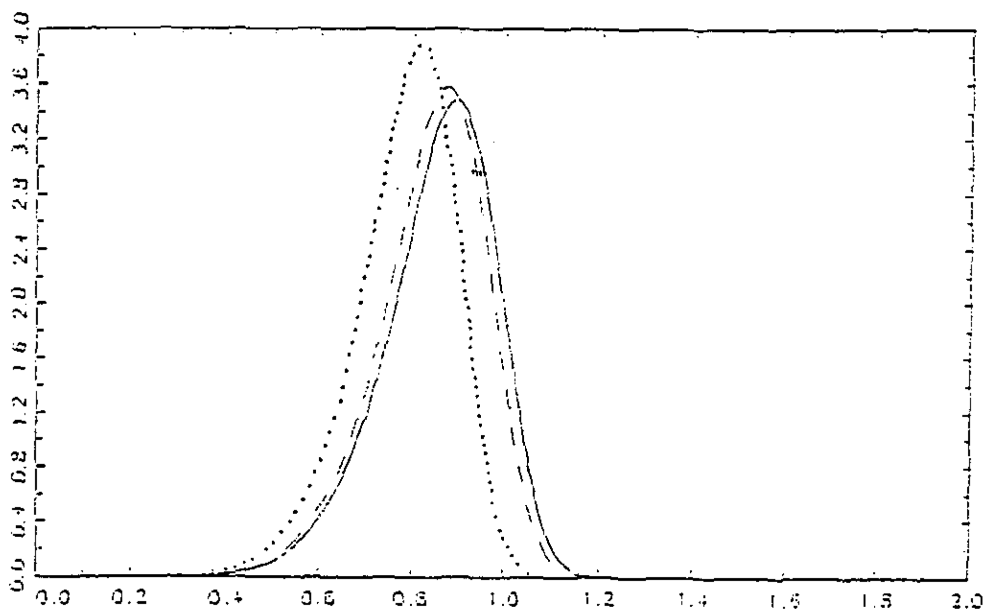


FIGURE 1.2. KERNEL ESTIMATES OF THE DENSITY OF THE ESTIMATORS OF THE AR PARAMETER IN AN AR(1) MODEL.
$\mu = 1$, $\sigma = 0.5$, $T = 50$, $H = 15,000$, 5.000 EXPERIMENTS.

$\alpha = 0.75$




$\alpha = 0.85$



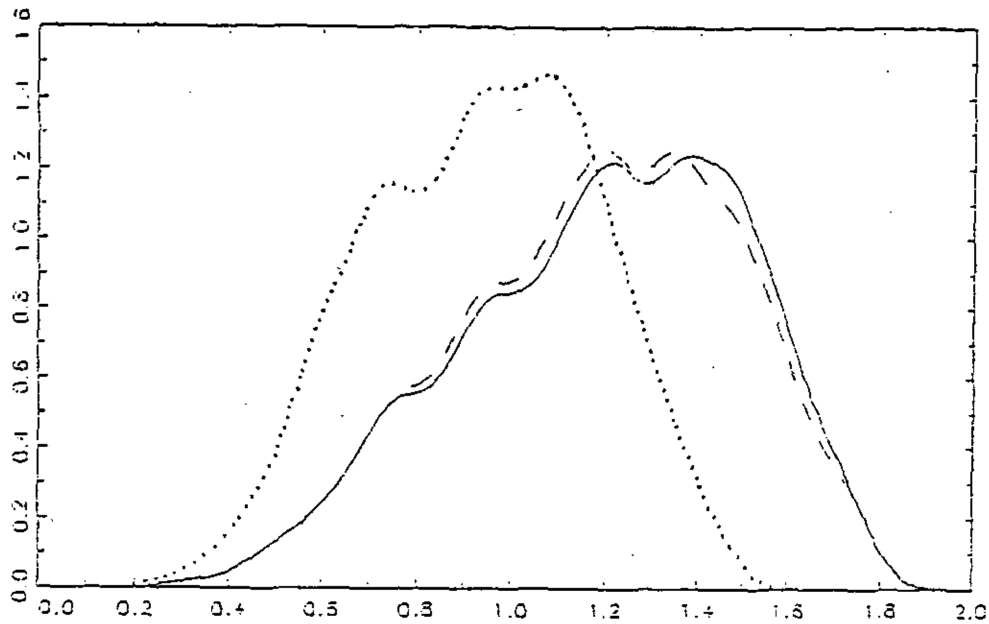

solid line : I.I. dashed line : M.U. dotted line : L.S.

FIGURE 2.1. KERNEL ESTIMATES OF THE DENSITY OF THE ESTIMATORS OF THE AR PARAMETERS IN AN AR(2) MODEL

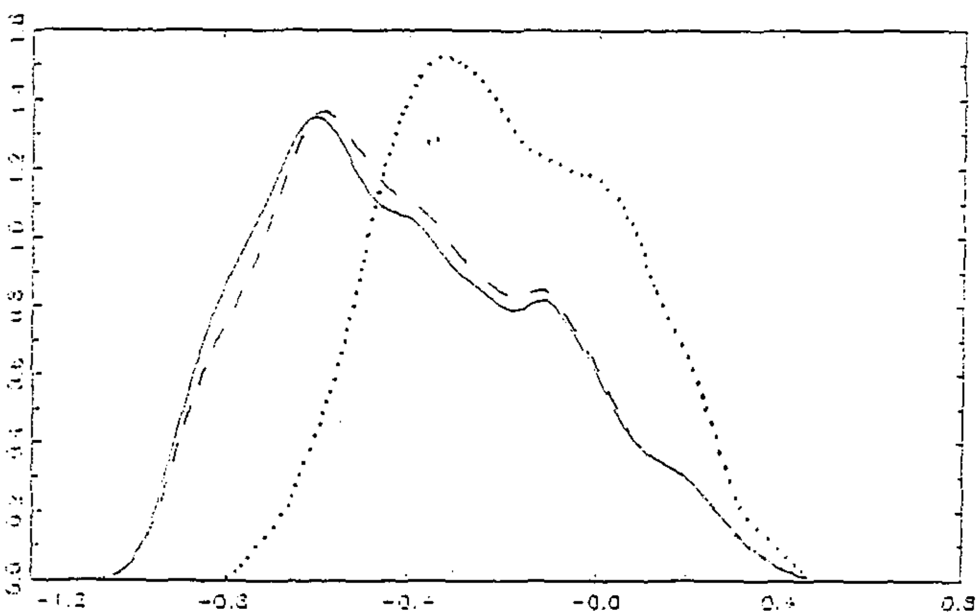
(REGULAR REGRESSION FORM)

$\mu = 1$, $\sigma = 0.5$, $\alpha_1 = 1.2$, $\alpha_2 = -0.4$, $T = 40$, $H = 5{,}000$, 1,000 EXPERIMENTS.

Estimators of $\alpha_1$



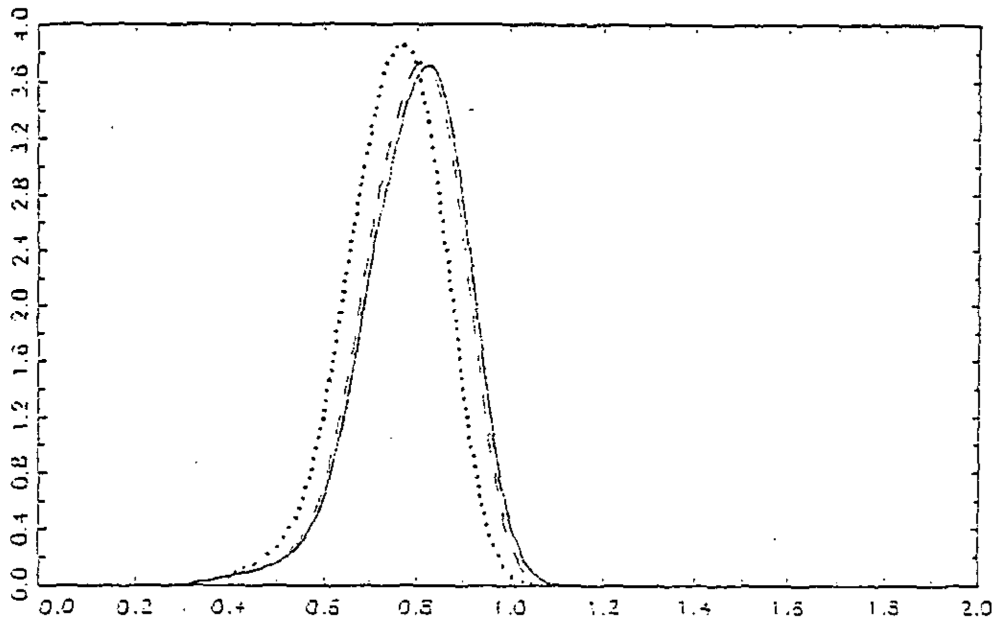

Estimators of $\alpha_2$



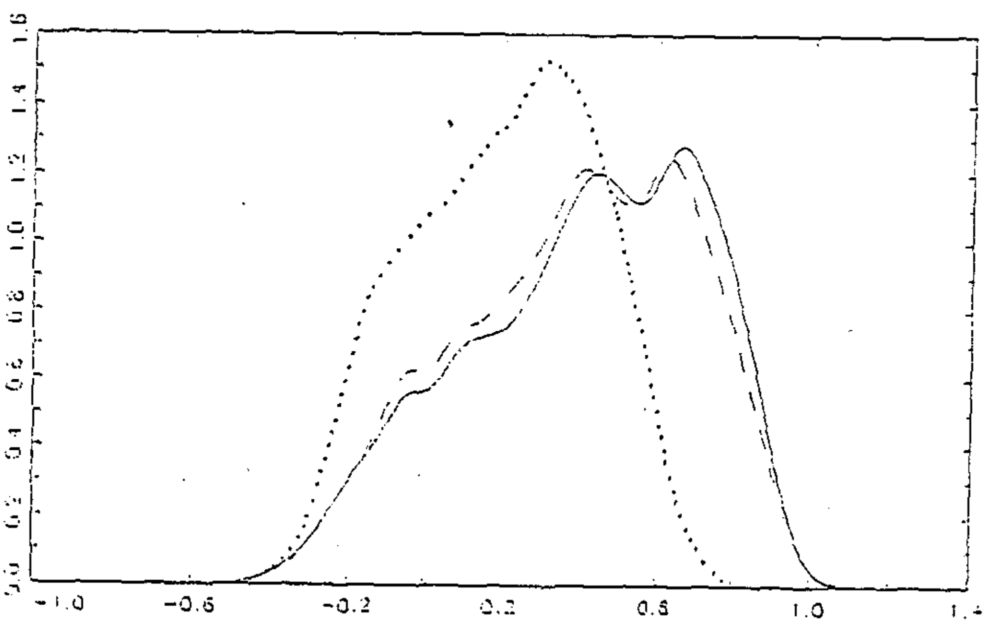

solid line : I.I. dashed line : M.U. dotted line : L.S.

FIGURE 2.2. KERNEL ESTIMATES OF THE DENSITY OF THE ESTIMATORS OF THE AR PARAMETERS IN AN AR(2) MODEL
(DICKEY-FULLER REGRESSION FORM)
$\mu = 1,\ \sigma = 0.5,\ \gamma_1 = 1.2,\ \gamma_2 = 0.4,\ T = 40,\ H = 5{,}000,\ 1{,}000$ EXPERIMENTS.

Estimators of $\gamma_1$



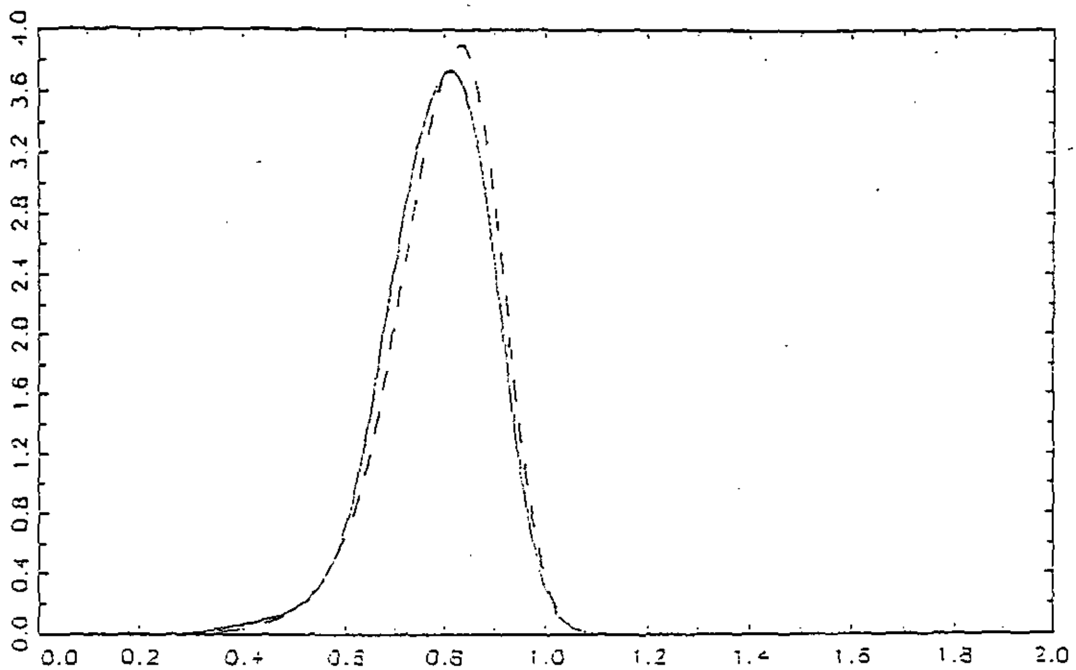

Estimators of $\gamma_2$



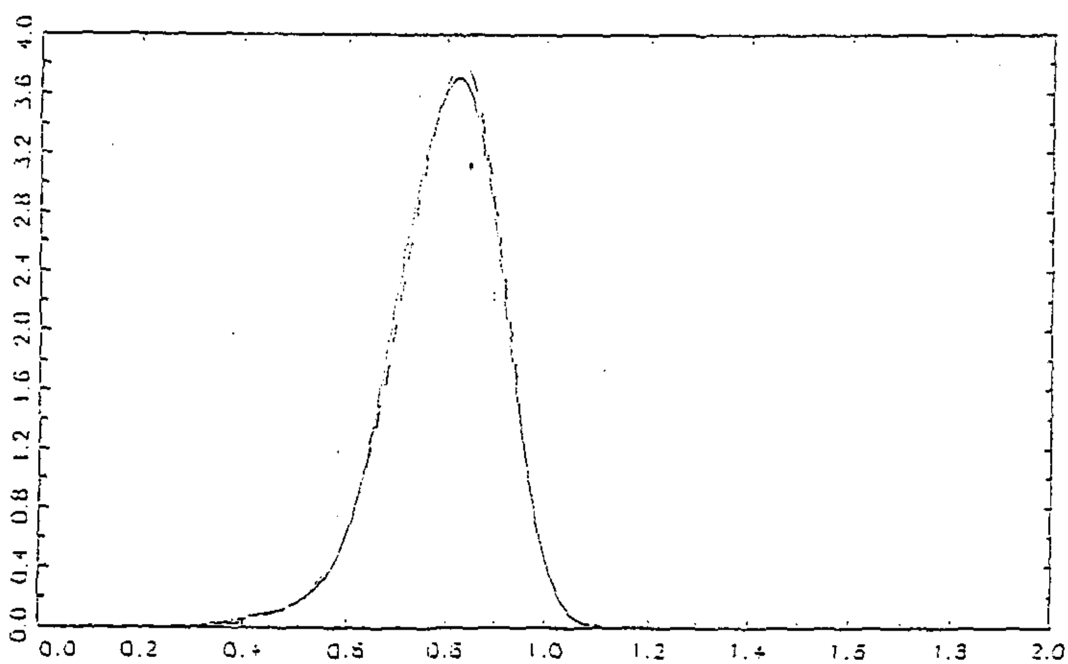

solid line : I.I. dashed line : M.U. dotted line : L.S.

FIGURE 2.3. ESTIMATION OF THE SUM OF THE AR COEF. IN AN AR(2) MODEL.
REGULAR VERSUS DICKEY-FULLER REGRESSION FORM.
$\mu = 1,\ \sigma = 0.5,\ \alpha_1 = 1.2,\ \alpha_2 = -0.4,\ T = 40,\ H = 5{,}000,\ 1{,}000$ EXPERIMENTS.

Approximately Median Unbiased (M.U.) Estimators




Indirect Inference (I.I.) Estimators




solid line : DF    dashed line : Regular AR(2)

## References


[1] Andrews D.W.K (1993), "Exactly Median-Unbiased Estimation of First Order Autoregressive / Unit Root Models", Econometrica 61, 1, pp. 139-165.

[2] Andrews D.W.K and H.Y. Chen (1992), "Approximately Median-Unbiased Estimation of Autoregressive Models with Application to U.S. Macroeconomic and Financial Time Series", Cowles Foundation Discussion Paper 1026, Yale University.

[3] Babu G.J. and C.R. Rao (1993), "Bootstrap Methodology", Hanbdbook of Statistics, vol. 10, Elsevier Science Publishers.

[4] Broze L, O. Scaillet and J.M. Zakoian (1993), "Testing for Continuous-Time Models of the Short-Term Interest Rate", CORE Discussion Paper 9331.

[5] Duffie D. and K.J. Singleton (1993), "Simulated Moments Estimation of Markov Models of Asset Prices", Econometrica 61, 3, pp.929-952.

[6] Efron B. (1979), "Bootstrap Methods : Another Look at the Jackknife", Ann. Statist. 7, pp. 1-26.

[7] Gallant R.A. and G. Tauchen (1994), "Which Moments to Match", Forthcoming in Econometric Theory.

[8] Gallant R.A. and H. White (1988), "Estimation and Inference for Nonlinear Dynamic Models", Blackwell.

[9] Gouriéroux C. and A. Monfort (1995), "Simulation Based Estimation Methods", CORE Lecture Series.

[10] Gouriéroux C, A. Monfort and A. Trognon (1989), "Pseudo-Maximum Likelihood Methods : Theory", Econometrica 52, pp. 681-700.

[11] Gouriéroux C, A. Monfort and E. Renault (1993), "Indirect Inference", Journal of Applied Econometrics, Vol. 8, S85-S118.

[12] Hall P. (1992), "The Bootstrap and Edgeworth Expansion", Springer Series in Statistics, Springer-Verlag.

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

[14] Heynen R, A. Kemna and T. Vorst (1991), "Analysis of the Term Structure of Implied Volatility", Discussion Paper, Erasmus Center for Financial Research, The Netherlands.

[15] Mariott F.H.C. and J.A. Pope (1954), "Bias in the Estimation of Autocorrelations", Biometrika 41, pp. 393-403.

[16] Orcutt G.H. and H.S. Winokur Jr. (1969), "First Order Autoregression : Inference, Estimation and Prediction", Econometrica 37, pp. 1-14.

[17] Pastorello S. (1994), "Estimating Random Variance Option Pricing Models : An Empirical Analysis", Discussion paper, University of Padova, Italy.

[18] Pastorello S, E. Renault and N. Touzi (1993), "Statistical Inference for Random Variance Option Pricing", Discussion Paper CREST and GREMAQ.

[19] Rudebusch G.D. (1992), "Trends and Random Walks in Macroeconomic Time Series : A Re-examination", International Economic Review 33, pp. 661-680.

[20] Rudebusch G.D. (1993), "The Uncertain Unit Root in Real GNP", The American Economic Review 83, 1, pp. 264-272.

[21] Scott L. (1987), "Option Pricing when The Variance Changes Randomly : Theory, Estimation and Application", Journal of Financial and Quantitative Analysis 22, pp. 419-438.

[22] Shaman P. and R.A. Stine (1989), "A Fixed Point Characterization for Bias of Autoregressive Estimators", Annals of Statistics 17, 3, pp. 1275-1284.

[23] Smith A.A. (1993), "Estimating Nonlinear Time-Series Models Using Simulated Vector Autoregressions", Journal of Applied Econometrics, Vol. 8, S63-S84.

[24] Stock J.H. (1991), "Confidence Intervals for the Largest Autoregressive Root in U.S. Macroeconomic Time Series", Journal of Monetary Economics 28, pp. 435-459.

[25] Vasicek O. (1977), "An Equilibrium Characterization of the Term Structure", Journal of Financial Economics 5, pp. 177-188.