GRÉDI
Groupe de Recherche en Économie et Développement International

Cahier de recherche / Working Paper
07-19

GLS Bias Correction for Low Order ARMA Models

Patrick Richard

UNIVERSITÉ DE SHERBROOKE

GLS Bias Correction for Low Order ARMA models

Patrick Richard
GREDI and Département de sciences économiques
Faculté d'administration, Université de Sherbrooke
Sherbrooke, Québec, Canada
patrick.richard2@usherbrooke.ca

Abstract

We study the problems of bias correction in the estimation of low order ARMA(p, q) time series models. We introduce a new method to estimate the bias of the parameters of ARMA(p, q) process based on the analytical form of the GLS transformation matrix of Galbraith and Zinde-Walsh (1992). We show that the resulting bias corrected estimator is consistent and asymptotically normal. We also argue that, in the case of an MA(q) model, our method may be considered as an iteration of the analytical indirect inference technique of Galbraith and Zinde-Walsh (1994). The potential of our method is illustrated through a series of Monte Carlo experiments. We find that it is particularly well suited for the estimation of MA(1) models.

I thank Russell Davidson, John Galbraith, James MacKinnon, David Stephens and Victoria Zinde-Walsh as well as the participants at the 2006 CIREQ Ph.D. student conference for their comments on various versions of this paper.

September 2007

1 Introduction

Accurate estimation of time series models belonging to the ARMA(p, q) family is of critical importance in several macroeconometric applications. Also, the accuracy of asymptotic and bootstrap inference is often function of the precision with which the models parameters are estimated. It is therefore not surprising that a lot of effort has been devoted throughout the years to the characterisation and correction of the bias of the most commonly used estimators.

The bias of the OLS estimator of the parameters of an AR(p) model has been studied by, among others, Hurwicz (1950), Kendall (1954), Marriott and Pope (1954), White (1961), Shenton and Johnson (1965), Tjøstheim and Paulsen (1983), Tanaka (1984), Yamamoto and Kunimoto (1984), Shaman and Stine (1988) and Stine and Shaman (1989). In all these papers, asymptotic expansions are used to derive approximate formulas of the bias as a function of the sample size and of the true parameter values. The most comprehensive of these papers is Shaman and Stine (1988), who provide such formulas for all AR(p) models up to order 6. They also derive bias equations for the Yule-Walker estimator and find that it is more biased than OLS. On the other hand, Cheang and Reinsel (2000) consider the bias of the ML estimator of the parameters of an AR(p) error process in a linear regression. They find that this bias is larger than that of a restricted MLE, which is obtained by modifying the log likelihood function so as to obtain an unbiased score vector. They also provide simulation evidence to the effect that large biases lead to less accurate inference on the regression parameters. All these papers assume stationarity of the AR process. LeBreton and Pham (1989) and Abadir (1993) study OLS bias for non-stationary models, while Zhang and McLoed (2006) show that the bias of the Burg estimator coincides with that of OLS for AR(p) models with $p = 1, 2, 3$.

Less attention has been devoted to the characterisation of the bias of estimators for ARMA models. The first important attempt was made by Tanaka (1984) who, using Edgeworth expansions, derives bias expressions for some low order ARMA(p, q) models estimated by maximum likelihood. Another widely cited paper in the statistical literature is Cordeiro and Klein (1994), who derive bias expressions which coincide to those of Tanaka (1984) for the ML estimator. Their method differs from that of Tanaka (1984) in that they use the covariance matrix of the ARMA process to find the bias. Unfortunately, this requires the computation of the inverse of the covariance matrix, as well as its first and second derivatives. Accordingly, their bias equations, just like those of Tanaka (1984), become very complicated as the order of the process increases.

Several bias corrected estimators for AR(p) models have been proposed. They either aim at providing mean or median-unbiased estimators. The difference between the two is that in the former case, the expectation of the estimator equals the true parameter value while in the latter, the median does. We will define these concepts more precisely in section 2. The first notable efforts in this direction are the jacknife estimator of Quenouille (1949, 1956) for AR(1) models, and the bias corrected estimator of Orcutt and Winokur (1969), which is based on the works of Mariott and Pope (1954) and Kendall (1954). Both these estimatiors, which will be discussed in section 2, are mean-unbiased through order $O(n^{-1})$. More recently, Rudebusch (1992), Andrews and Chen (1994), Fair (1996) and Fuller (1996) propose approximately median-unbiased estimators for AR(p) models. The last two papers are actually generalisations of the exact median-unbiased estimator proposed by Andrews (1993), which is itself based on a procedure first introduced by Lehmann (1959). Andrews' estimator will be discussed in section 2. Finally, Roy and Fuller (2001) provide a modified OLS estimator that is approximately median-unbiased for AR(1) parameter values 1 and -1.

Instead of modifying an existing estimator to make it unbiased, So and Shin (1999) propose the use of an estimator which is already approximately median-unbiased. This estimator, which is based on a method introduced by Cauchy (1836), uses the sign function of the first lag as an instrument for the estimation of the parameters of an AR(1) model. The paper of So and Shin (1999) is mostly concerned with the AR(1) case, with amphasis on hypothesis testing and confidence interval construction in the unit root case. Nevertheless, they provide a few results for more general AR(p) models. The analysis of the AR(1) case is extended by Phillips, Park and Chang (2001), who propose the use of other nonlinear functions of the data as instruments.

Bootstrap methods have also been devised to obtain unbiased estimators of AR(p) model parameters. The autoritative reference in this area is MacKinnon and Smith (1998), who show that one can use the bootstrap to obtain mean-unbiased estimators for general statistical models, including the AR(p) model. They briefly discuss how to extend their methods to obtain median-unbiased estimators. This idea was subsequently developped by Tanizaki (2000) and Tanizaki et al. (2005) for one of the estimators proposed by MacKinnon and Smith (1998). Other bootstrap-based bias corrected estimators were introduced by Killian (1998) and Hansen (1999). Since these last papers were mainly concerned with confidence intervals, we do not discuss them here.

Very little effort has been devoted to providing unbiased estimators specifically designed for MA(q) and ARMA(p, q) models. The expressions derived by Cordeiro and Klein (1994) certainly could be used to do this, but this idea does not seem to have been pursued anywhere in the econometrics and statistics literature. This may be due to their complexity for higher order models. The computational difficulties associated with the estimation of ARMA models compared to AR models may also be partly responsible. Finally, the fact that the bootstrap bias correction methods studied by MacKinnon and Smith (1998) are simple to implement may also explain this appearant lack of interest. The bias correction method we propose in this paper has the advantage of being even simpler to implement than MacKinnon and Smith (1998)'s estimators.

The rest of the paper is organised as follows. Section 2 provides the necessary theoretical background as well as some widely used bias correction techniques. It formally introduces several concepts such as the bias function, and the ideas of mean and median-unbiased estimators and discusses several of the bias corrected estimators mentionned in the present introduction. Section 3 introduces the GLS bias correction technique, establishes its properties and links it to the analytical indirect inference estimators of Galbraith and Zinde-Walsh (1994, 1997). We also discuss and compaire several estimation methods for ARMA models. Section 4 evaluates the finite sample performance of the GLS method and compaires it to several unbiased estimators. Section 5 concludes.

2 Bias correction

Consider a model characterised by a parameter θ which belongs to a parameter space D. Let $\hat{\theta}$ be an estimator of θ. Then, $\hat{\theta}$ is a mean-unbiased estimator of θ if and only if $\mathrm{E}_\theta(\hat{\theta}) = \theta$ for all $\theta \in$ D, where E_θ denotes the expectation when θ is the true parameter. On the other hand, $\hat{\theta}$ is a median-unbiased estimator of θ if and only if its median is θ for any $\theta \in$ D. An alternative definition for mean-unbiasedness is that

$$\mathrm{E}_\theta|\hat{\theta} - \theta| \leq \mathrm{E}_\theta|\hat{\theta} - \theta'|$$

for all θ and $\theta' \in$ D. This means that $\hat{\theta}$ is median unbiased if and only if it is, on average, closer to θ than to any other point in D. By definition, the fact that $\hat{\theta}$ is median-unbiased means that θ is the 50^{th} percentile of its distribution. This implies that there are equal probabilities that $\hat{\theta}$ over or under-estimates θ. This property is obviously desirable when the estimator's distribution is non-symmetric or has fat tails, for the sample median is a more robust measure of the center of the distribution than the mean. Median-unbiased estimators may also be useful when the parameter space is bounded and closed and the estimator is restricted to lie within it. Indeed, in such circumstances, mean-unbiased estimators do not exist because all estimators are mean-biased at the boundary points, see the discussion in Andrews (1993), p. 145.

Let us now suppose that $\hat{\theta}$ is estimated using a sample of n observations and let θ_0 be the true numerical value of θ. Following the notation in MacKinnon and Smith (1998, henceforth, MS), we write

$$\hat{\theta} = \theta_0 + b(\theta_0, n) + v(\theta_0, n) \tag{1}$$

where $v(\theta_0, n)$ is a mean 0 random disturbance and $b(\theta_0, n)$ is what MS call the bias function and is defined as $b(\theta_0, n) \equiv \mathrm{E}_{\theta_0}(\hat{\theta}) - \theta_0$. This formulation makes it clear that the bias of $\hat{\theta}$ is a function of the sample size, n, and of the true parameter value, θ_0, and that fixing one of these arguments allows one to plot the bias as a function of the other. Of course, for an estimator to be useful at all, it is necessary that its bias be a decreasing function of n, and we therefore expect the bias function of any consistent estimator to exhibit this characteristic for any value of θ_0. On the other hand, the form of $b(\theta_0, n)$ as a function of θ_0 can be just about anything, although some patterns tend to repeat themselves in certain classes of models. This feature makes bias correction a difficult task because, depending on whether $b(\theta_0, n)$ is a constant or a linear or non-linear function of θ_0, different bias correction methods should be used.

2.1 Bootstrap Bias Correction

We have mentioned in the introduction that the bootstrap may be used to correct the bias of a given estimator and even, in certain cases, to obtain mean or median-unbiased estimators. This essentially means that it can be used to estimate the bias function. In order to do this, and for a given sample size, it is necessary to provide the bootstrap algorithm with a value of θ. If it was known, then θ_0 would be the obvious choice. Since this is not the case, we must find another value of θ. It turns out that this choice is of critical importance in most cases of interest. In fact, the importance of this choice depends on the shape of $b(\theta_0, n)$ as a function of θ_0.

If $b(\theta_0, n)$ is a constant function throughout the parameter space, then the choice of θ at which we evaluate it is irrelevant. Indeed, the fact that $\mathrm{b}(\theta_0, n)$ is constant implies that we expect $\hat{\theta}$ to be biased in the same way whatever θ_0 really is. The simplest bootstrap bias correction method then results in what MS call the constant bias correcting (CBC) estimator. This simply consists of generating a large number (say, B) of bootstrap samples of size n from the model being studied and obtain an estimate of θ for each of them. This can be done using any values of θ, a common choice being $\hat{\theta}$. Because, in samples of reasonable size, $\hat{\theta}$ is generally not too far from θ_0, this choice has the advantage that the CBC still provides a fairly good bias correction even if the bias function is not a constant. If we denote the estimate of θ obtained from the j^{th} bootstrap sample as $\hat{\theta}_j$, then the bias can be estimated as:

$$\hat{b}(n) = \left[\frac{1}{B} \sum_{j=1}^{B} \hat{\theta}_j \right] - \hat{\theta} \tag{2}$$

where we have removed θ_0 from the bias expression to make explicit the fact that it does not depend on the true parameter value. The CBC estimator is therefore simply $\tilde{\theta} = \hat{\theta} - \hat{b}(n)$. As shown in MS, this estimator is mean-unbiased through order $O(n^{-1})$. A median-unbiased estimator is easily obtained by replacing $\frac{1}{B}\sum_{j=1}^{B} \hat{\theta}_j$ by the median of the set of $\hat{\theta}_j$s in the definition of b_n.

The CBC estimator is very commonly used in practice, no doubt because of its computational ease. It is however not often the best choice, for it is quite rare that an estimator has a flat bias function. Rather, it is more likely that $b(\theta_0, n)$ is linear or nonlinear. Both cases are investigated by MS. If the bias function is linear, then it can be written as follows:

$$b(\theta_0, n) = \alpha + \beta\theta_0. \tag{3}$$

Thus, if one knows the values of α and β, then one can evaluate the bias for any value of θ_0. In practice, these parameters are almost always unknown and must be estimated. Fortunately, this is a rather easy task which simply requires that we evaluate the function (3) at any two points. This can easily be accomplished by computing (2) at two points. What results is what MS call the linear bias correcting estimator (LBC):

$$\breve{\theta} = \frac{1}{1+\hat{\beta}}(\hat{\theta} - \hat{\alpha}) \tag{4}$$

where $\hat{\alpha}$ and $\hat{\beta}$ are the evaluated values of α and β. As shown by MS, the LBC estimator is exactly mean-unbiased whenever the bias function really is linear.

Once more, a median-unbiased estimator could be obtained by using the median instead of the average in equation (2).

The bias function may also be nonlinear. In such cases, MS show that the CBC and LBC are unbiased through order $O(n^{-1})$. They also introduce a nonlinear bias correcting (NBC) estimator. Rearranging (4), we find that $\breve{\theta}$ and $\hat{\theta}$ are related linearly by the function $\breve{\theta} = \hat{\theta} - \hat{\alpha} - \hat{\beta}\breve{\theta} = \hat{\theta} - b(\breve{\theta}, n)$. Let $\dot{\theta}$ be the NBC and define it such that $\dot{\theta} = \hat{\theta} - b(\dot{\theta})$, where we supress the dependence on n for convenience and the form of the bias function is now unknown. MS propose an iterative method to compute the NBC. First, compute an estimate of the bias at $\hat{\theta}$ using (2) and call it $b(\hat{\theta})$. Then, compute:

$$\dot{\theta}^{(j)} = (1-\gamma)\dot{\theta}^{(j-1)} + \gamma\left(\hat{\theta} - b(\dot{\theta}^{(j-1)})\right) \tag{5}$$

where $\dot{\theta}^{(0)} = \hat{\theta}$ and γ is a constant between 0 and 1. Evidently, if the iterations converge, then $\dot{\theta}^{(j)}$ will satisfy $\dot{\theta} = \hat{\theta} - b(\dot{\theta})$. As noted by MS, the sequence is less likely to converge for higher values of γ but its convergence will be faster if it occurs. Of course, γ need not remain constant through all iterations. For example, Tanizaki (2000) suggests $\gamma^j = 0.9^{j-1}$. MS show that the NBC and LBC estimators coincide to order $O(n^{-1})$, so that the NBC is approximately mean-unbiased. They also show that both estimators are mean-unbiased through order $O(n^{-2})$ if the second derivative of the bias function is 0 at θ_0. An exactly median-unbiased estimator could be obtained by replacing the average in (2) by the median, see MS for a discussion. This idea was applied by Tanizaki (2000), who surprisingly forgets to cite MS, and Tanizaki et al. (2005), who correct this rather important omission. We now review some of the less general bias-corrected estimators mentionned in the introduction.

2.2 Quenouille (1949, 1956)

Quenouille (1949, 1956) considers bias correction in the case of the AR(1) model. Consider the simple AR(1) model with no constant or trend:

$$y_t = \alpha y_{t-1} + \varepsilon_t \tag{6}$$

where $\varepsilon_t \sim iid(0, \sigma^2_\varepsilon)$. Quenouille's method relies on the observation that the bias of the OLS estimator is approximately proportional to n^{-1}. His bias corrected estimator is

$$\breve{\alpha} = 2\hat{\alpha} - \frac{(\hat{\alpha}' + \hat{\alpha}'')}{2} \tag{7}$$

where $\hat{\alpha}$ is the OLS estimator over the whole sample and $\hat{\alpha}'$ and $\hat{\alpha}''$ are the OLS estimators over the first and second half of the sample respectively. This estimator is approximately mean-unbiased.

2.3 Orcutt and Winokur (1969)

The second attempt at providing a bias corrected estimator for AR models is found in Orcutt and Winokur (1969), who restrict themselves to the stationary AR(1) case. Their work is based on Kendall (1954) and Mariott and Pope (1954) who show that the bias function of the OLS estimator of the parameter α in the AR(1) model with a constant is:

$$b(\alpha_0, n) = -\frac{1}{n}(1 + 3\alpha) + O(n^{-2})$$

Orcutt and Winokur (1969) therefore propose the approximately mean-unbiased estimator:

$$\breve{\alpha} = \frac{1}{n-3}(n\hat{\alpha} + 1). \qquad (8)$$

This estimator is, obviously, mean-unbiased to order $O(n^{-2})$. Notice that the approximating bias function is linear while the true bias function for $\hat{\alpha}$ is nonlinear (see the simulations below and MS). One would therefore expect this estimator to be very accurate in the regions of the parameter space where the bias function is almost linear (for α_0 between -0.75 and 0.75) and less accurate where the bias function is strongly curved. Orcutt and Winokur (1969) compaire their bias corrected estimator to that of Quenouille (1949, 1956) through a simulation experiment. They find that the estimators achieve comparable bias corrections but that Quenouille's estimator sometimes has a larger mean square error. We will not consider the Orcutt and Winokur (1969) bias corrected estimator in the simulations below because it only applies to AR(1) models with an unknown constant, which is a family of models we do not consider here. We will however propose a similar analytic bias correction procedure based on the equations of Tanaka (1984) and Cordeiro and Klein (1994).

2.4 Andrews (1993)

Andrews (1993) proposes an exactly median-unbiased estimator of the autoregressive parameter in AR(1) models with normally distributed errors. Contrarily to the previous two estimators, his is valid for unit root processes. Let $m(.)$ denote the median function. Then, Andrews (1993)'s estimator is:

$$\hat{\alpha}_U = \begin{cases} 1, & \text{if } \hat{\alpha} \geq m(1) \\ m^{-1}(\hat{\alpha}), & \text{if } m(-1) < \hat{\alpha} < m(1) \\ -1, & \text{if } \hat{\alpha} \leq m(-1). \end{cases} \tag{9}$$

where the inverse function $m^{-1}(.)$ satisfies $m^{-1}(m(\alpha)) = \alpha$. Thus, $\hat{\alpha}_U$ is the value of α for which $\hat{\alpha}$ is the median of the OLS estimator. $\hat{\alpha}_U$ is therefore median-unbiased (see Andrews, 2003, p. 145-46 for a proof). In the present paper, where we exclude unit root processes, $m(1)$ and $m(-1)$ are defined as $m(-1) = \lim_{\alpha \to -1} m(\alpha)$ and $m(1) = \lim_{\alpha \to 1} m(\alpha)$. For the same reason, we replace the strict inequality in Andrews (1993) by a $\geq$ in the first case. To compute $\hat{\alpha}_U$, we need to know the median of the OLS estimator $\hat{\alpha}$ for the relevant value of α and a given sample size. This may be done through simulations or using the tables provided by Andrews (1993). These tables are conditional on the normality of the errors asssumption, so that the resulting estimator is exactly median-unbiased only when the errors are indeed normal. Nevertheless, the simulations reported by Andrews (1993) indicate that the estimator's precision is quite robust to different error specifications. One case where it does not fare well is in the presence of thick tails. Note that similar tables could be built for different distributions, which would yield an exact estimator. This method is extended to the more general AR(p) case by Andrews and Chen (1994).

2.5 So and Shin (1999)

An approximately median-unbiased estimator for the sole parameter of an AR(1) process was proposed by So and Shin (1999). Their estimator, which is of the same form as the one introduced by Cauchy (1836), simply is:

$$\hat{\alpha}_c = ({y^s_{-1}}^\top y_{-1})^{-1} {y^s_{-1}}^\top y \tag{10}$$

where y_{-1} is the vector of the first lag of y and y^s_{-1} is a vector whose typical element is ${y^s_{-1}}_t = \text{sign}(y_{t-1})$, where $\text{sign}(y_{t-1}) = 1$ if $y_{t-1} \geq 0$ and $\text{sign}(y_{t-1}) = -1$ if $y_{t-1} < 0$. Thus, $\hat{\alpha}_c$ is an instrumental variable estimator of α with y^s_{-1} as an instrument. So and Shin (1999) use a Monte Carlo experiment to illustrate the fact that their estimator is median unbiased. According to this, the probability that their estimator underestimates α is approximately 50% for all values of α between 0.3 and 1.02. Another advantage of their estimator over OLS is that it has an asymptotic normal distribution when $|\alpha| = 1$. Phillips, Park and Chang (2001) suggest several alternatives to the sign function and discuss the properties of the resulting estimators.

2.6 Tanaka (1984) and Cordeiro and Klein (1994)

These two papers derive analytical expressions of the bias of ML estimators for some ARMA models through order $O(n^{-1})$. For the AR(1) model, the bias is $-n^{-1}2\alpha_1 + O(n^{-2})$, for the AR(2), the biases are $-n^{-1}\alpha_1 + O(n^{-2})$ and $-n^{-1}(1+3\alpha) + O(n^{-2})$. Therefore, the following mean-bias corrected estimators for the AR(1) model may be derived:

$$\breve{\alpha}_1 = \frac{n}{n-2}\hat{\alpha}$$

and for the AR(2) model:

$$\breve{\alpha}_1 = \frac{n}{n-1}\hat{\alpha}$$

$$\breve{\alpha}_2 = \frac{1}{n-3}(n\hat{\alpha}_2 + 1).$$

Similar bias-corrected estimators could be derived for higher order models. Now, consider an MA(q) model:

$$y_t = \theta_1\varepsilon_{t-1} + \ldots + \theta_q\varepsilon_{t-q} + \varepsilon_t$$

The bias derived by Tanaka (1984) and Cordeiro and Klien (1994) for the MA(1) model is $-n^{-1}\theta_1 + O(n^{-2})$ while the biases for the MA(2) are $-n^{-1}\theta_1 + O(n^{-2})$ and $-n^{-1}2\theta_2 + O(n^{-2})$. The following mean-unbiased estimator for the MA(1) can be derived:

$$\breve{\theta}_1 = \frac{n}{n-1}\hat{\theta}_1$$

and for the MA(2) model:

$$\breve{\theta}_1 = \frac{n}{n-1}\hat{\theta}_1$$

$$\breve{\theta}_2 = \frac{n}{n-2}\hat{\theta}_2.$$

Of course, similar bias-corrected estimators can be found for higher order MA models. Unfortunately, the bias expressions for ARMA models are very complicated and cannot easily be solved to yield a bias-corrected estimator. For example, even for the simple ARMA(1,1) model, the biases of the ML estimator of the AR and MA parameters involve second, third and fourth powers of $(1-\alpha_1\theta_1)$ and $(\alpha_1-\theta_1)$.

In comparaison, solving the equations necessary to obtain our GLS bias correction is very easy.

3 The GLS bias correction

We now show that it is possible to estimate the bias of any estimator in AR(p), MA(q) and ARMA(p,q) models using an analytical form of the GLS transformation matrix. We then use this bias estimator to define a bias corrected estimator of the model's parameters. Let $\boldsymbol{u}$ be a n-vector of observations generated by a stationary and invertible ARMA(p,q) process, that is, $u_t = \sum_{i=1}^{p} \alpha_i u_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t$, where ε_t is an *iid* innovation with mean 0 and variance σ_ε^2. Further, let $\boldsymbol{\Sigma}$ denote the covariance matrix of the vector $\boldsymbol{u}$. Then, the GLS transformation matrix $\boldsymbol{\Psi}$, which is a function of the parameters α_i and θ_i, is defined as a $n \times n$ matrix such that $\boldsymbol{\Psi}\boldsymbol{\Psi}^\top = \boldsymbol{\Sigma}^{-1}$. Evaluating $\boldsymbol{\Psi}^\top$ at the true parameter values and premultiplying it to the vector $\boldsymbol{u}$ yields a vector whose t^{th} element is ε_t. When the true parameter values are not known, $\boldsymbol{\Psi}$ can be evaluated using a set of parameter estimates. In this section, we investigate what happens when one uses a biased estimator.

There are several ways to build $\boldsymbol{\Psi}$. In all that follows, we use the estimator of Galbraith and Zinde-Walsh (1992). In this paper, the authors show that the transformation matrix $\boldsymbol{\Psi}$ can be constructed recursively, that is, the t^{th} row of $\boldsymbol{\Psi}^\top$ may be defined as a function of the $t-1$ first rows. In particular, for any stationary and invertible ARMA(p,q) process, they show that the element in position i,j of the lower triangular matrix $\boldsymbol{\Psi}^\top$ is

$$h_{i,j} = \begin{cases} 0, & \text{if } i < j \\ 1, & \text{if } i = j \\ -\sum_{k=1}^{\min\{i-1,q\}} \theta_k h_{i-k,j} - \alpha_{i-j}, & \text{otherwise.} \end{cases}$$

where θ_j and α_j denote the j^{th} MA and AR coefficient respectively. In practice, one replaces these unknown parameters by estimates obtained using some consistent method such as OLS or ML.

3.1 GLS bias correction for MA(q) models

Consider n observations of an invertible MA(q) process:

$$u_t = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q} + \varepsilon_t.$$

where the ε_ts are *iid* innovations with 0 mean and finite variance σ_ε^2. Let $\hat{\theta}$ be a consistent estimator of the q-vector of true parameters θ^0. Let θ_j^0 denote the true value of the j^{th} parameter in the model and assume that $\hat{\theta}$ is biased in small samples: $\mathrm{E}(\hat{\theta}_j) = \theta_j^0 + b_j$ where, as before, b_j is the bias of the j^{th} element of $\hat{\theta}$ and depends on n and θ^0. Let $\nu = \boldsymbol{\Psi}^\top(\mathrm{E}(\hat{\theta}))\boldsymbol{u}$ denote the $n \times 1$ vector of residuals obtained when $\boldsymbol{\Psi}$ is evaluated at the expected value of $\hat{\theta}$.

Observation 1.

Under standard regularity conditions and assuming that $\varepsilon_0 = \varepsilon_{-1} = ... = \varepsilon_{-q+1} = 0$, which is harmless asymptotically,

$$\varepsilon_t = \Phi(\theta^0, b, L)\nu_t$$

where L is the lag operator, Φ is an infinite order lag polynomial function of θ^0 and b, the vector of bias terms. More precisely,

$$\nu_t = \sum_{i=1}^{\infty} \gamma_i \nu_{t-i} + \varepsilon_t \tag{11}$$

where

$$\gamma_i = \sum_{k=1}^{\min\{i,q\}} -\gamma_{i-k}\theta_k^0 - b_i, \quad \gamma_0 = 0 \tag{12}$$

Proof.

Evaluating $\boldsymbol{\Psi}^\top$ at $\mathrm{E}(\hat{\theta})$ gives:

$$\boldsymbol{\Psi}^\top(\mathrm{E}(\hat{\theta})) = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots \\ -\theta_1^0 - b_1 & 1 & 0 & 0 & \cdots \\ (\theta_1^0 + b_1)^2 - \theta_2^0 - b_2 & -\theta_1^0 - b_1 & 1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \end{pmatrix}$$

or, algebraically, the element in position i, j is determined by the equation:

$$h_{i,j} = \begin{cases} 0, & \text{if } i < j \\ 1, & \text{if } i = j \\ -\sum_{k=1}^{\min\{i-1,q\}} (\theta_k^0 + b_k) h_{i-k,j}, & \text{otherwise.} \end{cases}$$

Then, we have the following equations, where we have suppressed the 0 superscript for ease of notation:

$$\nu_1 = \varepsilon_1 \tag{13}$$

$$\nu_2 = -b_1\varepsilon_1 + \varepsilon_2 \tag{14}$$

$$\nu_3 = (b_1^2 + \theta_1 b_1 - b_2)\varepsilon_1 - b_1\varepsilon_2 + \varepsilon_3 \tag{15}$$

$$\nu_4 = (\theta_1 b_2 + \theta_2 b_1 + 2b_1 b_2 - b_1^3 - 2\theta_1 b_1^2 - \theta_1^2 b_1 - b_3)\varepsilon_1 + (b_1^2 + \theta_1 b_1 - b_2)\varepsilon_2 - b_1\varepsilon_3 + \varepsilon_4 \tag{16}$$

$$\begin{aligned}\nu_5 = {} & (\theta_1^3 b_1 + 3\theta_1 b_1^3 + 3\theta_1^2 b_1^2 + b_1^4 - \theta_1^2 b_2 - 2\theta_1\theta_2 b_1 - 4\theta_1 b_1 b_2 - 2\theta_2 b_1^2 - 3b_1^2 b_2 + \theta_1 b_3 + \theta_3 b_1 \\ & + 2b_1 b_3 + \theta_2 b_2 + b_2^2 - b_4)\varepsilon_1 + \left(-\theta_1^2 b_1 - 2\theta_1 b_1^2 - b_1^3 + \theta_1 b_2 + \theta_2 b_1 + 2b_1 b_2 - b_3\right)\varepsilon_2 \\ & + \left(b_1^2 + \theta_1 b_1 - b_2\right)\varepsilon_3 - b_1\varepsilon_4 + \varepsilon_5\end{aligned} \tag{17}$$

If we solve equations (13), (14), (15) and (16) for ε_1, ε_2, ε_3 and ε_4 and substitute them in equation (17), we obtain:

$$\begin{aligned}\nu_5 = {} & -b_1\nu_4 + (\theta_1 b_1 - b_2)\,\nu_3 + \left(-\theta_1^2 b_1 + \theta_1 b_2 + \theta_2 b_1 - b_3\right)\nu_2 \\ & + \left(\theta_1^3 b_1 - \theta_1^2 b_2 - 2\theta_1\theta_2 b_1 + \theta_1 b_3 + \theta_3 b_1 + \theta_2 b_2\right)\nu_1 + \varepsilon_5\end{aligned} \tag{18}$$

which has the expected form. Obviously, generalizing this expression by further substitutions yields the stated result. •

The bias equations (12) are generalisations of the equations used by Galbraith and Zinde-Walsh (1994) to develop their analytical indirect inference estimator of MA parameters through the fitting of a long autoregression to the data. This can be seen by placing an original estimator such that $\mathrm{E}(\hat{\theta}_i) = 0$ for all i in the GLS transformation matrix so that $b_i = -\theta_i^0$ for all i. Hence, using these equations to estimate the bias terms can be considered as applying GZW (1994)'s method to the residuals obtained from a first stage biased estimator. If this first stage estimator is itself obtained by analytical indirect inference, then estimating its bias through equations (12) could be considered as some sort of iteration of the method. However, any estimator at all can be used as an initial value of θ, even an inconsistent one, as long as it converges to a non-stochastic limit within the invertibility region. This last restriction is necessary because $\mathbf{\Psi}$ is only valid for invertible processes.

There are several possible ways one could use equations (12) to obtain bias corrected estimators. In the case of MA models, we propose to estimate the bias of each parameter one at a time and to define the bias corrected estimator in the following recursive way:

1. Use the initial estimator to obtain a vector of filtered data: $\hat{\nu} = \mathbf{\Psi}^{\top}(\hat{\theta})\boldsymbol{u}$. Because of the properties of $\hat{\theta}$, this vector is expected to have the autoregressive structure identified in observation 1.

2. Fit a long autoregression to $\hat{\nu}_t$. Then, estimate the bias of $\hat{\theta}_1$ as $\hat{b}_1 = -\hat{\gamma}_1$ and compute the bias corrected estimator, which we define as $\tilde{\theta}_1 \equiv \hat{\theta}_1 - \hat{b}_1$.

3. Estimate b_2 as $\hat{b}_2 = -\hat{\gamma}_1 - \hat{\gamma}_2\tilde{\theta}_1$. It is preferable to use the bias corrected estimate of θ_1 instead of $\hat{\theta}_1$ because it is likely to be closer to θ_1^0 than the original estimate, so that the estimate of $\hat{b}_2$ should be more precise. Compute the bias corrected estimator of θ_2, which we define as $\tilde{\theta}_2 \equiv \hat{\theta}_2 - \hat{b}_2$.

4. Use steps similar to 2 to get bias corrected estimates of any other parameter. That is, compute the bias corrected estimators $\tilde{\theta}_j \equiv \hat{\theta}_j - \hat{b}_j$ where $\hat{b}_j = -\hat{\gamma}_i - \sum_{j=1}^{\min\{i,q\}} \hat{\gamma}_{i-j}\tilde{\theta}_j$ for $j = 3, ..., q$.

This bias correction scheme can be iterated by using $\tilde{\theta}$ in the GLS transformation matrix so as to obtain a new vector of filtered data $\tilde{\nu} = \mathbf{\Psi}^{\top}(\tilde{\theta})\boldsymbol{u}$ and going through steps 2 to 4 with the new filtered data. We explore this possibility in the simulations below.

3.2 GLS bias correction for AR(p) models

Consider n observations of a stationary autoregression:

$$u_t = \alpha_1 u_{t-1} + ... + \alpha_p u_{t-p} + \varepsilon_t$$

and let $\hat{\alpha}$ be a biased estimator of the true parameter vector α^0 with typical element α_i^0 and let $\mathrm{E}(\hat{\alpha}_i) = \alpha_i^0 + b_i$. Finally, let $\nu = \mathbf{\Psi}^{\top}(\mathrm{E}(\hat{\alpha}))\boldsymbol{u}$ be the vector of filtered observations at the expected value of $\hat{\alpha}$.

Observation 2.

Under the same conditions as in observation 1 plus that $u_{t-j} = 0$ for all $0 < j \leq p$,

$$\varepsilon_t = \Theta(\alpha^0, b, L)\nu_t$$

where Θ is an infinite order lag polynomial function of α^0, the vector of true parameters of the AR process and b, the vector of bias terms and L is the lag operator. More precisely,

$$\nu_t = \sum_{i=1}^{\infty} \gamma_i \nu_{t-i} + \varepsilon_t \tag{19}$$

where

$$\gamma_j = \sum_{i=1}^{\min(j,p)} (\alpha_i^0 + b_i)\gamma_{j-i} - b_j, \quad \gamma_0 = 0 \tag{20}$$

Proof.

Evaluating $\mathbf{\Psi}^\top$ at $\mathrm{E}(\hat{\alpha})$ yields:

$$\mathbf{\Psi}^\top(\mathrm{E}(\hat{\alpha})) = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots \\ -\alpha_1^0 - b_1 & 1 & 0 & 0 & \cdots \\ -\alpha_2^0 - b_2 & -\alpha_1^0 - b_1 & 1 & 0 & \cdots \\ -\alpha_3^0 - b_3 & -\alpha_2^0 - b_2 & -\alpha_1^0 - b_1 & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \end{pmatrix}$$

Then, we get the following equations:

$$\nu_1 = \varepsilon_1 \tag{21}$$

$$\nu_2 = -b_1\varepsilon_1 + \varepsilon_2 \tag{22}$$

$$\nu_3 = -(b_2 + b_1\alpha_1)\varepsilon_1 - b_1\varepsilon_2 + \varepsilon_3 \tag{23}$$

$$\nu_4 = -(b_3 + b_2\alpha_1 + b_1\alpha_1^2 + b_1\alpha_2)\varepsilon_1 - (b_2 + b_1\alpha_1)\varepsilon_2 - b_1\varepsilon_3 + \varepsilon_4. \tag{24}$$

Estimating a moving average model to correct the bias of an AR(p) model may seem cumbersome and we may wish to avoid this. Substituting (21), (22) and (23) in (24), we obtain:

$$\begin{aligned}\nu_4 = &-(b_3 + b_2\alpha_1 + b_1\alpha_1^2 + b_1\alpha_2 + 2b_1b_2 + 2b_1^2\alpha_1 + b_1^3)\nu_1 \\ &-(b_2 + b_1\alpha_1 + b_1^2)\nu_2 - b_1\nu_3 + \varepsilon_4.\end{aligned}$$

It is easy to see that the coefficients of this autoregression have the form given in equations (20). •

Based on the results of observation 2, we can obtain a bias-corrected estimator in much the same way as we did for MA(q) models. We first need to compute $\hat{\nu} = \mathbf{\Psi}^\top(\hat{\alpha})\boldsymbol{u}$. Then, we can estimate the bias terms by fitting a long AR(k) to $\hat{\nu}_t$ and following steps similar to those described in observation 1. The bias corrected estimator is then defined as $\tilde{\alpha}_i = \hat{\alpha}_i - \hat{b}_i$, where $\hat{b}_i$ is the bias estimate. Of course, this can be iterated.

3.3 GLS bias correction for ARMA(p, q) models

It is easy to extend the results of observations 1 and 2 to find a similar result for ARMA(p, q) models. Let us consider the following process:

$$u_t = \alpha_1 u_{t-1} + ... + \alpha_p u_{t-p} + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

which we assume to be invertible and stationary. Let $\hat{\theta}$ and $\hat{\alpha}$ be biased estimators defined as above. Then, it can be shown that the process $\nu = \mathbf{\Psi}^\top(\mathrm{E}(\hat{\theta}), \mathrm{E}(\hat{\alpha}))\boldsymbol{u}$ has an infinite autoregressive form with parameters functions of the true parameters and of the bias terms. Unfortunately, these functions are not simple and they involve products and squares of the bias terms and the true parameter values. For example, in the case of an ARMA(1,1), the first coefficient of this infinite AR process is equal to -$(b_\alpha + b_\theta)$, that is, minus the sum of the bias terms of the AR and MA parameters. The second coefficient is much more complicated: -$(\alpha b_\alpha - \theta b_\theta + b_\theta b_\alpha + b_\alpha^2)$. This yields the following expressions for the biases:

$$b_\alpha = \left[\frac{\gamma_2 - \theta\gamma_1}{\gamma_1 - \theta - \alpha}\right]$$

$$b_\theta = -\left[\frac{\gamma_2 - \theta\gamma_1}{\gamma_1 - \theta - \alpha}\right] - \gamma_1$$

where γ_1 and γ_2 are the first two parameters of the infinite autoregression. It is therefore possible to estimate the biases by replacing γ_1, γ_2, θ and α by consistent estimates. Since the bias terms are here functions of the ratio of several parameters which must be estimated, we should expect these bias estimates to have a high degree of variability. This is confirmed by our simulations. As we will now show, the resulting bias-corrected estimators are nevertheless consistent.

3.4 Properties of the bias-corrected estimator

In this section, we discuss the properties of the GLS bias-corrected estimator for MA(q) models. It is very easy to extend this discussion to the bias-corrected estimators for the AR(p) and ARMA(p, q) models. Let us define the vector of bias-corrected estimators for the parameters of a MA(q) model as $\tilde{\theta} \equiv \hat{\theta} - \hat{b}$, where $\tilde{\theta}$ has typical element $\hat{\theta}_i - \hat{b}_i$. We will now show that $\tilde{\theta}$ is a consistent estimator of θ but not an exactly unbiased one. We make the following assumptions:

Assumptions A

-u_t is an invertible MA(q) process.

-The order of the approximating autoregression used to estimate the bias increases with the sample size at the rate $o(n^{1/3})$.

Let us consider the expectation of $\tilde{\theta}$:

$$\begin{aligned} \mathrm{E}(\tilde{\theta}) &= \mathrm{E}(\hat{\theta}) - \mathrm{E}(\hat{b}) \\ &= \theta^0 + b - \mathrm{E}(\hat{b}). \end{aligned}$$

Thus, for $\tilde{\theta}$ to be mean-unbiased, it is necessary that $\hat{b}$ be an unbiased estimator of b. This is evidently not the case because each element of b appears as part of the coefficients in an infinite autoregression. Thus, $\hat{b}$ has to be estimated from a finite order approximation of this infinite order model. Hence, the regression model from which the elements of b are estimated is always underspecified and $\hat{b}$ consequently suffers from omitted variable bias. In fact, even if it were somehow possible to estimate the true AR(∞) regression (11), $\hat{b}$ would still be biased because the regressors in (11) are obviously not exogenous. A similar argument can be made for the AR(p) and ARMA(p, q) cases.

Suppose now that $\hat{\theta}$ is a consistent estimator of θ. One such estimator for MA(q) models is the simple estimator of GZW. Then, it is possible to show that our GLS bias corrected estimator is consistent. Indeed, we have:

$$\begin{aligned} \text{plim } \tilde{\theta} &= \text{plim } \hat{\theta} - \text{plim } \hat{b} \\ &= \theta^0 - \text{plim } \hat{b}. \end{aligned}$$

Consistency therefore follows if plim $\hat{b} = 0$. This follows if we show that the results of Berk (1974) can be applied to the approximating autoregression (11). Indeed, Berk shows that OLS estimators of the parameters of an AR(k) approximation of an AR(∞) model are consistent, provided that we let k increase at a proper rate of the sample size. The proof, which we present in the appendix, is quite simple, but we take the trouble of going through it because Berk (1974) considers infinite autoregressions with fixed coefficients, whereas the coefficients of regression (11) go to zero as $n \to \infty$ and are therefore not fixed as the sample size increases. Indeed, if we once more take the analytical indirect inference estimator of GZW (1994) as an example, then $\hat{b}$ can be shown to go to 0 as k and n increase. For example, GZW (1994) show that the asymptotic bias of the estimator of the sole parameter of an MA(1) model is of order $O(\theta^{2k+1})$, where θ is the true parameter value. Whenever $\theta \in (-1, 1)$, this means that the asymptotic bias goes to 0 as k and n go to infinity because k is a function of n. Thus, the result of Berk (1974) implies that plim $\hat{b} = 0$. This result can easily be extended to model (19) as well as to the bias corrected ARMA(p, q) parameter estimates. Hence, we conclude that the GLS bias corrected estimator is consistent.

It is shown in Berk (1974) and in Galbraith and Zinde-Walsh (2001) that the OLS estimator of the parameters of an AR(∞) model based on an AR(k) regression has a limiting normal distribution as $n \to \infty$ under assumptions A. Thus, the OLS estimator of the parameters γ_i in (11) has this property. This implies that the estimator of the first bias term, namely $\hat{b}_1$, is asymptotically normal. In turn, this means that $\tilde{\theta}_1$ is asymptotically normal because it is the sum of two independent asymptotically normal random variables.

The asymptotic normality of $\hat{b}_i$, $i = 2, 3, ...$ is also easy to establish. Consider as an example $\hat{b}_2 = \tilde{\theta}_1 \hat{b}_1 - \hat{\gamma}_2$. Using a first order Taylor expansion, we have:

$$\hat{b}_2 - b_2 \overset{a}{=} \theta_1 b_1 - \gamma_2 + \left(\tilde{\theta}_1 - \theta_1\right) b_1 + \left(\hat{b}_1 - b_1\right) \theta_1 - (\hat{\gamma}_2 - \gamma_2) - \theta_1 b_1 + \gamma_2.$$

Thus,

$$n^{1/2}\left(\hat{b}_2 - b_2\right) \overset{a}{=} n^{1/2}\left(\tilde{\theta}_1 - \theta_1\right) b_1 + n^{1/2}\left(\hat{b}_1 - b_1\right) - n^{1/2}\left(\hat{\gamma}_2 - \gamma_2\right).$$

Hence, if the joint asymptotic distribution of $n^{1/2}\left(\tilde{\theta}_1 - \theta_1\right)$, $n^{1/2}\left(\hat{b}_1 - b_1\right)$ and $n^{1/2}\left(\hat{\gamma}_2 - \gamma_2\right)$ is multivariate normal, which is so under usual assumptions, then $n^{1/2}\left(\hat{b}_2 - b_2\right)$ is asymptotically normal and so is $n^{1/2}\left(\tilde{\theta}_2 - \theta_2\right)$. A similar argument can be made for $\tilde{\theta}_j$, $j = 3, 4, ...$

4 Simulations

In this section, we use Monte Carlo simulation to assess the quality of the proposed bias correction method in small samples. For simplicity, and to keep this section as short as possible, we mainly study AR(1), MA(1) and ARMA(1,1) models.

We first consider simulated samples of a simple MA(1) model with uncorrelated N(0,1) error terms. The most common estimation technique for MA models is maximum likelihood but other estimators have been suggested. Among these is the analytical indirect inference estimator of Galbraith and Zinde-Walsh (1994) (GZW), where the MA parameters are deduced from the coefficients of an auxiliary AR(k) model through analytical binding functions. As GZW's simulations indicate, this estimator is likely to be less accurate than ML for processes close to the non-invertibility region. Also, the choice of specification of the alternative AR(k) model on which the indirect estimator is based may affect its accuracy. These facts are illustrated in the next two figures, which show the bias function of the ML and GZW estimators, the latter based on an AR(k) approximation with $k = 4$, 6, 8, 10 and 12 or chosen by the AIC with a minimum of 4 lags and a maximum of 10. These curves are based on 65 000 Monte Carlo samples.

[INSERT FIGURES 1 AND 2 ABOUT HERE]

Figure 1 shows that, interestingly enough, the GZW and MLE estimators are not biased in the same direction. The bias and MSE are generally much larger for the GZW than for the MLE, especially for extreme values of θ. A notable exception is when $k = 4$ in the middle of the parameter space. The figures also illustrate the main weakness of the GZW estimator, which is that its accuracy greatly depends on k. This feature is most clearly visible in figure 2, where a low value of k results in a high MSE in the extreme parts of the parameter space but a large one needlessly increases the MSE in the middle of the space.

Turning to the performances of the proposed GLS bias correction, and remembering that it is closely related to the GZW method, we first take a look at how well it works for this estimator. Figures 3 and 4 show the bias and MSE functions of the GZW estimator with $k = 8$ along with, in figure 3, the bias estimated by the GLS, iterated GLS (GLSIT), CBC and NBC methods and, in figure 4, the MSE of the bias corrected estimators. The results on the bootstrap methods were obtained from 5 000 Monte Carlo samples and 500 bootstrap repetitions while the other results are based on 25 000 Monte Carlo samples.

[INSERT FIGURES 3 AND 4 ABOUT HERE]

A common feature of all the bias correction methods is that they provide virtually mean-unbiased estimates for θ between -0.7 and 0.7 at the cost of a slightly higher MSE. On the other hand, for extreme values of θ, the GLS method provides lower MSE than the bootstrap and iterating the GLS correction appears to be useful. These results were expected because of the similar nature of the GZW estimator and GLS bias correction.

Figures 5 and 6 show the bias and MSE functions of the MLE estimator along with, in figure 5, the bias estimated by the GLS, iterated GLS and CBC methods and, in figure 6, the MSE of the bias corrected estimators. We have restricted our attention to the CBC estimator in order to minimise the computational cost of the experiment. The results for the MLE, GLS and GLSIT are based on 40 000 Monte Carlo samples while those on the bootstrap come from 1 500 Monte Carlo samples with 300 bootstrap repetitions each.

[INSERT FIGURES 5 AND 6 ABOUT HERE]

The GLS method is extremely accurate when θ is between -0.7 and 0.7. Indeed, figure 5 shows that it provides virtually unbiased estimation in that range, at only a very small cost in terms of MSE. The CBC is not as accurate in terms of mean estimated bias but it has somewhat smaller MSE. Things are quite different in the extreme regions of the parameter space. The GLS method completely misses the curves of the bias function, which results in a much higher MSE. The CBC, on the other hand, provides quite acceptable bias correction, even though it tends to undershoot the absolute value of the bias.

Based on our findings so far, it would appear to be a good idea to apply the iterated GLS bias correction whenever we use the GZW estimator since it very accurately estimates the bias and significantly reduces MSE for extreme values. Furthermore, based on the results of figure 4, the GLS method provides more accurate estimates than the bootstrap methods over the whole parameter space. On the other hand, if one uses ML the CBC estimator is preferable, unless one is willing to trade variability for mean-bias reduction.

Next, we consider the performances of our GLS method with only a few of the bias corrected estimators for AR(1) models introduced above, namely, the analytical bias correction of Orcutt and Winokur (1969), the CBC and NBC. The bootstrap results are based on 5 000 Monte Carlo samples with 500 bootstrap repetitions each while the others were computed with 100 000 Monte Carlo samples.

[INSERT FIGURES 7 AND 8 ABOUT HERE]

Once again, all the methods considered provide a very good approximation of the bias for moderate values of the parameter. In particular, iterating the GLS bias correction appears quite useful when $|\alpha|$ is around 0.75. Unfortunately, the GLS procedure does not behave well at all in the extreme regions of the parameter space, especially so when $|\alpha| > 0.9$. In spite of its fairly good results in most of the parameter space, there may not be much sense in using the GLS bias correction since the analytical correction and the bootstrap procedures always perform at least as well with much lower MSE.

In order to investigate bias correction for ARMA models, we have used a set of ARMA(1,1) models whith the AR and MA parameters always equal to each other ($\alpha = \theta$) and values of these parameters covering the stationarity and invertibility regions. Figures 9 and 10 show the bias functions of the ML estimator of these parameters, obtained from 10 000 Monte Carlo samples, along with the biases estimated by the GLS, GLSIT and CBC methods.

[INSERT FIGURES 9 AND 10 ABOUT HERE]

According to figures 9 and 10, the iterated GLS method provides acceptable estimates of the bias of the AR and MA components through most of the parameter space. In particular, it yields practically mean-unbiased estimates of the MA part for moderate parameter values. It is interesting to note that the bias functions of the ML estimator of α and θ are close to mirroring each others. This is especially remarquable at 0, where the estimator of θ is biased downward by approximately 0.08 while the estimator of α is biased upward by approximately 0.085. These two biases produce near-cancelling roots and the overall dynamics of the model estimated by ML is quite close to what it should be.

[INSERT FIGURE 11 ABOUT HERE]

Since this is a two parameter model, it is interesting to take a look at the MSE of the vector of estimated parameters. Figure 11 shows the determinant of the 2×2 MSE matrix of the different estimators of the vector $[\alpha, \theta]^\top$. All the bias correction techniques increase the MSE with the notable exception of the GLS at $\theta = \alpha = 0$. Since the CBC greatly increases the MSE around 0 and is comparable to the GLS elsewhere, it may be a good idea to use the latter. On the other hand, iterating the GLS is nowhere profitable in terms of MSE.

Finally, we briefly investigate the performance of the GLS bias correction for higher order MA models. For the sake of brevity, we only consider an MA(2) model. Figure 12 plots the determinant of the MSE matrix of the two parameters as estimated by MLE and by MLE with the GLS bias correction for various models. Those models are represented here by the modulus of their root closest to 1. It can be seen that the GLS correction increases the MSE for all models and more so when the modulus is close to 1. Figure 13 shows that the GLS bias estimates are quite accurate, except for modulus close to 1.

[INSERT FIGURES 12 AND 13 ABOUT HERE]

5 Conclusion

We introduce a method to estimate the bias of any estimator of the parameters of a low order ARMA(p,q) model based on the exact GLS transformation matrix introduced by Galbraith and Zinde-Walsh (1992). It is that it is very easy to compute, both in terms of programming and computational cost, and it may be seen as a natural extension of the analytical indirect inference estimator of Galbraith and Zinde-Walsh (1994). The resulting bias corrected estimator is consistent and has a limiting normal distribution. This GLS bias correction is particularly useful in the case of MA(1) models estimated by maximum likelihood where, according to our simulations, it provides a virtually mean-unbiased estimator for most of the parameter space. It also greatly improves the finite sample properties of the GZW estimator for MA(1) models throughout the parameter space.

Applying the GLS bias correction to AR(1) models allows one to obtain very accurate estimates of the OLS bias through most of the stationarity region. This, however, comes at the cost of substantially increased MSE. Our simulations suggest that usual bootstrap based bias correction is preferable. The GLS bias correction also works fairly well for ARMA(1,1) and MA(2) models.

References

Abadir, K. M. (1993). "The limiting distribution of the autocorrelation coefficient under a unit root," *The Annals of Statistics*, **21**, 1058-70.

Andrews, D. W. K. (1993). "Exactly median-unbiased estimation of first order autoregressive/ unit root models," *Econometrica*, **61**, 139-65.

Andrews, D. W. K. and H. Y. Chen (1994). "Approximately median-unbiased estimation of autoregressive models," *Journal of Business and Economic Statistics*, **12**, 187-204.

Berk, K. N. (1974). "Consistent autoregressive spectral estimates." *Annals of Statistics*, **2**, 489-502.

Cauchy, A. (1836). "On a formula for solving the problem of interpolation in a manner applicable to physical investigations," *Philosophical Magazine*, **8**, 459-68.

Cheang, K. W. and G. C. Reinsel (2000). "Bias reduction of autoregressive estimates in time series regression model through restricted maximum likelihood," *Journal of the American Statistical Association*, **95**, 1173-84.

Cordeiro, G. M. and R. Klein (1994). "Bias correction in ARMA models," *Statistics and Probability Letters*, **19**, 169-76.

Galbraith, J. W. and V. Zinde-Walsh (1992). "The GLS transformation matrix and a semi-recursive estimator for the linear regression model with ARMA errors," *Econometric Theory*, **8**, 143-55.

Galbraith, J. W. and V. Zinde-Walsh (1994). "A simple noniterative estimator for moving average models," *Biometrika*, **81**, 143-55.

Galbraith, J. W. and V. Zinde-Walsh (1997). "On some simple, autoregression-based estimation and identification techniques for ARMA models," *Biometrika*, **84**, 685-96.

Galbraith, J. W. and V. Zinde-Walsh (2001). "Analytical indirect inference," Working paper, McGill University.

Hansen, B. E. (1999). "The grid bootstrap and the autoregressive model," *The review of Economics and Statistics*, **81**, 594-607.

Hurvicz, L. (1950). "Least squares bias in time series," in . Koopmans (ed.) *Statistical inference in dynamic economic models*. Wiley, New-York, 365-83.

Kendall, M.G. (1954). "Note on the bias in the estimation of autocorrelation," *Biometrika*, **41**, 207-14.

Killian, L. (1998). "Small sample confidence intervals for impulse response functions," *The Review of Economics and Statistics*, **80**, 218-30.

LeBreton, E. L. and D. T. Pham (1989). "On the bias of the least squares estimator for the first order autoregressive process," *Annals of the Institute of Statistical Mathematics*, **41**, 555-63.

Lehmann, E. L. (1959). *Testing Statistical Hypothesis*, Wiley, New-York.

MacKinnon, J. G. and A. A. Smith (1998). "Approximate bias correction in econometrics," *Journal of Econometrics*, **85**, 205-30.

Mariott, F. H. C. and J. A. Pope (1954). "Bias in the estimation of autocorrelations," *Biometrika*, **41**, 393-403.

Orcutt, G. H. and H. S. Winokur (1969). "First order autoregression: inference, estimation, and prediction," *Econometrica*, **37**, 1-14.

Phillips, P. C. B., J. Y. Park and Y. Chang (2001). "Nonlinear instrumental variable estimation of an autoregression," Cowles Foundation Discussion Paper 1331.

Quenouille, M. H. (1949). "Approximate tests of correlation in time series," *Journal of the Royal Statistical Society Series B*, **11**, 68-84.

Quenouille, M. H. (1956). "Notes on bias in estimation," *Biometrika*, **43**, 353-60.

Rudebusch, G. (1992). "Trends and random walks in macroeconomic time series: a re-examination," *International Economic Review*, **33**, 661-80.

Roy, A. and W. A. Fuller (2001). "Estimation for autoregressive time series with a root near 1," *Journal of Business and Economics Statistics*, **19**, 482-93.

Shamman, P. and R. A. Stine (1988). "The bias of autoregressive coefficients estimators," *Journal of the American Statistical Association*, **83**, 842-88.

Shenton, L. R. and W. L. Johnson (1965). "Moments of a serial correlation coefficient," *Journal of the Royal Statistical Society Series B*, **27**, 308-20.

So, B. S. and D. W. Shin (1999). "Cauchy estimators for autoregressive processes with applications to unit root tests and confidence intervals," *Econometric Theory*, **15**, 165-76.

Stine, R. A. and P. Shamman (1989). "A fixed point characterisation for the bias of autoregressive estimators," *The Annals of Statistics*, **17**, 1275-84.

Tanaka, K. (1984). "An asymptotic expansion associated with the maximum likelihood estimators in ARMA models," *Journal of the Royal Statistical Society Series B*, **46**, 58-67.

Tanizaki, H. (2000). "Bias correction of OLSE in the regression model with lagged dependent variables," *Computational Statistics and Data Analysis*, **34**, 495-511.

Tanizaki, H., S. Hamori and Y. Matsubayashi (2005). "On least squares bias in the AR(p) models: bias correction using the bootstrap methods," *Statistical Papers*, **47**, 109-24.

Tjøstheim, D. and J. Paulsen (1983). "Bias of some commonly-used time series estimates," *Biometrika*, **70**, 389-99.

White, J. S. (1961). "Asymptotic expansions for the mean and variance of the serial correlation coefficient," *Biometrika*, **48**, 85-94.

Yamamoto, T. and N. Kunimoto (1984). "Asymptotic bias of the least squares estimator for multivariate autoregressive models," *Annals of the Institute of Statistical Mathematics*, **36**, 419-30.

Zhang, Y. and A. I. McLoed (2006). "Computer algebra derivation of the bias of linear estimators of autoregressive models," *Journal of Time Series Analysis*, **27**, 157-65.

Figure 1. Bias Functions, MA(1) Model.

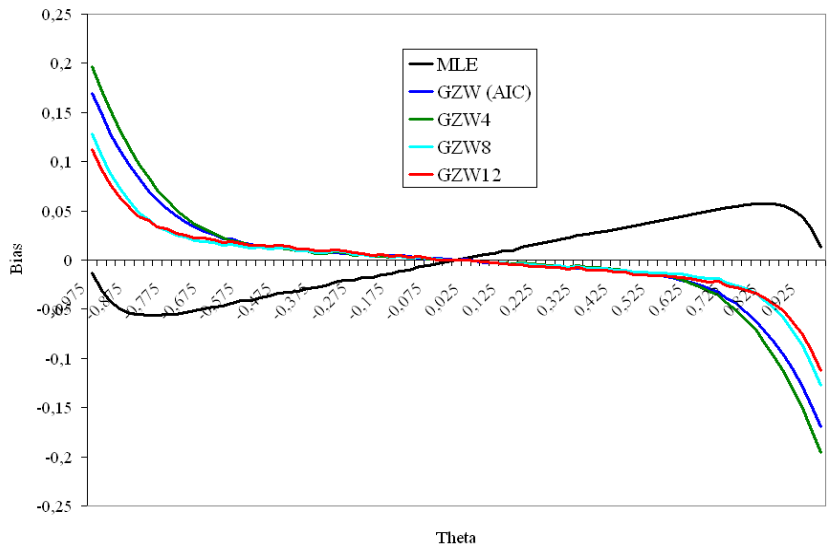


Figure 2. MSE Functions, MA(1) Model.

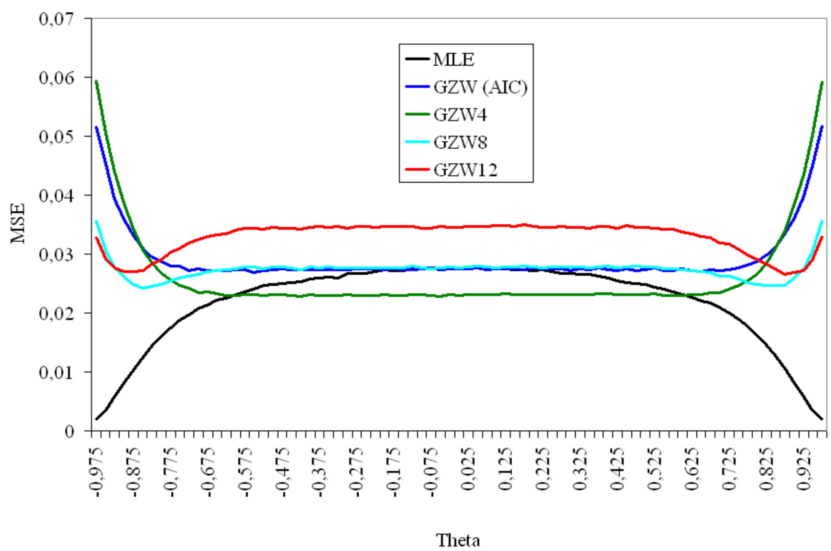


Figure 3. Bias Functions, GZW Estimator, MA(1) Model.

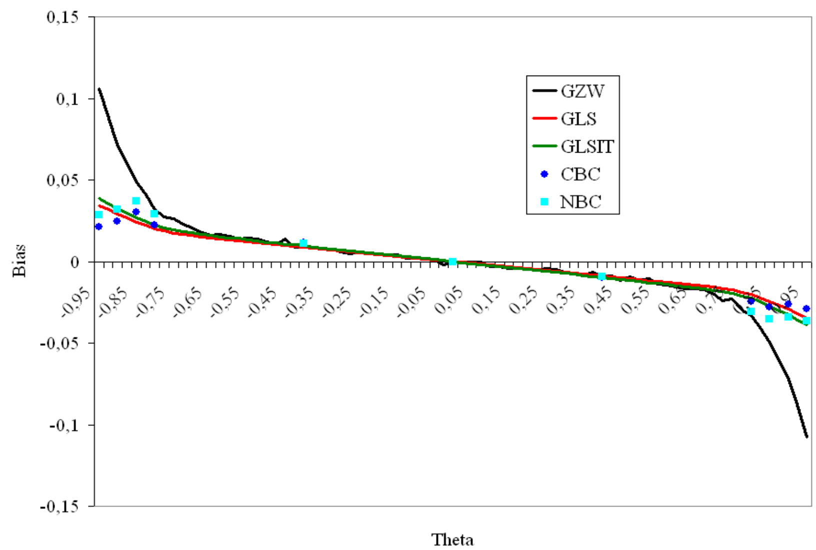


Figure 4. MSE Functions, GZW Estimator, MA(1) Model.

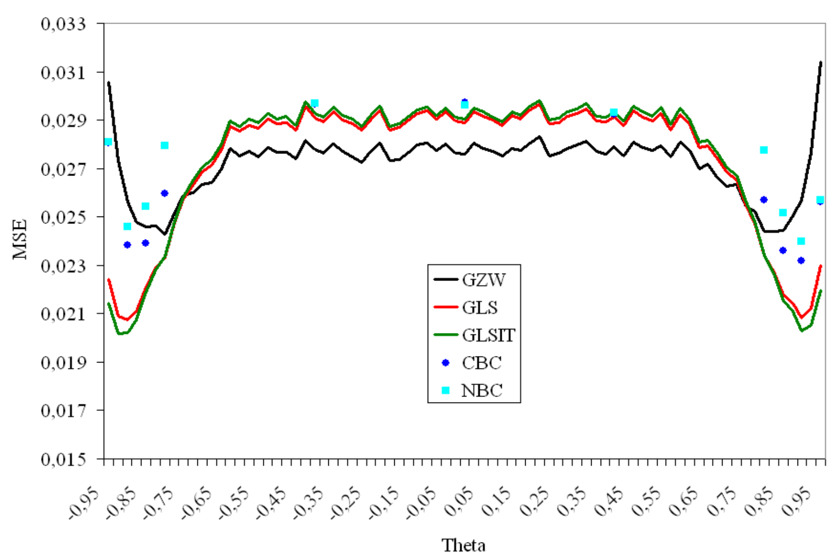


Figure 5. Bias Functions, ML Estimator, MA(1) Model.

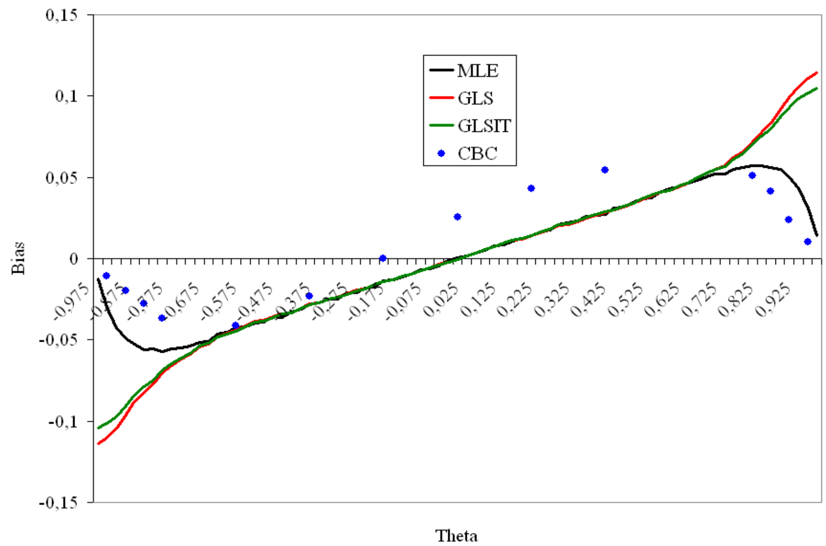


Figure 6. MSE Functions, ML Estimator, MA(1) Model.

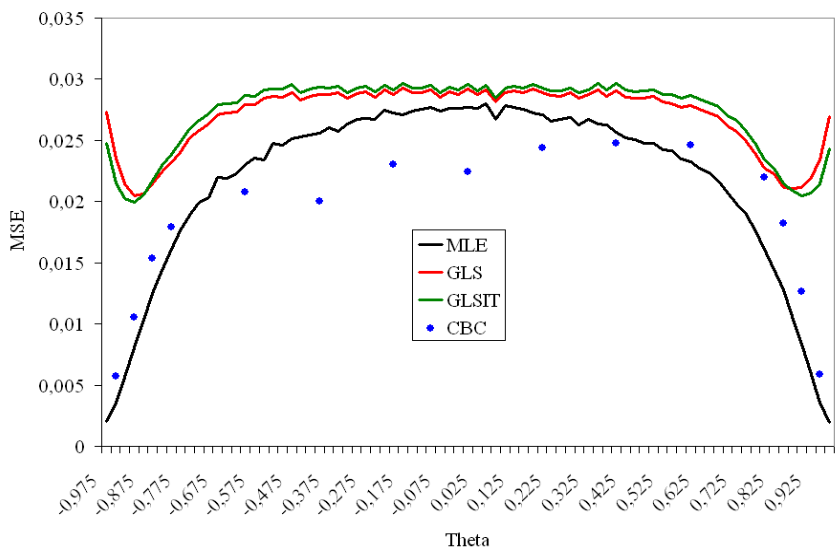


Figure 7. Bias Functions, OLS Estimator, AR(1) Model.

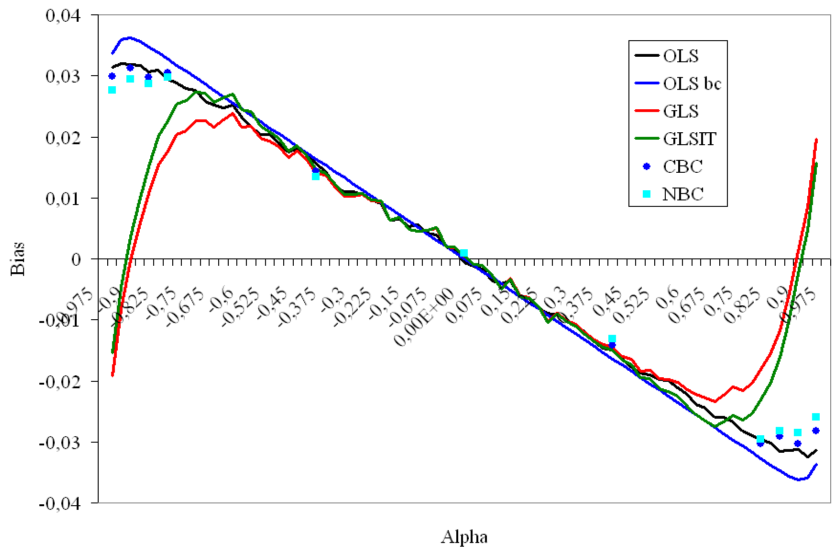


Figure 8. MSE Functions, OLS Estimator, AR(1) Model.

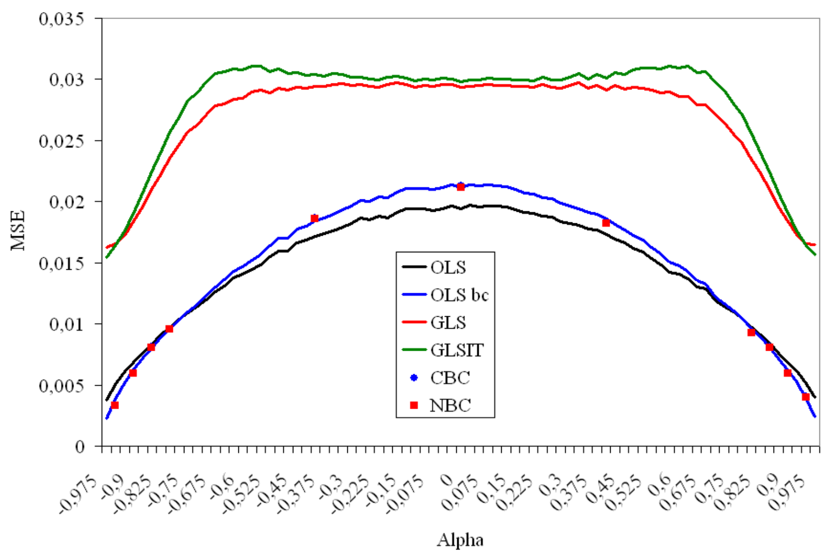


Figure 9. Bias Functions, ML Estimator, ARMA(1,1) Model.

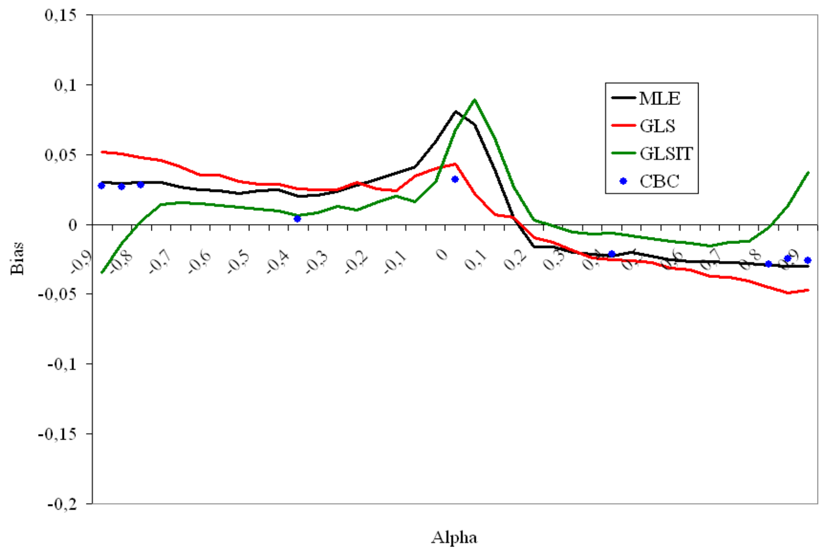


Figure 10. Bias Functions, ML Estimator, ARMA(1,1) Model.

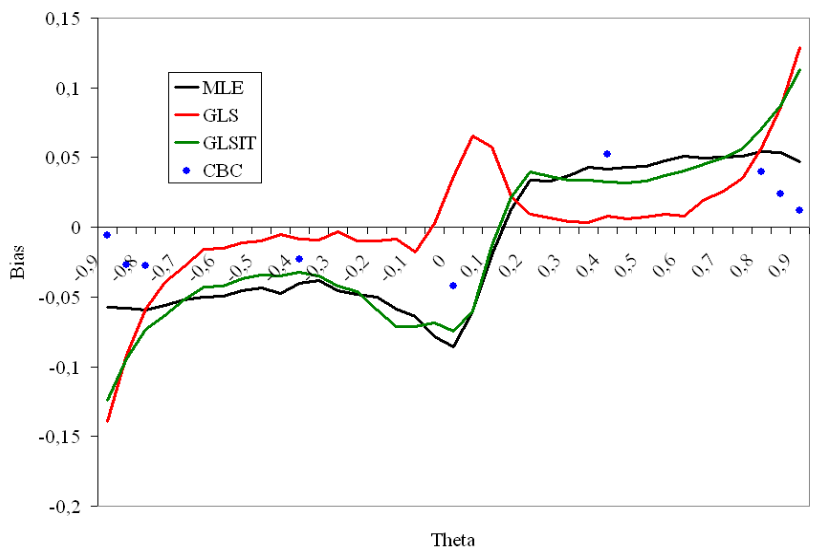


Figure 11. Det(MSE) Functions, ML Estimator, ARMA(1,1) Model.

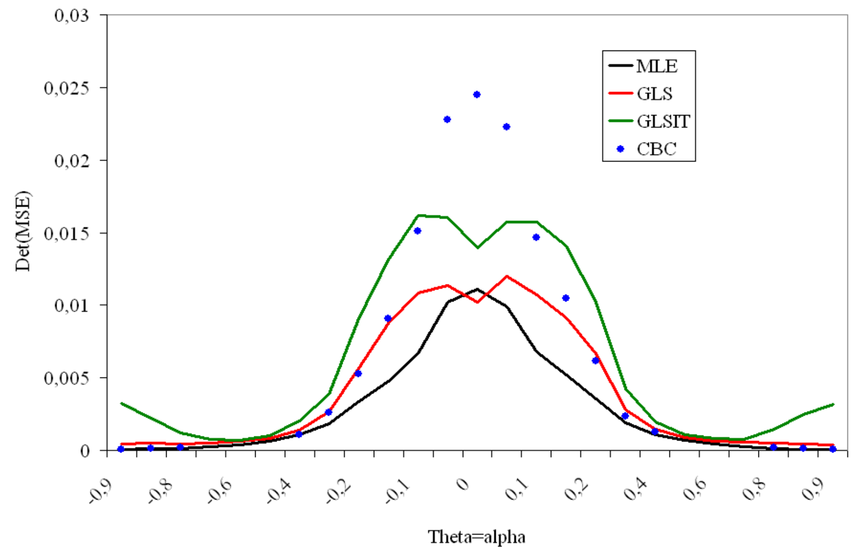


Figure 12. Det(MSE) Functions, ML Estimator, MA(2) Model.

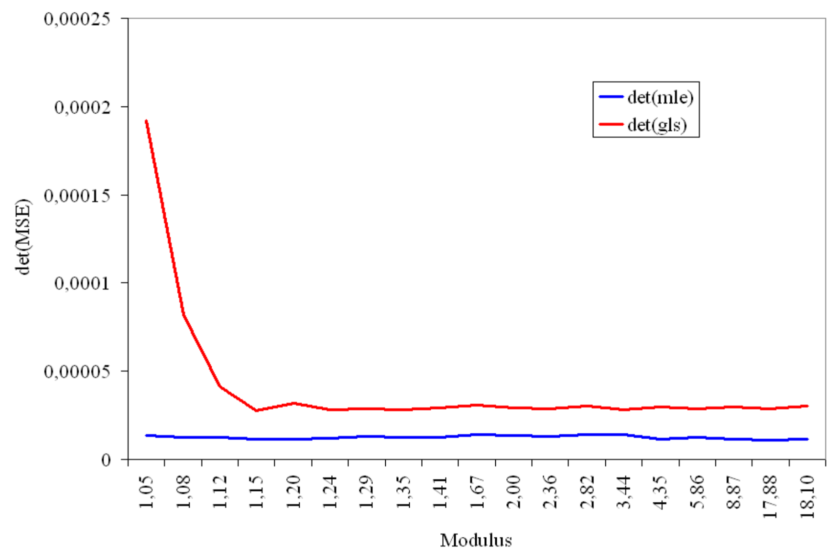
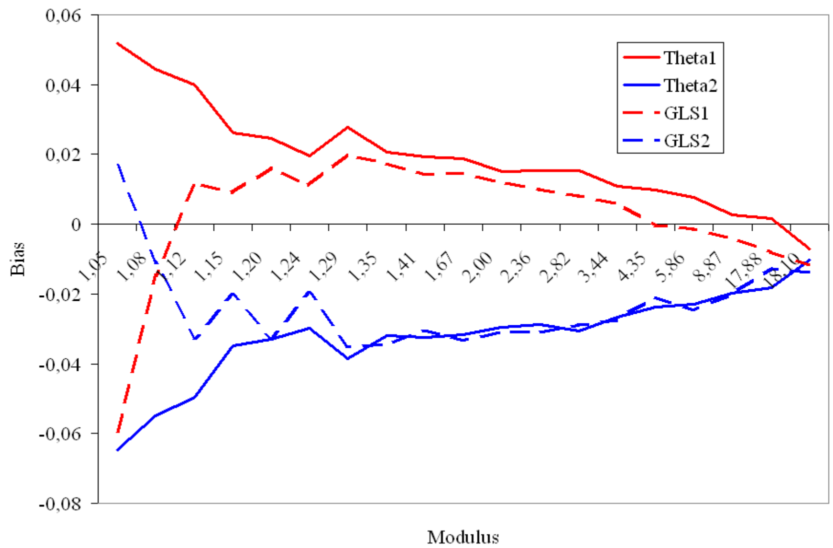


Figure 13. Bias Functions, ML Estimator, MA(2) Model.



A Proof of Consistency of GLS bias corrected estimator

Theorem 1, Berk, equation 2.17, p. 493.)

Under the assumptions specified above and letting $\mathbf{a}(k)$ denote the k-vector of parameters of the long autoregression used to estimate the bias terms,

$$\|\hat{\mathbf{a}}(k) - \mathbf{a}(k)\| \xrightarrow{p} 0.$$

Proof.

In the notation of Berk, his equation 2,8 becomes:

$$\mathbf{a}(k) - \hat{\mathbf{a}}(k) = \hat{\mathbf{R}}(k)^{-1} \sum_{j=k}^{n-1} \mathbf{X}_j(k)\varepsilon_{j+1,k}/(n-k)$$

where $\varepsilon_{j+1,k} = \nu_{j+1} - a_1\nu_j - ... - a_k\nu_{j-k+1}$. Then, adding and subtracting $\mathbf{R}(k)^{-1} \sum_{j=k}^{n-1} \mathbf{X}_j(k)$ $\varepsilon_{j+1,k}/(n-k)$ and $\mathbf{R}(k)^{-1} \sum_{j=k}^{n-1} \mathbf{X}_j(k)\varepsilon_{j+1}/(n-k)$ to the right hand side of this equation and using a triangle inequality yields:

$$\|\hat{\mathbf{a}}(k) - \mathbf{a}(k)\| \leq \left\|\hat{\mathbf{R}}(k)^{-1} - \mathbf{R}(k)\right\| \left\|\sum_{j=k}^{n-1} \mathbf{X}_j(k)\varepsilon_{j+1,k}/(n-k)\right\| \tag{25}$$

$$+ \left\|\mathbf{R}(k)^{-1}\right\| \left\|\sum_{j=k}^{n-1} \mathbf{X}_j(k)\left(\varepsilon_{j+1,k} - \varepsilon_{j+1}\right)/(n-k)\right\| + \left\|\mathbf{R}(k)^{-1}\right\| \left\|\sum_{j=k}^{n-1} \mathbf{X}_j(k)\varepsilon_{j+1}/(n-k)\right\|$$

which is equation (2.17, p. 493) in Berk. Under the assumption that $k = o(n^{1/3})$, Berk shows that $\mathbf{R}(k)^{-1}$ is bounded (equation 2.14, p. 493), and that $k^{1/2}\left\|\hat{\mathbf{R}}(k)^{-1} - \mathbf{R}(k)\right\|$ goes to 0 in probability (lemma 3, p. 493). He also shows that $\mathrm{E}\left\|\sum_{j=k}^{n-1} \mathbf{X}_j(k)\varepsilon_{j+1}/(n-k)\right\|^2$ is finite and that

$$\mathrm{E}\left\|\sum_{j=k}^{n-1} \mathbf{X}_j(k)\left(\varepsilon_{j+1,k} - \varepsilon_{j+1}\right)/(n-k)\right\| \leq \kappa k\left(a_{k+1}^2 + a_{k+2}^2 + ...\right)$$

where κ is a constant. Thus, the convergence of $\hat{a}(k)$ depends on whether $k\left(a_{k+1}^2 + a_{k+2}^2 + ...\right)$ goes to 0. This is obviously the case here because the bias terms on the original estimates vanish asymptotically.

•