RT-MAE-9314

ON RESIDUAL VARIANCE ESTIMATION IN ARMA MODELS

by

R.P. Mentz, P.A. Morettin

and

C.M.C. Toloi

Palavras Chaves: ARMA models, bias, least squares estimator, maximum likelihood estimator, moment estimator, residual variance, time series.
(Key words)

Classificação AMS: 62M10
(AMS Classification)

- Julho de 1993 -

ON RESIDUAL VARIANCE ESTIMATION IN ARMA MODELS

[1]R.P. Mentz, [2]P.A. Morettin and [2]C.M.C. Toloi

[1]University of Tucumán, Argentina and [2]University of São Paulo,Brazil

ABSTRACT

In this paper we consider time series models belonging to the ARMA familiy and deal with the estimation of the residual variance. This is important because estimates of the variance enter, for example, into confidence sets for the parameters of the model, in the estimation of the spectrum and in expressions for the estimated error of prediction.We consider the asymptotic biases for moment,least squares and maximum likelihood estimators of the residual variance for some simple,but useful models.Simulation results are also presented.

Key words: ARMA models,bias,least squares estimator, maximum likelihood estimator, moment estimator,residual variance,time series.

1. Introduction.

We consider time series models belonging to the ARMA(p,q) family, defined as follows: The observable stochastic process $\{X_t\}$ satisfies

$$\sum_{j=0}^{p} \beta_j (X_{t-j} - \mu) = \sum_{k=0}^{q} \alpha_k a_{t-k}, \quad t = \ldots, -1, 0, 1, \ldots, \tag{1.1}$$

where $\mu, \beta_0 = 1, \beta_1, \ldots, \beta_p, \alpha_0 = 1, \alpha_1, \ldots, \alpha_q$ are real parameters and the process $\{a_t\}$ is white noise with variance $0 < \sigma_a^2 < \infty$ .For stationarity we require that the roots of the polynomial equation

$$\sum_{j=0}^{p} \beta_j z^{p-j} = 0 \tag{1.2}$$

satisfy $|z_j| < 1, j = 1, \ldots, p$; this also gives invertibility of (1.1) into an infinite moving average(the model is causal). No restriction on the α_k's is needed for stationarity, but if the roots of

$$\sum_{k=0}^{q} \alpha_k w^{q-k} = 0 \tag{1.3}$$

satisfy $|w_k| < 1, k = 1, \ldots, q$, then (1.1) is invertible into an infinite autoregression.

In this paper we consider the estimation of the parameter σ_a^2 . This is important because estimates of σ_a^2 enter, for example, into confidence sets for the parameters, in the estimation of the spectrum and in expressions for the estimated error of prediction.

Estimates of σ_a^2 come from the methods of moments(MM), least squares(LS) or maximum likelihood(ML) under normality, and also from frequency domain arguments. In spite of its inferential role, not many papers have been written about the determination of (large-sample) biases of the estimators of the variance in ARMA models.

The main purposes of this paper are to review the literature on the subject and present some new material.

2. Review of the Literature

Some of the papers deal with the estimation of coefficients of the model (1.1) only, some deal with the estimation of the residual variance and some with both. In this section we comment briefly on each paper and present the main results in sections 3, 4 and 5.

The bias for Yule-Walker(YW)(or moment) and least squares estimators, for univariate and multivariate autoregressive(AR) processes was considered by Tjøstheim and Paulsen(1983). They obtain explicit formulas for the large-sample bias of YW estimators in the scalar AR(1) and AR(2) cases and for LS estimators in the general case. Both theory and simulations indicate that YW estimators are inferior to LS estimators. For strongly autocorrelated processes, YW estimators can be severely biased even for comparatively large sample sizes.

In a recent paper Shaman and Stine(1988) present simple expressions for the biases of estimators of coefficients of an AR(p) model, p considered arbitrary but known. Results include models both with and without a constant term. Emphasis is on YW and LS estimators. An easily programmed algorithm for generating these expressions is given, for any AR(p) process and for LS estimators. The same kind of conclusion arises:YW estimators tend to have greater biases than LS estimators.

Paulsen and Tjøstheim(1985) consider the bias for the residual variance of AR(p) models for YW, LS and Burg-type estimators. The same conclusion that they obtained for the estimators of the coefficients of such models hold here.An explicit expression for the bias of the estimator of σ_a^2 in the case of YW is given for the AR(2) model, in terms of the roots of the associated equation (1.2) above. We derive an expression for this bias in terms of the coefficients(see formula (3.15)).

Ansley and Newbold(1981) treat the case of bias in estimators of forecast mean-square error; for one-step ahead forecasts one obtains the bias of the estimate of σ_a^2 . It was found that when LS estimators were used to estimate model parameters, bias in the resulting estimates of the residual variance was typically more severe than when using ML estimators.

Yamamoto and Kunitomo(1984) derive asymptotic bias for LS estimators for multivariate AR models, with a constant term, up to order T^{-1} ,where T is the sample size.Tuan(1992) obtains the exact and asymptotic biases for LS and forward-backward LS estimators of the coefficients and the variance of the bias for the AR(1) model.

Tanaka(1984) suggests a technique for obtaining the Edgeworth type asymptotic expansion associated with ML estimators in ARMA models. He obtains bias up to order T^{-1} for AR(1), AR(2),MA(1),MA(2) and ARMA(1,1) models with and without constant terms. Bias for the residual variance are also derived.

Cordeiro and Klein(1993) present a general procedure to obtain the bias of ML estimators in ARMA models. It turns out that the formula is difficult to obtain for models other than the lower order ones,but the authors state that the formula can be easily computed numerically.

The estimation of the innovation variance through frequency domain considerations was considered by Pukkila and Nyquist(1985). This will not be considered here.

Further references are Marriott and Pope(1954), Walker(1962),Bhansali(1981), Kunitomo and Yamamoto(1985), Lysne and Tjøstheim(1987),Stine and Shaman(1990).

Good references of techniques and results for asymptotic analysis in time series models are Anderson(1971) and Fuller(1976).

3. Autoregressive Models

The AR(p) model can be written as

$$\sum_{j=0}^{p} \beta_j (X_{t-j} - \mu) = a_t , \quad t = \dots, -1, 0, 1, \dots \tag{3.1}$$

with $\beta_0 = 1$. The autocovariance sequence is

$$\gamma_s = E(X_t - \mu)(X_{t+s} - \mu) , \quad s = 0, \pm 1, \pm 2, \dots \tag{3.2}$$

and satisfies the (theoretical) Yule-Walker equations

$$\sum_{j=0}^{p} \beta_j \gamma_j = \sigma_a^2 , \tag{3.3}$$

and

$$\sum_{j=0}^{p} \beta_j \gamma_{j-s} = 0 , \quad s = 1, 2, \dots . \tag{3.4}$$

It also satisfies the "inversion formula "

$$\gamma_s = \int_{-\pi}^{+\pi} e^{i\lambda s} f(\lambda) d\lambda, \tag{3.5}$$

where $f(\lambda)$ is the spectral density of the process,

$$f(\lambda) = \frac{\sigma_a^2}{2\pi} \left| \sum_{j=0}^{p} \beta_j e^{i\lambda j} \right|^{-2}, \qquad -\pi \le \lambda \le \pi . \tag{3.6}$$

The correlation sequence is $\rho_s = \gamma_s / \gamma_0$, $s = 0, \pm 1, \dots$ and for purposes of inference we have a sample $X_1, \dots, X_T$ from (3.1).

We shall denote by $\hat{\sigma}_{Rk}^2$ $k = 1, 2 \dots$ the various estimators that we consider in the case of AR(p) models and by $\hat{\sigma}_{Rk_c}^2$ the "corrected "estimators,to be defined below.

3.1. Moment (Yule-Walker) Estimators

A Yule-Walker(YW) or moment estimator, $\hat{\sigma}_{R1}^2$, is provided by the sample analog of (3.3), namely

$$\hat{\sigma}_{R1}^2 = \sum_{j=0}^{p} \hat{\beta}_j^{(1)} c_j , \tag{3.7}$$

where

$$c_j = \frac{1}{T} \sum_{t=1}^{T-j} (X_t - \bar{X})(X_{t+j} - \bar{X}) . \tag{3.8}$$

and $\hat{\mu}_1 = \bar{X} = \frac{1}{T}\sum_{t=1}^{T} X_t$. In(3.7) the $\hat{\beta}_j^{(1)}$ come from the sample analog of the first p equations in (3.4), namely

$$\hat{\Gamma}_p \hat{\beta}^{(1)} = -\mathbf{c}_p , \tag{3.9}$$

where $\hat{\Gamma}_p = [c(i-j)]_{i,j=1}^{p}$,$\mathbf{c}_p = (c(1), \ldots, c(p))'$ and $\hat{\beta}^{(1)} = (\hat{\beta}_1^{(1)}, \ldots, \hat{\beta}_p^{(1)})'$. These are the sample YW equations.

A corrected moment estimator is

$$\hat{\sigma}^2_{R1c} = \frac{T}{T-p-1}\hat{\sigma}^2_{R1} , \tag{3.10}$$

where the correcting factor is motivated by the finite-sample, bias-correcting factor used in standard variance estimation.

Let us now turn to the consideration of simple models. For an AR(1) model, with $\beta_1 = \beta$,Shaman and Stine(1988) deduce that

$$E(\hat{\beta}^{(1)} - \beta) = T^{-1}(1 - 4\beta) + O(T^{-2}) . \tag{3.11}$$

Using (3.11) we can prove the following result, whose proof is given in the Appendix.

Proposition 3.1. *For the AR(1) model, we have*

$$E(\hat{\sigma}^2_{R1} - \sigma^2_a) = \frac{-\sigma^2_a}{T}\frac{1-7\beta^2}{1-\beta^2} + O(T^{-2}) . \tag{3.12}$$

Shaman and Stine(1988,p.846) give the following formulas for the asymptotic bias of $\hat{\beta}_1^{(1)}$ and $\hat{\beta}_2^{(1)}$ for the AR(2) model:

$$E(\hat{\beta}_1^{(1)} - \beta_1) = T^{-1}\{(1 - 2\beta_1 - \beta_2) - \frac{2\beta_1\beta_2(1+\beta_2)}{(1+\beta_2)^2 - \beta_1^2}\} + O(T^{-2}), \tag{3.13}$$

$$E(\hat{\beta}_2^{(1)} - \beta_2) = T^{-1}\{(2 - 4\beta_2) - \frac{2\beta_2(1+\beta_2)^2}{(1+\beta_2)^2 - \beta_1^2}\} + O(T^{-2}). \tag{3.14}$$

Using these results we obtain

Proposition 3.2. *For the AR(2) model we have*

$$E(\hat{\sigma}^2_{R1} - \sigma^2_a) = \frac{-\sigma^2_a}{T}\frac{P_{MM}(2.0)}{(\beta_2+1)(\beta_2-1)(1-\beta_1+\beta_2)(1+\beta_1+\beta_2)} + O(T^{-2}). \tag{3.15}$$

where the polymonial $P_{MM}(2,0)$ is given in the Appendix.

We consider the model (3.1) with unknown mean μ , estimated by $\bar{X}$ .Assuming that μ is known(e.g. equal to 0) has an effect over some of the results, for example over (3.11),(3.13) and (3.14).

The autocovariances γ_j are estimated by the c_j defined in (3.8).Several other estimators are often encountered in the literature. The c_j have the property that $\hat{\Gamma}_p$ in (3.9) is positive definite, and that the $\hat{\beta}^{(1)}$ defined in (3.9) are such that (1.2) with $\hat{\beta}_j^{(1)}$ replaced for β_j has all roots less than one in absolute value, so that the fitted model is causal.

In most cases, moment estimators are less efficient than LS or ML estimators. For an AR(p) model, the YW estimator $\hat{\beta}^{(1)}$ has the same asymptotic distribution as the ML estimator. Moreover, the Durbin-Levinson recursions can be used to compute $\hat{\beta}^{(1)}$,avoiding the matrix inversions in (3.9). See Morettin(1984) for details.

For the AR(2) model, Tjøstheim and Paulsen(1983) have expressions for the bias of $\hat{\beta}_1^{(1)}$ and $\hat{\beta}_2^{(1)}$ in terms of the roots z_1 and z_2 of (1.2). As a consequence they find the bias of the estimator of the residual variance in terms of these roots. This is given in Paulsen and Tjøstheim(1985), equation (3.10).

If $\beta = 0$ in (3.12) or $\beta_1 = \beta_2 = 0$ in (3.15), we obtain $-\sigma_a^2/T$, which is the bias of the moment estimator of the variance of i.i.d observations.

An interesting question is whether setting $\beta_2 = 0$ in (3.15) we obtain (3.12). This is true in this particular case, but is not expected in general. For example, setting $\beta_2 = 0$ in (3.13) we do not obtain (3.11). Further comments on this question appear at the beginning of the Appendix.

3.2. Least Squares Estimators

The LS procedure consists in estimating $\mu, \beta_1, \ldots, \beta_p$ by minimizing

$$S = \sum_{t=p+1}^{T} a_t^{*2} = \sum_{t=p+1}^{T} \{(X_t - \mu^*) + \beta_1^*(X_{t-1} - \mu^*) + \ldots + \beta_p^*(X_{t-p} - \mu^*)\}^2, \quad (3.16)$$

among all relevant choices of $\mu^*, \beta_1^*, \ldots, \beta_p^*$. An estimate of μ can be taken to be

$$\hat{\mu}_2 = \frac{\bar{X}_1 + \hat{\beta}_1^{(2)}\bar{X}_2 + \ldots + \hat{\beta}_p^{(2)}\bar{X}_p}{1 + \hat{\beta}_1^{(2)} + \ldots + \hat{\beta}_p^{(2)}}, \quad (3.17)$$

where

$$\bar{X}_{j+1} = \frac{1}{T-p} \sum_{t=p+1-j}^{T-j} X_t. \quad (3.18)$$

This will be a non-linear estimation procedure, that can be proposed as a recursive procedure,leading to the estimate $\hat{\beta}^{(2)} = (\hat{\beta}_1^{(2)}, \ldots, \hat{\beta}_p^{(2)})'$, for β and the estimate

$$\hat{\sigma}_{R2}^2 = \frac{1}{T-2p-1} \sum_{t=p+1}^{T} \{(X_t - \hat{\mu}_2) + \hat{\beta}_1^{(2)}(X_{t-1} - \hat{\mu}_2) + \ldots + \hat{\beta}_p^{(2)}(X_{t-p} - \hat{\mu}_2)\}^2, \quad (3.19)$$

for σ_a^2. The divisor T-2p-1 comes from T-p equations with p+1 parameters.

Alternatively, the estimation procedure can be simplified by estimating μ by $\bar{X}$, the β_j by their ordinary LS estimators $\hat{\beta}_j^{(3)}$ based on mean-corrected observations $X_t - \bar{X}$, and σ_a^2 by

$$\hat{\sigma}_{R3}^2 = \frac{1}{T-2p-1} \sum_{t=p+1}^{T} \{(X_t - \bar{X}) + \hat{\beta}_1^{(3)}(X_{t-1} - \bar{X}) + \ldots + \hat{\beta}_p^{(3)}(X_{t-p} - \bar{X})\}^2. \quad (3.20)$$

Let us turn now to some simple models.For an AR(1) model, Kendall(1954), Yamamoto and Kunitomo(1984) and Shaman and Stine(1988) give for $\beta_1 = \beta$

$$E(\hat{\beta}^{(2)} - \beta) = T^{-1}(1 - 3\beta) + O(T^{-2}). \quad (3.21)$$

For the AR(2) model,Tjøstheim and Paulsen(1983), Yamamoto and Kunitomo(1984) and Shaman and Stine(1988) give

$$E\{[\hat{\beta}_1^{(2)}, \hat{\beta}_2^{(2)}] - [\beta_1, \beta_2]\} = T^{-1}\{[1 - \beta_1 - \beta_2, 2 - 4\beta_2]\} + O(T^{-2}). \quad (3.22)$$

For the general AR(p) model, Yamamoto and Kunitomo(1984) and Shaman and Stine(1988) derive general procedures to obtain the asymptotic biases of the LS estimators of the parameters, but no expressions for the biases of estimators of the variance are given.

Tuan(1992) considers the case of AR(1), with $\mu = 0$ and obtains $-2\beta/T$ for the bias of the coefficient estimator and

$$Var\{\hat{\beta}^{(2)} - \beta\} = \frac{-\sigma_a^2}{T}(1-\beta^2) + O(T^{-2}).$$

Proposition 3.3. *For the AR(1) model, we have*

$$E(\hat{\sigma}_{R3}^2 - \sigma_a^2) = \frac{-\sigma_a^2}{T}\frac{1-3\beta^2}{1-\beta^2} + O(T^{-2}). \tag{3.23}$$

Proposition 3.4.*For the AR(2) model, we have*

$$E(\hat{\sigma}_{R3} - \sigma_a^2) = \frac{-2\sigma_a^2}{T}\frac{P_{LS}(2,0)}{(\beta_2 - 1)(1-\beta_1+\beta_2)(1+\beta_1+\beta_2)} + O(T^{-2}), \tag{3.24}$$

where $P_{LS}(2,0)$ is given in the Appendix.

We notice that (3.24) reduces to (3.23) for $\beta_2 = 0$.

The simulations in this paper are done using the ITSM package(see Brockwell and Davis,1991b), which minimizes the weighted sum of squares

$$S = \sum_{j=1}^{T}(X_t - \hat{X}_t)^2/r_{j-1}, \tag{3.25}$$

where $\hat{X}_t$ are the one-step predictors and $E(X_t - \hat{X}_t)^2 = \sigma_a^2 r_{j-1}$.See section 8.7 of Brockwell and Davis(1991a) for details.Minimization of (3.25) will be equivalent to the minimization of the log-likelihood and LS and ML estimators will have similar asymptotic properties.

It follows then that it is reasonable to assume that the LS estimator for σ_a^2 is given by

$$\hat{\sigma}_{LS}^2 = \frac{1}{T-p}S, \tag{3.26}$$

where (3.26) differs from (3.20) by the denominator and the number of terms in the sum.For small p and reasonably large T this will have no essential effect.

We also remark that the asymptotic biases of the LS and ML estimators of the coefficients for the AR(1) and AR(2) models coincide. See formulas (3.21) and (3.22) above and next section.Therefore, it would be reasonable to assume that

$$\hat{\sigma}^2_{LS} \simeq \frac{T}{T-p}\hat{\sigma}^2_{ML}. \tag{3.27}$$

Hence, for an AR(p) model, we have

$$E(\hat{\sigma}^2_{LS} - \sigma^2_a) \simeq E(\frac{T}{T-p}\hat{\sigma}^2_{ML} - \sigma^2_a) = \frac{T}{T-p}E(\hat{\sigma}^2_{ML} - \sigma^2_a) + \sigma^2_a\frac{p}{T-p}. \tag{3.28}$$

In the simulations, we have taken $\sigma^2_a = 1$, hence for an AR(1) model, for example, we can write briefly

$$BIAS(LS) = \frac{T}{T-1}BIAS(ML) + \frac{1}{T-1} \simeq BIAS(ML) + \frac{1}{T}. \tag{3.29}$$

This result can be checked in Table 2, where we can see that it holds with a great degree of accuracy. The same is true for p=2.

3.3. Maximum Likelihood Estimators

In this section we consider maximum likelihood estimation when the underlying process is assumed to be Gaussian.

In one approach, the likelihood function is obtained from the joint density of $\mathbf{X} = (X_1, \ldots, X_T)'$ conditioned on the "initial values" $X_{-p+1}, \ldots, X_0$, and it is

$$(2\pi\sigma_a^2)^{-T/2} exp\{\frac{-1}{2\sigma_a^2}\sum_{t=1}^{T}[(X_t - \mu) + \beta_1(X_{t-1} - \mu) + \ldots + \beta_p(X_{t-p} - \mu)]^2\}. \qquad (3.30)$$

We call this the conditional likelihood function and the set at which it is maximized the conditional maximum likelihood estimates(MLE) of the parameters. The situation is analogous to that faced in the first part of Section 3.2: The estimation problem may be solved recursively for estimates $\hat{\mu}_4$ and $\hat{\beta}_j^{(4)}$, and then estimate σ_a^2 by

$$\hat{\sigma}_{R4}^2 = \frac{1}{T}\sum_{t=1}^{T}\{(X_t - \hat{\mu}_4) + \hat{\beta}_1^{(4)}(X_{t-1} - \hat{\mu}_4) + \ldots + \hat{\beta}_p^{(4)}(X_{t-p} - \hat{\mu}_4)\}^2. \qquad (3.31)$$

We can consider a corrected estimator

$$\hat{\sigma}_{R4c}^2 = \frac{T}{T-p-1}\hat{\sigma}_{R4}^2. \qquad (3.32)$$

The estimation procedure can be simplified by estimating μ by $\bar{X}$, the β_j by the $\hat{\beta}_j^{(5)}$ obtained by minimizing (3.30) with $\bar{X}$ substituted for μ, and σ_a^2 by

$$\hat{\sigma}_{R5}^2 = \frac{1}{T}\sum_{t=1}^{T}\{(X_t - \bar{X}) + \hat{\beta}_1^{(5)}(X_{t-1} - \bar{X}) + \ldots + \hat{\beta}_p^{(5)}(X_{t-p} - \bar{X})\}^2. \qquad (3.33)$$

Then a corrected estimator is obtained as in (3.32).

In a second approach, the underlying stochastic process is taken as Gaussian, stationary and stable, so that the joint density of $\mathbf{X}$ is

$$(2\pi)^{-T/2}|\Sigma|^{-1/2} exp\{\frac{-1}{2}[(\mathbf{X} - \mu)'\Sigma^{-1}(\mathbf{X} - \mu)]\}, \qquad (3.34)$$

where $\mu = E(\mathbf{X}) = (\mu, \ldots, \mu)'$ is the vector of expected values, and Σ is the covariance matrix of $\mathbf{X}$. We may consider here the estimation of μ by $\bar{X}$ and of σ_a^2 by

$$\hat{\sigma}_{R6}^2 = \frac{1}{T}(\mathbf{X} - \bar{\mathbf{X}})'\hat{\mathbf{Q}}^{-1}(\mathbf{X} - \bar{\mathbf{X}}), \qquad (3.35)$$

where $\bar{\mathbf{X}} = (\bar{X}, \ldots, \bar{X})'$ and $\Sigma = \sigma_a^2\mathbf{Q}$, so that $\hat{\mathbf{Q}}$ is the estimated matrix formed with estimates $\hat{\beta}_j^{(6)}$ of the β_j parameters. A corrected estimator $\hat{\sigma}_{R6c}^2$ may also be formed.

In view of the difference between this proposal and the first one in this section, we may call the approach unconditional maximum likelihood, which will then apply to (3.35) and the corrected estimator.

Asymptotic biases for ML estimator of parameters and residual variance have been considered by Tanaka(1984) and Cordeiro and Klein(1993). For the AR(1) and AR(2) models they obtained the same expressions as those given in formulas (3.21) and (3.22), respectively, for the asymptotic biases of the coefficients. For the asymptotic biases of the variance the results are

$$E(\hat{\sigma}^2_{R6} - \sigma^2_a) = \frac{-2}{T}\sigma^2_a + O(T^{-2}), \tag{3.36}$$

for the AR(1) model and

$$E(\hat{\sigma}^2_{R6} - \sigma^2_a) = \frac{-3}{T}\sigma^2_a + O(T^{-2}), \tag{3.37}$$

for the AR(2) model.Expressions for higher order models are also available.See Section 5.2 below for further considerations on the ML method.

The results (3.36) and (3.37) seem intriguing. In Figure 1(a) we have the plot of $TE(\hat{\sigma}^2_a - \sigma^2_a)$, when $\sigma^2_a = 1$, for MM,LS and ML estimators.It shows that $BIAS(ML) < BIAS(LS) < BIAS(MM)$ as expected. However, note that $|BIAS(ML)| > |BIAS(LS)|$ or $|BIAS(ML)| > |BIAS(MM)|$ if, for example, $|\beta| > 0.6$, which seems strange.

4. Moving Average Models

The MA(q) model can be written as

$$X_t - \mu = \sum_{k=0}^{q} \alpha_k a_{t-k} , \qquad t = \ldots, -1, 0, 1, \ldots , \tag{4.1}$$

with $\alpha_0 = 1$. The covariance sequence is

$$\gamma_s = E(X_t - \mu)(X_{t+s} - \mu) = \sigma_a^2 \sum_{k=0}^{q-|s|} \alpha_k \alpha_{k+|s|} , \quad s = 0, \pm 1, \ldots , \tag{4.2}$$

and it satisfies the "inversion formula "(3.5), where the spectral density of the process is now

$$f(\lambda) = \frac{\sigma_a^2}{2\pi} | \sum_{k=0}^{q} \alpha_k e^{i\lambda k} |^2 , \quad -\pi \le \lambda \le \pi . \tag{4.3}$$

The correlation sequence is

$$\rho_s = \frac{\gamma_s}{\gamma_0} = \frac{\sum_{k=0}^{q-|s|} \alpha_k \alpha_{k+|s|}}{1 + \alpha_1^2 + \ldots \alpha_q^2} , \quad s = \pm 1, \pm 2 \ldots . \tag{4.4}$$

We shall denote by $\hat{\sigma}^2_{Mk}$, $k = 1, 2, \ldots$ the various estimators in the MA(q) case and by $\hat{\sigma}^2_{Mkc}$ the corrected estimators.

4.1 Moment Estimators

The spectral density (4.3) can be written as

$$\frac{\sigma_a^2}{2\pi} | \sum_{k=0}^{q} \alpha_k e^{i\lambda k} |^2 = \frac{\sigma_a^2}{2\pi} (\sum_{k=0}^{q} \alpha_k e^{i\lambda k}) (\sum_{j=0}^{q} \alpha_j e^{-i\lambda j}) = \frac{\gamma_0}{2\pi} \sum_{s=-q}^{q} \rho_s e^{-i\lambda s} . \tag{4.5}$$

The correlations ρ_s can be estimated by the sample correlations $r_s = c_s / c_0$, $s = 1, \ldots, q$, where c_s was defined in (3.8). The moment estimators $\hat{\alpha}_k^{(1)}$ of the α_k are those satisfying

$$\frac{1}{1 + \hat{\alpha}_1^{(1)2} + \ldots + \hat{\alpha}_q^{(1)2}} (\sum_{k=0}^{q} \hat{\alpha}_k^{(1)} z^k) (\sum_{j=0}^{q} \hat{\alpha}_j^{(1)} z^{-j}) = \sum_{s=-q}^{q} r_s z^{-s} . \tag{4.6}$$

A moment estimator for σ_a^2 is then

$$\hat{\sigma}^2_{M1} = \frac{c_0}{1 + \hat{\alpha}_1^{(1)2} + \ldots + \hat{\alpha}_q^{(1)2}} , \tag{4.7}$$

obtained from (4.2), and a corrected moment estimator is

$$\hat{\sigma}^2_{M1c} = \frac{T}{T-p-1}\hat{\sigma}^2_{M1}. \tag{4.8}$$

Biases of moment estimators of the residual variance for the MA case have not been considered in the literature, probably due to the complications involved. For the MA(1) model we have the following result, whose proof is again given in the Appendix.

Proposition 4.1. *In the MA(1) model $X_t - \mu = a_t + \alpha a_{t-1}$,the moment estimator defined in (4.7) has*

$$E(\hat{\sigma}^2_{M1} - \sigma^2_a) = \frac{-\sigma^2_a}{T}\frac{1-7\alpha^2-2\alpha^3-9\alpha^4-4\alpha^5-9\alpha^6-2\alpha^7}{(1+\alpha^2)(1-\alpha^4)} + O(T^{-2}). \tag{4.9}$$

If $\alpha = 0$ we get the result for i.i.d observations.
We now turn to the MA(2) model,

$$X_t - \mu = a_t + \alpha_1 a_{t-1} + \alpha_2 a_{t-2}. \tag{4.10}$$

Proposition 4.2. *For the MA(2) model defined by (4.10), the moment estimator (4.7) is such that*

$$E(\hat{\sigma}^2_{M1} - \sigma^2_a) = \frac{-\sigma^2_a}{T}\frac{P_{MM}(0,2)}{(1-\alpha_2)(1-\alpha_1+\alpha_2)(1+\alpha_1+\alpha_2)(1+\alpha_1^2+\alpha_2^2)^2} + O(T^{-2}), \tag{4.11}$$

where $P_{MM}(0,2)$ is given in the Appendix.

If $\alpha_2 = 0$ in (4.11) we do not get the approximation (4.9).See comments in the Appendix.

4.2. Least Squares and Maximum Likelihood Estimators

Simple criterion functions like (3.16), involving the unknown parameters and the observable X_t's, are not available for the MA(q) model. However, several approximations to the likelihood function under normality can be regarded as exactly or approximately least squares procedures, and as such are identified in the literature. Some of these will be considered below.

We follow Priestley(1981,Section 5.4.2). As in our Section 3.3 we distinguish between a conditional and an unconditional maximum likelihood under normality. In the conditional approach we set $a_{-q+1} = \ldots = a_0 = 0$ and maximize the resulting quadratic form in the exponent of the joint normal density. Let $\hat{\mu_2}, \hat{\alpha_1}^{(2)}, \ldots, \hat{\alpha_q}^{(2)}$ be the estimates obtained by using this approach. Then σ_a^2 is estimated by

$$\hat{\sigma}^2_{M2} = \frac{1}{T}\sum_{t=1}^{T}[\hat{a}_t^{(2)}]^2, \tag{4.12}$$

where $\hat{a}_t^{(2)}$ estimates the error a_t of the model. A corrected estimator is

$$\hat{\sigma}^2_{M2c} = \frac{T}{T-q-1}\hat{\sigma}^2_{M2}. \tag{4.13}$$

The unconditional approach is similar to that described in Section 3.3, and hence

$$\hat{\sigma}^2_{M3} = \frac{1}{T}(\mathbf{X}-\bar{\mathbf{X}})'\hat{\mathbf{P}}^{-1}(\mathbf{X}-\bar{\mathbf{X}}), \tag{4.14}$$

where now the covariance matrix of the process having components given by (4.2) is denoted as $\sigma_a^2\mathbf{P}$, and $\hat{\mathbf{P}}$ is obtained by replacing α_k by $\hat{\alpha}_k^{(3)}$ so that $\hat{\mathbf{P}}$ has components $\sum_{k=0}^{q-|s|}\hat{\alpha}_k^{(3)}\hat{\alpha}_{k+|s|}^{(3)}$ along diagonals s=0,±1,...,±q, respectively, and 0's elsewhere. A corrected estimator is

$$\hat{\sigma}^2_{M3c} = \frac{T}{T-q-1}\hat{\sigma}^2_{M3}. \tag{4.15}$$

Some results for simple models are given by Tanaka(1984) and by Cordeiro and Klein(1993) for $\mu = 0$ and μ unknown.

For the MA(1) case, in the case of μ unknown, with $\alpha_1 = \alpha$,

$$E(\hat{\alpha}-\alpha) = \frac{1+2\alpha}{T} + O(T^{-2}), \tag{4.16}$$

$$E(\hat{\sigma}^2_{M2}-\sigma_a^2) = \frac{-2\sigma_a^2}{T} + O(T^{-2}). \tag{4.17}$$

For the MA(2) case,

$$E([\hat{\alpha_1},\hat{\alpha_2}]-[\alpha_1,\alpha_2]) = T^{-1}\left[1+\alpha_1+\alpha_2, 1+3\alpha_2\right] + O(T^{-2}), \tag{4.18}$$

$$E(\hat{\sigma}^2_{M2} - \sigma^2_a) = \frac{-3\sigma^2_a}{T} + O(T^{-2}). \qquad (4.19)$$

Note that the biases of the variance estimators do not depend on the coefficients of the models.

We now turn to the consideration of LS estimators for the MA(1) model. Using a "long autoregression approximation "to a moving average process, as suggested by Durbin(1959), we can get

$$E(\hat{\sigma}^2_{LS} - \sigma^2_a) \simeq \frac{-\sigma^2_a}{T}\frac{1-3\alpha^2}{(1-\alpha^2)} + 4\sigma^2_a \frac{\alpha}{1-\alpha^2} E(\hat{\alpha}_{LS} - \alpha). \qquad (4.20)$$

Two comments are in order. The first factor in the RHS of (4.20) is the bias for the LS estimator of σ^2_a in the AR(1) model and the bias of $\hat{\alpha}_{LS}$ is not available.If we use for this bias the expression (4.16) we obtain for (4.20)

$$E(\hat{\sigma}^2_{LS} - \sigma^2_a) = \frac{-\sigma^2_a}{T}\frac{1-4\alpha-11\alpha^2}{1-\alpha^2} + O(T^{-2}). \qquad (4.21)$$

The proof of (4.21) is given in the Appendix. This value will be used in the table for the MA(1) model. In Figure 1(b) we have plotted again $TE(\hat{\sigma}^2_a - \sigma^2_a)$, when $\sigma^2_a = 1$, for MM,LS and ML estimators,for the first order moving-average.It shows a similar situation to Figure 1(a).

The argument in (3.27)-(3.29) holds for all models, that is, we have approximately,

$$BIAS(LS) \simeq BIAS(ML) + (p+q)/T \quad . \qquad (4.22)$$

This can be checked in Tables 2-6.

5. Autoregressive Moving Average Models

The ARMA(p,q) model was given in (1.1), with (1.2) and (1.3) holding. The covariance sequence satisfies

$$\gamma_j = -\beta_1\gamma_{j-1} - \cdots - \beta_p\gamma_{j-p} + \gamma_{xa}(j) + \alpha_1\gamma_{xa}(j-1) + \cdots + \alpha_q\gamma_{xa}(j-q)\,, \tag{5.1}$$

where $\gamma_{xa}(j) = E([X_{t-j} - \mu]a_t)$. This implies

$$\gamma_j = -\beta_1\gamma_{j-1} - \cdots - \beta_p\gamma_{j-p}, \quad j \geq p+1 \tag{5.2}$$

and hence $\rho_j = \gamma_j/\gamma_0$ satisfy the same equation. For $j = 0$ we get the variance

$$\gamma_0 = -\beta_1\gamma_1 - \cdots - \beta_p\gamma_p + \sigma_a^2 + \alpha_1\gamma_{xa}(-1) + \ldots + \alpha_q\gamma_{xa}(-q). \tag{5.3}$$

The spectral density of the process is

$$f(\lambda) = \frac{\sigma_a^2}{2\pi}\frac{|\sum_{j=0}^{q}\alpha_j e^{ij\lambda}|^2}{|\sum_{j=0}^{p}\beta_j e^{ij\lambda}|^2}\,, \quad -\pi \leq \lambda \leq \pi. \tag{5.4}$$

Of considerable practical importance is the ARMA(1,1) model,

$$X_t - \mu + \beta(X_{t-1} - \mu) = a_t + \alpha a_{t-1}\,. \tag{5.5}$$

for which we obtain

$$\gamma_0 = -\beta\gamma_1 + \sigma_a^2 + \alpha\gamma_{xa}(-1),$$

$$\gamma_1 = -\beta\gamma_0 + \alpha\sigma_a^2,$$

$$\gamma_j = -\beta\gamma_{j-1}, \quad j \geq 2 \tag{5.6}$$

and since $\gamma_{xa}(-1) = (\alpha - \beta)\sigma_a^2$, we have

$$\gamma_0 = \frac{1 + \alpha^2 - 2\alpha\beta}{1 - \beta^2}\sigma_a^2,$$

$$\gamma_1 = \frac{(1 - \alpha\beta)(\alpha - \beta)}{1 - \beta^2}\sigma_a^2,$$

$$\gamma_j = -\beta\gamma_{j-1}, \quad j \geq 2. \tag{5.7}$$

From (5.7), the autocorrelations are

$$\rho_j = \begin{cases} 1, & j=0 \\ \frac{(1-\alpha\beta)(\alpha-\beta)}{1+\alpha^2-2\alpha\beta}(-\beta)^{j-1}, & j \geq 1. \end{cases} \tag{5.8}$$

5.1. Moment Estimators

Box and Jenkins(1976,p.201) give a general method for obtaining initial estimates of the parameters of an ARMA(p,q) model, which can be viewed as moment-type estimators. A related reference is Brockwell and Davis(1991a,p.250).

In this section we concentrate in the case of the ARMA(1,1) model, given by (5.5). From (5.7) we find that an estimate for σ_a^2 is given by

$$\hat{\sigma}_a^2 = \frac{1-\hat{\beta}^{(1)2}}{1+\hat{\alpha}^{(1)2}-2\hat{\alpha}^{(1)}\hat{\beta}^{(1)}} c_0 , \tag{5.9}$$

where $\hat{\alpha^{(1)}}$ and $\hat{\beta}^{(1)}$ are the moment estimators of α and β, coming from (5.8), for $j = 1,2$ and ρ_1 and ρ_2 estimated by the sample correlations r_1 and r_2. As in previous cases, call $\hat{\sigma}^2_{Ak}, k = 1,2,\ldots$ the various estimators that we consider in the ARMA(p,q) models and $\hat{\sigma}^2_{Akc}$ the corrected estimators.Denote the estimators of α and β simply by $\hat{\alpha}$ and $\hat{\beta}$, respectively.

The following result can then be derived.

Proposition 5.1. *For the ARMA(1,1) model, we have*

$$E(\hat{\sigma}^2_{A1} - \sigma_a^2) = \frac{-\sigma_a^2}{T} \frac{P_{MM}(1,1)}{(\alpha^2-1)(\beta^2-1)(\alpha\beta-1)(\alpha-\beta)} + O(T^{-2}) , \tag{5.10}$$

where the polynomial $P_{MM}(1,1)$ is given in the Appendix.

If $\alpha = 0$ in (5.10) we get (3.12), the AR(1) case. But if $\beta = 0$ we do not get (4.9), the MA(1) case, the same fact that occurred for the MA(2) model.

5.2. Least Squares and Maximum Likelihood Estimators

Consider the general ARMA(p,q) model given by (1.1). As in Sections 3.3 and 4.2, we may consider the conditional and unconditional approaches to maximum likelihood estimation. See Box and Jenkins(1976) and Priestley(1981) for further details.

A convenient way to treat the problem is by using the prediction error decomposition of the likelihood function. This avoids the direct calculation of the determinant and of the inverse of the covariance function of the vector $\mathbf{X}$ in

$$L(\mu,\Sigma) = (2\pi)^{-T/2}|\Sigma|^{-1/2} exp\{-\frac{1}{2}(\mathbf{X}-\mu)'\Sigma^{-1}(\mathbf{X}-\mu)\}. \tag{5.11}$$

The prediction error decomposition gives

$$L(\beta,\alpha,\sigma_a^2) = (2\pi\sigma_a^2)^{-T/2}(r_0\cdot r_1\cdots r_{T-1}) exp\{-\frac{1}{2}\sigma_a^{-2}\sum_{j=1}^{T}(X_j-\hat{X}_j)^2/r_{j-1}\}, \tag{5.12}$$

where $r_n = v_n/\sigma_a^2$, $v_n = E(X_{n+1}-\hat{X}_{n+1})^2$ is the mean square error of prediction and the predictors $\hat{X}_j$ can be computed recursively through the Durbin-Levinson recursions or the Innovations Algorithm. See Brockwell and Davis(1991a) for details.

The ML estimator of the residual variance is then

$$\hat{\sigma}_{A2}^2 = \frac{S(\hat{\beta},\hat{\alpha})}{T}, \tag{5.13}$$

where

$$S(\hat{\beta},\hat{\alpha}) = \sum_{j=1}^{T}(X_j-\hat{X}_j)^2/r_{j-1}, \tag{5.14}$$

and $\hat{\beta}$, $\hat{\alpha}$ are the values of β and α which minimizes the reduced likelihood

$$\ell(\beta,\alpha) = \ell n\frac{S(\beta,\alpha)}{T} + \frac{1}{T}\sum_{j=1}^{T}\ell n(r_{j-1}). \tag{5.15}$$

In our simulations, the program PEST of the ITSM package together with the innovations algorithm give, from (5.15), the ML estimates of β and α , and then from (5.13) the ML estimate of the variance is obtained.

The LS estimators of the parameters are obtained minimizing the sum in (5.14), with the estimated parameters replaced by the parameters, with respect to β and α. The LS estimator of the variance is then

$$\hat{\sigma}^2_{A3} = \frac{1}{T-p-q} S(\hat{\beta}, \hat{\alpha}). \tag{5.16}$$

In the case of invertible models the minimization of S and ℓ are equivalent and then ML estimators and LS estimators will have similar asymptotic properties.

Tanaka(1984) and Cordeiro and Klein(1993) derived formulas for the biases of ML estimators for the coefficients and residual variance for ARMA models. In particular, for the ARMA(1,1) case we have that

$$E(\hat{\sigma}^2_{A2} - \sigma^2_a) = \frac{-3\sigma^2_a}{T} + O(T^{-2}). \tag{5.17}$$

Using the same kind of approach as for the MA(1) model, an approximate bias for $\hat{\sigma}^2_{LS}$ could eventually be derived. We do not pursue this here.

6. Simulations

Table 1 gives the models that were generated in order to verify empirically the conclusions of the theoretical results presented in the paper. In all cases $\sigma_\varepsilon^2 = 1$. Four different sample sizes were considered:$T = 50, 100, 200, 400$. For each sample size, 100 replicates were taken, for each model.

For the computations, we have used the ITSM package(Brockwell and Davis,1991b). We now describe briefly how the estimators are computed.

*a) Moment Estimators.*For the AR models, these are the Yule-Walker estimators, computed via the Durbin-Levinson algorithm.For MA and ARMA models, the innovations algorithm is used.

*b) Least Squares estimators.*The program computes the least squares estimators for any ARMA model by minimizing (5.14).

*c) Maximum Likelihood Estimators.*Exact ML estimators are computed via the prediction error decomposition in conjunction with the innovations algorithm. See Brockwell and Davis(1991a,1991b) for details.

Tables 2-6 give the main results of the simulations.A total of 24 models were generated: three for each of AR(1),AR(2),MA(1),MA(2) and ARMA(1,1) models.

For each table and model, the following notation is used:

T: sample size
MM: method of moments
LS: method of least squares
ML: method of maximum likelihood
EST. BIAS: estimated bias,obtained by the average of the 100 replicates
ST.ERROR: estimated standard error of the estimated bias, computed from the 100 replicates
ASYM. BIAS: asymptotic bias, given by the theoretical formulas.

Note that we have no evaluation of the asymptotic biases for LS estimates in MA(2) and ARMA(1,1) models.

The estimated standard error is given by $s/10$, where

$$s^2 = \sum_{i=1}^{100} (\hat{b}_i - \bar{b})^2/100 \quad , \hat{b}_i = \hat{\sigma}_i^2 - 1 \quad , i = 1, \ldots, 100 \quad , \bar{b} = \sum_{i=1}^{100} \hat{b}_i/100 \quad ,$$

where $\hat{\sigma}_i^2$ is the variance estimate for each method.

Figure 2 shows the estimated bias, for each model and each method. It provides a convenient way of comparing the estimation methods within each model for the three sets of parameters, and among models.

To facilitate the interpretation of the tables of results, we computed the intervals $\bar{b} \pm 3s/10$, and marked with * those estimates for which the interval does not include the corresponding asymptotic bias. The results of this analysis can be summarized as follows:

Table	Model	Mod. of roots	Analysis of results
2	AR(1)	0.30	* at ML for T=50
		0.60	no *
		0.90	* at MM all T, LS for T=50
3	AR(2)	0.84;0.84	* at MM for T=50
		0.85;0.36	* at MM and LS for all T
		0.95;0.95	* at MM and LS for all T
4	MA(1)	0.20	* at MM all T,LS for T=50
		0.40	* at MM for T=50,100,200
		0.80	* at MM and LS for all T
5	MA(2)	0.87;0.23	* at MM all T, ML for T=100
		0.87;0.23	* at MM for T=50,100,400
		0.94;0.64	* at MM for all T
6	ARMA(1,1)	0.70;0.40	*at MM all T, ML for T=50
		0.80;0.60	no *
		0.90;0.30	* at MM for all M

Some conclusions are:

1. Moment estimators tend to have larger bias than least squares estimators, and these larger biases than maximum likelihood estimators, in agreement with what is expected.
2. Estimated biases tend to come closer to the asymptotic values, and hence to become smaller, as T increases, also in agreement with what is expected. However, for moment estimators, even T=400 is inadequate to give a good fit of the simulation to the asymptotic theory in some cases.
3. The estimated biases can differ considerably from the theoretical values, even for large sample sizes.
4. For given sample sizes, the variability of the estimated biases,as measured by the estimated standard errors, tend to have similar values.
5. We detected only a few cases of lack of fit when using the method of maximum likelihood,and except for one case,they all occurred when T=50,the smallest sample size that we studied.
6. When we analyzed the method of least squares, for which no asymptotic biases are available in Tables 5 and 6, cases of lack of fit were observed for all sample sizes,for the MA(1) when its root is close to the boundary of the region of invertibility, and for the AR(2) model when at least one of its roots has this property. Another case. for T=50, occured for the AR(1) model with parameter 0.90.

7. The method of moments led to many more cases of lack of fit. In some tables these cases increase as we approach the boundary of the region of invertibility. In some tables the cases do not tend to desappear with the increase in sample size.

7.Concluding Remarks

In this paper we considered the estimation of the residual variance in ARMA models, that is, the variance of the white noise component of the models defined by (1.1).This variance is a nuisance parameter, and its estimation is important because estimates enter into prediction and confidence intervals, tests of hypothesis, spectral estimates, and other inferential procedures.

In spite of the indicated usefulness, not many results are available about properties of estimators of the residual variance in ARMA models of frequent use. In this paper we considered estimation by three standard methods, namely moments,least squares, and maximum likelihood under normality. We considered five models, namely ARMA(p,q) for (p,q)=(1,0),(2,0),(0,1),(0,2) and (1,1). This gives a total of fifteen needed estimators, of which only those for maximum likelihood procedures had been studied. For the remaining ten cases, we provided asymptotic results for all, except for least squares in the MA(2) and ARMA(1,1) models.

In the analytical part of our study we concentrated in the study of the asymptotic biases of the estimators, by using certain approximation procedures.In general, our approach is to rely on Taylor-type expansions, and use asymptotic results in the literature for means, variances, autocovariances or autocorrelations of linear processes, under different definitions of these sample quantities.

We have no results either exact or approximated for the variances of the variance estimators, and hence we cannot analyze figures of merit such as mean square errors. It has been argued(see,for example, Simonoff,1993),that in the estimation of the variance of a statistic, attention should concentrate in the bias: a negatively biased estimator will lead to misleadingly short prediction or confidence intervals, which is a dangerous situation in an inferencial context.

We studied the asymptotic biases of the ten estimators metioned above. In the case of the one-parameter models AR(1) and MA(1), the asymptotic biases can be graphed easily, and the theoretical behavior studied and compared for varying parameter values; see graphs of Figures 1(a) and 1(b).In general maximum likelihood estimators have smaller biases than least squares estimators which in turn have smaller biases than moment estimators.

Figure 1(a) depicts $T\sigma_a^2(BIAS) = T(BIAS)$ because here $\sigma_a^2 = 1$, in the cases of the AR(1) models under MM, LS and ML. Hence, it corresponds to comparing (3.12),(3.23) and (3.36). We note that the curves for MM and LS are parameter dependent, while for ML the curve is a straight line at -2. Since the ML estimator was defined with denominator T , it pays to correct it, as we did for the LS estimator in (3.27)-(3.29).For paramete values close to the region of invertibility($|\beta|$ close to 1) MM has sensibly larger values than LS.

We conclude that for AR(1) models a ML estimator corrected for bias can be recommended for any values of the parameter.

A similar argument is supported by Figure 1(b), in the case of MA(1) model, only that now LS has larger biases than MM and the graphs for MM and LS are not symmetric, while the corresponding curves were symmetric in Figure 1(a).

These large-sample results were reasonably well supported by our simulations presented in Section 6. For example, the connection between LS and ML established in equations (3.27)-(3.29), or the parameter-independence of ML results reported in equations

(3.36),(3.37),(4.17),(4.19) and (5.17), were well fitted for all models, choices of parameters, and sample sizes.

Even the largest sample size of T=400 was not good enough to exhibit a good fit of the MM estimation procedure.

Appendix: Proofs of the Theorems

Concerning the method of moments, we notice that the nature of the proofs is different(and simpler) for the AR(1) and AR(2) cases. In effect, to derive asymptotic expressions for $E(\hat{\sigma}^2_{R1} - \sigma^2_a)$ we only need those for $E(c_0 - \gamma_0)$ and $E(\hat{\beta}_i - \beta_i)$: The former is well known, and the latter are given, for example, in Shaman and Stine(1988). It is interesting to note that in all cases $AR(2)(\beta_1, 0) = AR(1)(\beta_1)$.

In the MA(1) and MA(2) cases, $E(c_0 - \gamma_0)$ and $E(r_i - \rho_i)$ are known, for example from Fuller(1976) or Anderson(1971). However expressions for $E(\hat{\alpha}_i - \alpha_i)$ have not been written in the literature. In this case, we resort to simple Taylor expansions, starting from the known expressions for the ρ_i's in terms of the α_j's.These results have the peculiarity that for example,$MA(2)(\alpha_1, 0) \neq MA(1)(\alpha_1)$.

This can be explained in terms of a simple example. In the MA(2) case, $\gamma_2 = \alpha_2$, so that $\frac{\partial \gamma_2}{\partial \alpha_2} = 1$;however, in MA(1) $\gamma_2 = 0$ and γ_2 is not a function of α_2 , so that the corresponding derivative will not become zero and the two results will not agree.In consequence, MA(2) will not reproduce MA(1) by setting $\alpha_2 = 0$, and ARMA(1,1) will not reproduce MA(1) by setting $\beta = 0$.

In the proofs of the theorems,expressions like $P_{MM}(0,2)$ and $P_{MM}(1,1)$ presented below, were derived by using the Mathematica programs(Wolfram,1988) to avoid algebraic mistakes.

Proposition 3.1. The YW equations for the AR(1) model are

$$\gamma_0 + \beta\gamma_1 = \sigma^2_a,$$

$$\gamma_1 + \beta\gamma_0 = 0,$$

so that $\sigma^2_a = \gamma_0(1 - \beta^2)$. The sample YW equations yield

$$\hat{\sigma}^2_{R1} = c_0(1 - \hat{\beta}^2).$$

Hence,

$$\hat{\sigma}^2_{R1} - \sigma^2_a = (c_0 - \gamma_0) - (c_0\hat{\beta}^2 - \gamma_0\beta^2). \tag{A.1}$$

Using a Taylor expansion we have

$$c_0\hat{\beta}^2 - \gamma_0\beta^2 \simeq \beta^2(c_0 - \gamma_0) + 2\gamma_0\beta(\hat{\beta} - \beta),$$

and substituting in (A.1) we get

$$E(\hat{\sigma}^2_{R1} - \sigma^2_a) \simeq (1 - \beta^2)E(c_0 - \gamma_0) - 2\gamma_0\beta E(\hat{\beta} - \beta). \tag{A.2}$$

Now (Fuller,1976,Th.6.2.2)

$$E(c_0 - \gamma_0) \simeq -Var(\bar{X}) = -2\pi f(0)/T + O(T^{-2}),$$

where $f(\lambda)$ is the spectral density of X_t, and hence

$$E(c_0 - \gamma_0) = \frac{-\sigma_e^2}{T}\frac{1}{(1+\beta)^2} + O(T^{-2}).$$

Also, the asymptotic bias of $\hat{\beta}$ is (Shaman and Stine, 1988),

$$E(\hat{\beta} - \beta) = \frac{1-4\beta}{T} + O(T^{-2}),$$

for an unknown μ. By replacing these results in (A.2) we obtain (3.12).

Proposition 3.2. In the AR(2) model the YW equations are

$$\gamma_0 + \beta_1\gamma_1 + \beta_2\gamma_2 = \sigma_a^2,$$

$$\gamma_1 + \beta_1\gamma_0 + \beta_2\gamma_1 = 0, \qquad (A.3)$$

$$\gamma_2 + \beta_1\gamma_1 + \beta_2\gamma_0 = 0.$$

The sample YW equations are obtained replacing γ_j by c_j and β_j by $\hat{\beta}_j^{(1)} = \hat{\beta}_j$. Hence,

$$\gamma_0 = \frac{\sigma_a^2}{1 + \beta_1\rho_1 + \beta_2\rho_2},$$

and then

$$\gamma_0 = \frac{1+\beta_2}{1-\beta_2}\frac{\sigma_a^2}{(1+\beta_2)^2 - \beta_1^2}. \qquad (A.4)$$

From Shaman and Stine(1988,p.846), we have

$$E(\hat{\beta}_1 - \beta_1) = \frac{1 - 2\beta_1 - \beta_2}{T} - \frac{2\beta_1\beta_2(1+\beta_2)}{T[(1+\beta_2)^2 - \beta_1^2]} + O(T^{-2}).$$

$$E(\hat{\beta}_2 - \beta_2) = \frac{2 - 4\beta_2}{T} - \frac{2\beta_2(1+\beta_2)^2}{T[(1+\beta_2)^2 - \beta_1^2]} + O(T^{-2}). \qquad (A.5)$$

From (A.3) we have

$$\hat{\sigma}_{R1}^2 - \sigma_a^2 = (c_0 - \gamma_0) - (c_0\hat{\beta}_2^2 - \gamma_0\beta_2^2) - (c_0\hat{\beta}_1^2\frac{1-\hat{\beta}_2}{1+\hat{\beta}_2} - \gamma_0\beta_1^2\frac{1-\beta_2}{1+\beta_2}). \qquad (A.6)$$

Now,

$$E(c_0 - \gamma_0) \simeq -2\pi f(0)/T \simeq \frac{-\sigma_a^2}{T}\frac{1}{(1+\beta_1+\beta_2)^2}.$$

For the second term in (A.6), expanding in a Taylor series up to order one,

$$E(c_0\hat{\beta}_2^2 - \gamma_0\beta_2^2) \simeq \beta_2^2 E(c_0 - \gamma_0) + 2\gamma_0\beta_2 E(\hat{\beta}_2 - \beta_2). \qquad (A.7)$$

Similarly, for the third term in (A.6), expanding up to order one, the expectation of the third term becomes approximately

$$\beta_1^2\frac{1-\beta_2}{1+\beta_2}E(c_0-\gamma_0)+2\gamma_0\beta_1\frac{1-\beta_2}{1+\beta_2}E(\hat{\beta}_1-\beta_1)-\frac{2\gamma_0\beta_1^2}{(1+\beta_2)^2}E(\hat{\beta}_2-\beta_2). \qquad (A.8)$$

Substituting the expressions appropriatelly and using (A.6) we get the desired result, with

$$P_{MM}(2,0)=-1+7\beta_1^2-4\beta_2-6\beta_1^2\beta_2+6\beta_2^2-5\beta_1^2\beta_2^2+20\beta_2^3+11\beta_2^4.$$

Proposition 3.3. The ordinary LS estimator of the residual variance is(with $\hat{\beta}_1^{(3)}=\hat{\beta}$)

$$\hat{\sigma}_{R3}^2=\frac{1}{T-3}\sum_{t=2}^{T}\{(X_t-\bar{X})+\hat{\beta}(X_{t-1}-\bar{X})\}^2\simeq c_0+c_0\hat{\beta}^2+2c_1\hat{\beta}.$$

Hence,

$$\hat{\sigma}_{R3}^2-\sigma_a^2\simeq(c_0-\gamma_0)+2(c_1\hat{\beta}-\gamma_1\beta)+\beta\gamma_1+c_0\hat{\beta}^2$$

and since $\gamma_1=-\beta\gamma_0$,we have

$$\hat{\sigma}_{R3}^2-\sigma_a^2\simeq(c_0-\gamma_0)+2(c_1\hat{\beta}-\gamma_1\beta)+(c_0\hat{\beta}^2-\gamma_0\beta^2). \qquad (A.9)$$

Considering the Taylor expansions of the second and third terms of (A.9) and taking expectation we obtain

$$E(\hat{\sigma}_{R3}^2-\sigma_a^2)\simeq(1+\beta^2)E(c_0-\gamma_0)+2\beta E(c_1-\gamma_1)+2(\beta\gamma_0+\gamma_1)E(\hat{\beta}-\beta). \qquad (A.10)$$

Now, the last term of (A.10) is zero since $\gamma_1+\beta\gamma_0=0$ and

$$E(c_0-\gamma_0)\simeq -Var(\bar{X})\simeq-\frac{\sigma_a^2}{T}\frac{1}{(1+\beta)^2},$$

$$E(c_1-\gamma_1)=\frac{-\gamma_1}{T}-\frac{\sigma_a^2}{T}\frac{1}{(1+\beta)^2}+O(T^{-2}), \quad \text{(Fuller,1976).}$$

Substituting into (A.10) we obtain (3.23).

Proposition 3.4. Follows along the same lines as the proof of Proposition 3.3, taking into account the expressions of $E(c_i-\gamma_i), i=0,1,2$, given by Fuller(1976). The polynomial in (3.24) is given by

$$P_{LS}(2,0)=-1+3\beta_1^2-\beta_2-\beta_1^2\beta_2+3\beta_2^2+3\beta_2^3.$$

Proposition 4.1. The autocovariance function of an MA(q) model is

$$\gamma_h = \sigma_a^2 \sum_{j=0}^{q-|h|} \alpha_j \alpha_{j+|h|}, \quad |h| \leq q,$$

while $\gamma_h = 0$,for $|h| > q$.Hence

$$\gamma_o = \sigma_a^2(1+\alpha^2),$$

$$\gamma_1 = \alpha\sigma_a^2.$$

Hence,if $\hat{\sigma}^2_{M1}$ is the moment estimator,

$$\hat{\sigma}^2_{M1} = \frac{c_0}{1+\hat{\alpha}^2}$$

and

$$c_0 - \gamma_0 = (\hat{\sigma}^2_{M1} - \sigma_a^2) + (\hat{\sigma}^2_{M1}\hat{\alpha}^2 - \sigma_a^2\alpha^2).$$

Expanding in a Taylor series up to order one the second term of the RHS,

$$c_0 - \gamma_o \simeq (1+\alpha^2)(\hat{\sigma}^2_{M1} - \sigma_a^2) + 2\alpha\sigma_a^2(\hat{\alpha} - \alpha)$$

and hence

$$E(\hat{\sigma}^2_{M1} - \sigma_a^2) \simeq \frac{1}{1+\alpha^2}E(c_0 - \gamma_0) - \frac{2\alpha\sigma_a^2}{1+\alpha^2}E(\hat{\alpha} - \alpha).$$

Now, from (4.6) we obtain(calling $r = r_1 = c_1/c_0$),

$$\hat{\alpha} = \frac{1 - \sqrt{1-4r^2}}{2r},$$

a similar relation holding between α and ρ.

Using a Taylor expansion,

$$\hat{\alpha} - \alpha \simeq (r-\rho)\frac{d\alpha}{d\rho} = (r-\rho)\frac{(1+\alpha^2)^2}{1-\alpha^2}$$

and hence

$$E(\hat{\alpha} - \alpha) \simeq \frac{(1+\alpha^2)^2}{1-\alpha^2}E(r-\rho).$$

Using Theorem 6.2.3 and results of page 239 of Fuller(1976) we get:

$$Var(\tilde{\gamma}_0) \simeq \frac{2\sigma_a^4}{T}(1+\alpha^4+4\alpha^2),$$

$$Cov(\tilde{\gamma}_1,\tilde{\gamma}_0) \simeq \frac{4\sigma_a^4(T-1)}{T^2}\alpha(1+\alpha^2),$$

$$E(r) = \rho - \frac{1}{T}\{\frac{\alpha}{1+\alpha^2} + \frac{(1-\alpha+\alpha^2)(1+\alpha)^2}{(1+\alpha^2)^2} - \frac{2\alpha(1+4\alpha^2+\alpha^4)-4\alpha(1+\alpha^2)^2}{(1+\alpha^2)^3}$$
$$- \frac{4\alpha}{T(1+\alpha^2)}\} + O(T^{-2})$$

and hence

$$E(r-\rho) = \frac{-1}{T}\frac{1}{(1+\alpha^2)^3}\{\alpha^6+4\alpha^5+\alpha^4+4\alpha^3+\alpha^2+4\alpha+1\} + O(T^{-2}).$$

Also,

$$E(c_0-\gamma_0) = -\frac{\sigma_a^2}{T}(1+\alpha)^2 + O(T^{-2}).$$

Substituting everything together we obtain (4.9).

Proposition 4.2. From the general formula of γ_h given in Proposition 4.1, in the case of an MA(2) model we get that

$$\rho_1 = \frac{\alpha_1(1+\alpha_2)}{1+\alpha_1^2+\alpha_2^2}$$

and

$$\rho_2 = \frac{\alpha_2}{1+\alpha_1^2+\alpha_2^2}. \qquad (A.11)$$

Also, an estimator for the residual variance is

$$\hat{\sigma}^2_{M1} = \frac{c_0}{1+\hat{\alpha}_1^2+\hat{\alpha}_2^2}. \qquad (A.12)$$

So, from (A.11) we have $\rho_1 = f_1(\alpha_1, \alpha_2)$ and $\rho_2 = f_2(\alpha_1, \alpha_2)$, from which we can obtain α_1 and α_2 as functions of ρ_1 and ρ_2.Then get the respective moment estimators of the alpha's, using r_1 and r_2.

Now,

$$c_0 - \gamma_0 = (\hat{\sigma}^2_{M1} - \sigma^2_a) + (\hat{\sigma}^2_{M1}\hat{\alpha}_1^2 - \sigma^2_a\alpha_1^2) + (\hat{\sigma}^2_{M1}\hat{\alpha}_2^2 - \sigma^2_a\alpha_2^2). \qquad (A.13)$$

As in the MA(1) case, by expanding each term of (A.13) in a Taylor series up tp order one, we obtain

$$E(\hat{\sigma}^2_{M1} - \sigma^2_a) \simeq \frac{1}{1+\alpha_1^2+\alpha_2^2}E(c_0-\gamma_0) - \frac{2\alpha_1\sigma^2_a}{1+\alpha_1^2+\alpha_2^2}E(\hat{\alpha}_1-\alpha_1) - \frac{2\alpha_2\sigma^2_a}{1+\alpha_1^2+\alpha_2^2}E(\hat{\alpha}_2-\alpha_2). \qquad (A.14)$$

Now, as before,

$$E(c_0-\gamma_0) \simeq -Var(\bar{X}) = \frac{-\sigma^2_a}{T}(1+\alpha_1+\alpha_2)^2 + O(T^{-2})$$

and

$$E(\hat{\alpha}_i - \alpha_i) \simeq E(r_1-\rho_1)\frac{\partial\alpha_i}{\partial\rho_1} + E(r_2-\rho_2)\frac{\partial\alpha_i}{\partial\rho_2}, \qquad i = 1,2. \qquad (A.15)$$

The partial derivatives are obtained through

$$\begin{pmatrix} \frac{\partial\alpha_1}{\partial\rho_1} & \frac{\partial\alpha_1}{\partial\rho_2} \\ \frac{\partial\alpha_2}{\partial\rho_1} & \frac{\partial\alpha_2}{\partial\rho_2} \end{pmatrix} = \begin{pmatrix} \frac{\partial\rho_1}{\partial\alpha_1} & \frac{\partial\rho_1}{\partial\alpha_2} \\ \frac{\partial\rho_2}{\partial\alpha_1} & \frac{\partial\rho_2}{\partial\alpha_2} \end{pmatrix}^{-1}$$

using(A.11).

To obtain $E(r_i - \rho_i).i = 1,2$ proceed as in Proposition 4.1. using the results of Fuller(1976). We have

$$Var(\tilde{\gamma_0}) \simeq \frac{1}{T}(2\gamma_0^2 + 4\gamma_1^2 + 4\gamma_2^2),$$

$$Cov(\hat{\gamma}_1, \hat{\gamma}_0) \simeq \frac{1}{T}(4\gamma_0\gamma_1 + 4\gamma_1\gamma_2),$$

$$Cov(\hat{\gamma}_2, \hat{\gamma}_0) \simeq \frac{1}{T}(4\gamma_0\gamma_2 + 2\gamma_1^2).$$

From these, we obtain

$$E(r_1 - \rho_1) = \frac{-\rho_1}{T} - (1-\rho_1)Var(\bar{X})/\gamma_0 + \frac{\rho_1}{T}(2 + 4\rho_1^2 + 4\rho_2^2) - (4\rho_1 + 4\rho_1\rho_2)/T + O(T^{-2})$$

and

$$E(r_2 - \rho_2) = \frac{-2\rho_2}{T} - (1-\rho_2)Var(\bar{X})/\gamma_0 + \frac{\rho_2}{T}(2 + 4\rho_1^2 + 4\rho_2^2) - (4\rho_2 + 2\rho_1^2)/T + O(T^{-2}).$$

The derivatives in (A.15) are given by:

$$\frac{\partial \alpha_1}{\partial \rho_1} = -(1 + \alpha_1^2 - \alpha_2^2)(1 + \alpha_1^2 + \alpha_2^2)/d,$$

$$\frac{\partial \alpha_1}{\partial \rho_2} = \alpha_1(1 + \alpha_1^2 + \alpha_2^2)(1 + \alpha_1^2 - 2\alpha_2 - \alpha_2^2)/d,$$

$$\frac{\partial \alpha_2}{\partial \rho_1} = -2\alpha_1\alpha_2(1 + \alpha_1^2 + \alpha_2^2)/d,$$

$$\frac{\partial \alpha_2}{\partial \rho_2} = -(1 + \alpha_2)(1 - \alpha_1^2 + \alpha_2^2)(1 + \alpha_1^2 + \alpha_2^2)/d,$$

where

$$d = (\alpha_2 - 1)(1 - \alpha_1 + \alpha_2)(1 + \alpha_1 + \alpha_2).$$

Now, use (A.11) to obtain $E(r_i - \rho_i)$ in terms of α_1, α_2. With the partial derivatives and the expectations above obtain (A.15).Finally,substituting in (A.14) obtain(4.11),with

$$\begin{aligned} P_{MM}(0,2) = 1 &- 5\alpha_1^2 + 2\alpha_1^3 + \alpha_1^4 + 4\alpha_1^5 + \alpha_1^6 + 2\alpha_1^7 + 2\alpha_1^8 + \alpha_2 - 4\alpha_1\alpha_2 \\ &- 9\alpha_1^2\alpha_2 - 4\alpha_1^3\alpha_2 + 15\alpha_1^4\alpha_2 + 4\alpha_1^5\alpha_2 + 9\alpha_1^6\alpha_2 + 4\alpha_1^7\alpha_2 - 9\alpha_2^2 \\ &- 8\alpha_1\alpha_2^2 - 38\alpha_1^2\alpha_2^2 - 12\alpha_1^3\alpha_2^2 - 25\alpha_1^4\alpha_2^2 - 4\alpha_1^5\alpha_2^2 + 8\alpha_1^6\alpha_2^2 \\ &- 11\alpha_2^3 - 12\alpha_1\alpha_2^3 - 22\alpha_1^2\alpha_2^3 - 8\alpha_1^3\alpha_2^3 - 15\alpha_1^4\alpha_2^3 + 4\alpha_1^5\alpha_2^3 - 15\alpha_2^4 \\ &- 16\alpha_1\alpha_2^4 - 33\alpha_1^2\alpha_2^4 - 14\alpha_1^3\alpha_2^4 - 17\alpha_2^5 - 12\alpha_1\alpha_2^5 - 29\alpha_1^2\alpha_2^5 - 4\alpha_1^3\alpha_2^5 \\ &- 15\alpha_2^6 - 8\alpha_1\alpha_2^6 - 8\alpha_1^2\alpha_2^6 - 13\alpha_2^7 - 4\alpha_1\alpha_2^7 - 2\alpha_2^8 \end{aligned}$$

Proof of (4.21): We use the "long AR "approximation to a MA process, suggested by Durbin(1959). We approximate

$$X_t - \mu = a_t + \alpha a_{t-1}$$

by

$$\sum_{j=0}^{B_T} (-\alpha)^j (X_t - \mu) = a_t,$$

where $B_T \to \infty$ as $T \to \infty$, but in such a way that B_T (or a power of it) is dominated by T: $B_T/T \to 0$, for examplе. Then

$$\hat{\sigma}^2_{LS} = \frac{1}{T - B_T - 2} \sum_{t=B_T+1}^{T} [\sum_{j=0}^{B_T} (-\alpha)^j (X_{t-j} - \bar{X})]^2$$

$$\simeq c_0 \sum_{j=0}^{B_T} (-\alpha)^j + 2 \sum_{j=0}^{B_T} \sum_{j'=0}^{B_T} (-\alpha)^{j+j'} c_{j-j'}$$

where in the second sum $j > j'$. After a change in variables, we obtain

$$\hat{\sigma}^2_{LS} \simeq c_0 \sum_{j=0}^{B_T} \hat{\alpha}^{2j} + 2 \sum_{k=1}^{B_T} c_k \sum_{j=k}^{B_T} (-\hat{\alpha})^{2j-k}$$

$$\simeq \frac{c_0}{1 - \hat{\alpha}^2} + 2 \sum_{j=1}^{\infty} c_j \frac{(-\hat{\alpha})^j}{1 - \hat{\alpha}^2}, \qquad (A.16)$$

as $B_T \to \infty$. Since $\gamma_0 = \sigma^2_a(1 + \alpha^2)$, we have that

$$\frac{c_0}{1 - \hat{\alpha}^2} - \sigma^2_a \simeq c_0 \sum_{j=0}^{B_T} \alpha^{2j} - \gamma_0 \sum_{j=0}^{B_T} (-\alpha)^{2j}$$

$$\simeq \sum_{j=0}^{\infty} (\hat{\alpha}^{2j} c_0 - \alpha^{2j} \gamma_0) + 2\gamma_0 \frac{\alpha^2}{1 - \alpha^4}, \qquad (A.17)$$

and hence, expanding this sum, taking into account that $\gamma_j = 0, j > 1$ and that $\gamma_0 = \sigma^2_a(1 + \alpha^2)$, we have that some constant terms cancell out and we are left with

$$\hat{\sigma}^2_{LS} - \sigma^2_a = \frac{c_0 - \gamma_0}{1 - \alpha^2} + [\frac{2\alpha\sigma^2_a(1 + \alpha^2)}{(1 - \alpha^2)^2} + \frac{2\alpha\sigma^2_a(1 - 3\alpha^2)}{(1 - \alpha^2)^2}](\hat{\alpha} - \alpha) +$$

$$\frac{-2\alpha}{1-\alpha^2}(c_1-\gamma_1)+\frac{2}{1-\alpha^2}\sum_{j=2}^{\infty}(-\alpha)^j c_j. \tag{A.18}$$

In taking expected values, we use from Fuller(1976):

$$E(c_0-\gamma_0)\simeq -\frac{\sigma_a^2}{T}(1+\alpha)^2,$$

$$E(c_1-\gamma_1)\simeq -\frac{\sigma_a^2}{T}(1+3\alpha+\alpha^2),$$

$$E(c_j)\simeq -\frac{\sigma_a^2}{T}(1+\alpha)^2, j\geq 2.$$

Then we obtain from (A.18),

$$E(\hat{\sigma}_{LS}^2-\sigma_a^2)\simeq -\frac{\sigma_a^2}{T}\frac{1-3\alpha^2}{1-\alpha^2}+\frac{4\alpha\sigma_a^2}{1-\alpha^2}E(\hat{\alpha}_{LS}-\alpha),$$

which is (4.20). We obtain (4.21) by replacing the bias of the estimator of α by (4.16), taking into account that we write the model slightly differently from Tanaka(1984).

Proposition 5.1..The autocovariances of an ARMA(1,1) model are given by (5.7),or

$$\gamma_h = \begin{cases} \frac{1+\alpha^2-2\alpha\beta}{1-\beta^2}\sigma_a^2, & \text{if } h=0; \\ \frac{(1-\alpha\beta)(\alpha-\beta)}{1-\beta^2}(-\beta)^{h-1}\sigma_a^2, & \text{if } h=1,2,\ldots \end{cases}, \tag{A.19}$$

and an estimator for σ_a^2 is given by (5.9).

We note that

$$\sigma_a^2 = \gamma_0 \frac{1-\beta^2}{(1+\alpha^2)(1-2\alpha\beta/(1+\alpha^2))} \simeq$$

$$\gamma_0 \Big[\frac{1-\beta^2}{1+\alpha^2}(1+\frac{2\alpha\beta}{1+\alpha^2})\Big].$$

This agrees with AR(1) when $\alpha = 0$ and with MA(1) when $\beta = 0$. Then

$$\begin{aligned} \hat\sigma_{A1}^2 - \sigma_a^2 \simeq & \left(c_0\frac{1-\hat\beta^2}{1+\hat\alpha^2} - \gamma_0\frac{1-\beta^2}{1+\alpha^2}\right) \\ & + \left(c_0\frac{2\hat\alpha\hat\beta(1-\hat\beta^2)}{(1+\hat\alpha^2)^2} - \gamma_0\frac{2\alpha\beta(1-\beta^2)}{(1+\alpha^2)^2}\right) \end{aligned}. \tag{A.20}$$

If we note that the spectrum computed at the origin is given by

$$f_X(0) = \frac{\sigma_a^2}{2\pi}\frac{(1+\alpha)^2}{(1+\beta)^2};$$

then

$$E(c_0 - \gamma_0) \simeq -Var(\bar X) \simeq \frac{-\sigma_a^2}{T}\frac{(1+\alpha)^2}{(1+\beta)^2}.$$

Expanding the two terms in (A.20) up to first order and taking expectations we get

$$E(\hat\sigma_{A1}^2 - \sigma_a^2) \simeq \frac{(1+\alpha^2)(1-\beta^2)+2\alpha\beta(1-\beta^2)}{(1+\alpha^2)^2}E(c_0-\gamma_0)+$$

$$2\sigma_a^2\frac{1+\alpha^2-2\alpha\beta}{1-\beta^2}\frac{\alpha(1-3\beta^2)-\beta(1+\alpha^2)}{(1+\alpha^2)^2}E(\hat\beta-\beta)+ \tag{A.21}$$

$$+2\sigma_a^2\frac{(1+\alpha^2-2\alpha\beta)}{1-\beta^2}(1-\beta^2)\frac{\beta(1-3\alpha^2)-\alpha(1+\alpha^2)}{(1+\alpha^2)^3}E(\hat\alpha-\alpha).$$

From (5.8) we obtain ρ_1 and ρ_2 and as in the MA(2) case, we can obtain the partial derivatives of the alpha's relative to the rho's and proceed to calculate

$$E(\hat\alpha - \alpha) \simeq E(r_1-\rho_1)\frac{\partial\alpha}{\partial\rho_1} + E(r_2-\rho_2)\frac{\partial\alpha}{\partial\rho_2}.$$

$$E(\hat{\beta} - \beta) \simeq E(r_1 - \rho_1)\frac{\partial \beta}{\partial \rho_1} + E(r_2 - \rho_2)\frac{\partial \beta}{\partial \rho_2}. \tag{A.22}$$

As in Proposition 4.2, the expectations $E(r_i - \rho_i), i = 1,2$ are computed from the formulas in Fuller(1976), taking into account that:

$$\begin{aligned} Var(\tilde{\gamma}_0) &\simeq \frac{2}{T}\sum_j \gamma_j^2 = \frac{2\sigma_a^4}{T}\frac{1}{(1-\beta^2)^3}\{1 + 4\alpha^2 + \alpha^4 - 8\alpha\beta - 8\alpha^3\beta \\ &+ \beta^2 + 10\alpha^2\beta^2 + \alpha^4\beta^2 - 2\alpha^2\beta^4\} \end{aligned}$$

$$\begin{aligned} Cov(\hat{\gamma}_1, \hat{\gamma}_0) &\simeq \frac{2(T-1)}{T^2}\sum_{j=1}^{\infty} 2\gamma_j\gamma_{j-1} \\ &= \frac{4(T-1)}{T^2}\sigma_a^4\frac{(1-\alpha\beta)(\alpha-\beta)}{(1-\beta^2)^3}\{-\beta(1-\alpha\beta)(\alpha-\beta) \\ &+ (1-\beta^2)(1+\alpha^2-2\alpha\beta)\} \end{aligned}$$

and

$$\begin{aligned} Cov(\hat{\gamma}_2, \hat{\gamma}_0) &\simeq \frac{2(T-2)}{T^2}\sum_j \gamma_j\gamma_{j-2} \\ &= \frac{2(T-2)}{T^2}\sigma_a^4\frac{(1-\alpha\beta)(\alpha-\beta)}{(1-\beta^2)^3}\{(1-\beta^2)(1-\alpha\beta)(\alpha-\beta) \\ &- 2\beta(1+\alpha^2-2\alpha\beta)(1-\beta^2) + 2\beta^2(1-\alpha\beta)(\alpha-\beta)\} \end{aligned}$$

Making the necessary substitutions we get (5.10),with

$$\begin{aligned} P_{MM}(1,1) = &-\alpha + 7\alpha^3 + 2\alpha^4 - \alpha^5 - 8\alpha^6 - 3\alpha^7 - 2\alpha^8 - 2\alpha^9 + \beta \\ &- 14\alpha^2\beta - 4\alpha^3\beta - 16\alpha^4\beta + 24\alpha^5\beta + 6\alpha^6\beta + 12\alpha^7\beta \\ &+ 7\alpha^8\beta + 10\alpha\beta^2 - 2\alpha^2\beta^2 + 28\alpha^3\beta^2 - 26\alpha^4\beta^2 + 28\alpha^5\beta^2 \\ &- 14\alpha^6\beta^2 + 12\alpha^7\beta^2 + 2\alpha^8\beta^2 + 2\alpha^9\beta^2 - 7\beta^3 - 6\alpha^2\beta^3 \\ &+ 20\alpha^3\beta^3 - 40\alpha^4\beta^3 - 8\alpha^5\beta^3 - 90\alpha^6\beta^3 - 12\alpha^7\beta^3 - 17\alpha^8\beta^3 \\ &+ 7\alpha\beta^4 + 2\alpha^2\beta^4 - 19\alpha^3\beta^4 + 16\alpha^4\beta^4 + 117\alpha^5\beta^4 + 22\alpha^6\beta^4 \\ &+ 55\alpha^7\beta^4 + 16\alpha^2\beta^5 - 16\alpha^3\beta^5 - 24\alpha^4\beta^5 - 16\alpha^5\beta^5 - 72\alpha^6\beta^5 \\ &- 16\alpha^3\beta^6 + 8\alpha^4\beta^6 + 32\alpha^5\beta^6 \end{aligned}$$

References

Anderson,T.W.(1971). *The Statistical Analysis of Time Series*. New York: John Wiley and Sons.

Ansley,C.F. and Newbold,P.(1981). On the bias in estimates of forecast mean square error. *Journal of The American Statistical Association*,76,569-578.

Bhansali,R.J.(1981). Effects of not knowing the order of an autoregressive process on the mean squared error of prediction-I. *Journal of The American Statistical Association*,76,588-597.

Box,G.E.P. and Jenkins,G.M.(1976). *Time Series Analysis:Forecasting and Control*. San Francisco:Holden Day.

Brockwell,P.J. and Davis,R.A.(1991a). *Time Series:Theory and Methods*. New York: Springer Verlag.

Brockwell,P.J. and Davis,R.A.(1991b). *ITSM:An Interactive Time Series Modelling Package for the PC*. New York: Springer-Verlag

Cordeiro,G.M. and Klein,R.(1993). Bias correction of maximum likelihood estimates for ARMA models.To appear, *Probability and Statistics Letters*.

Durbin,J.(1959).Efficient estimation of parameters in moving average models. *Biometrika*,46,

306-316.

Fuller,W.A.(1976). *Introduction to Statistical Time Series*. New York:John Wiley and Sons.

Kendall,M.G.(1954). Note on bias in the estimation of autocorrelation. *Biometrika*, 41,403-404.

Kunitomo,N. and Yamamoto,T.(1985). Properties of predictors in misspecified autoregressive time series models. *Journal of The American Statistical Association*, 80,941-950.

Lysne,D. and Tjøstheim,D.(1987). Loss of spectral peakes in autoregressive spectral estimation. *Biometrika*,74,200-206.

Marriott,F.H. and Pope,J.A.(1954). Bias in the estimation of autocorrelation.

Biometrika,41,390-402.

Morettin,P.A.(1984). The Levinson algorithm and its applications in time series analysis. *International Statistical Revue*,52,83-92.

Paulsen,J. and Tjøstheim,D.(1985). On the estimation of residual variance and order in autoregressive time series. *The Journal of the Royal Statistical Society,Series B*, 47,216-228.

Priestley,M.B.(1981).*Spectral Analysis and Time Series*.London:Academic Press.

Pukkila, T. and Nyquist,H.(1985). On the frequency domain estimation of the innovation variance of a stationary univariate time series.*Biometrika*,72,317-323.

Shaman,P. and Stine,R.A.(1988). The bias of autoregressive coefficient estimators. *Journal of The American Statistical Association*,83,842-848.

Simonoff,J.S.(1993). The relative importance of bias and variability in the estimation of the variance of a statistic.*The Statistician*,42,3-7.

Stine,R.A. and Shaman,P.(1990).Bias of autoregressive spectral estimators. *Journal of The American Statistical Association*,85,1091-1098.

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.

Tjøstheim,D. and Paulsen,J.(1983). Bias of some commonly-used time series estimates. *Biometrika*,70,389-399.

Tuan,D.P.(1992). Approximate distribution of parameter estimators for first-order autoregressive models.*Journal of Time Series Analysis*,13,147-170.

Walker,A.M.(1962). Large-sample estimation of parameters for autoregressive processes with moving-average residuals.*Biometrika*,49,117-131.

Yamamoto,T. and Kunitomo,N.(1984). Asymptotic bias of the least squares estimators for multivariate autoregressive models.*Annals of the Institute of Statistical Mathematics*,36,419-430.

Wolfram,S.(1988). *Mathematica: A System for Doing Mathematics by Computer*.Reading, Addison-Wesley.

Table 1. Generated models, $\sigma_e^2 = 1$.

Parameters		Model I	Model II	Model III
AR(1)	β	-0.60	0.90	-0.30
AR(2)	β_1	-1.20	-0.40*	-0.95*
	β_2	0.30	0.70	0.90
MA(1)	α	-0.40	0.20	0.80
MA(2)	α_1	-1.10	1.10	0.30
	α_2	0.20	0.20	-0.60
ARMA(1,1)	β	0.30	-0.60	-0.70
	α	0.90	-0.80	0.40

* complex roots

Table 2. Estimated bias (with standard error) and asymptotic bias for moment, least squares and maximum likelihood estimators. AR(1) model.

T		$\beta = -0.6$			$\beta = 0.9$			$\beta = -0.3$		
		MM	LS	ML	MM	LS	ML	MM	LS	ML
50	EST.BIAS	-0.009	0.022	0.001	0.152*	0.024*	0.006	-0.007	0.035	0.015*
	(ST.ERROR)	(0.021)	(0.022)	(0.021)	(0.026)	(0.020)	(0.019)	(0.018)	(0.018)	(0.018)
	ASYM.BIAS	0.048	0.003	-0.040	0.492	0.150	-0.040	-0.008	-0.016	-0.040
100	EST.BIAS	0.018	0.027	0.017	0.094*	0.035	0.026	-0.004	0.018	0.006
	(ST.ERROR)	(0.016)	(0.016)	(0.016)	(0.017)	(0.014)	(0.014)	(0.013)	(0.013)	(0.013)
	ASYM.BIAS	0.024	0.001	-0.020	0.246	0.075	-0.020	-0.004	-0.008	-0.020
200	EST.BIAS	0.011	0.014	0.009	0.048*	0.016	0.011	-0.012	-0.001	-0.006
	(ST.ERROR)	(0.011)	(0.011)	(0.011)	(0.013)	(0.011)	(0.011)	(0.010)	(0.010)	(0.010)
	ASYM.BIAS	0.012	0.001	-0.010	0.123	0.038	-0.010	-0.002	-0.004	-0.010
400	EST.BIAS	0.007	0.009	0.006	0.023*	0.009	0.006	-0.013	-0.008	-0.010
	(ST.ERROR)	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)
	ASYM.BIAS	0.006	0.0003	-0.005	0.061	0.019	-0.005	-0.001	-0.002	-0.005

Table 3. Estimated bias (with standard error) and asymptotic bias for moment, least squares and maximum likelihood estimators. AR(2) model.

T		$\beta_1 = -1.2 \quad \beta_2 = 0.3$			$\beta_1 = -0.4 \quad \beta_2 = 0.7$			$\beta_1 = -0.95 \quad \beta_2 = 0.90$		
		MM	LS	ML	MM	LS	ML	MM	LS	ML
50	EST.BIAS	0.183*	0.017*	-0.031	0.043*	0.011	-0.027	0.429*	0.028*	-0.011
	(ST.ERROR)	(0.034)	(0.022)	(0.019)	(0.023)	(0.022)	(0.021)	(0.053)	(0.020)	(0.018)
	ASYM.BIAS	0.511	0.335	-0.060	0.125	0.028	-0.060	0.772	0.340	-0.060
100	EST.BIAS	0.103*	0.012*	-0.009	0.036	0.026	0.006	0.229*	0.016*	-0.001
	(ST.ERROR)	(0.021)	(0.014)	(0.014)	(0.018)	(0.016)	(0.016)	(0.025)	(0.013)	(0.013)
	ASYM.BIAS	0.255	0.167	-0.030	0.062	0.014	-0.030	0.386	0.170	-0.030
200	EST.BIAS	0.047*	-0.003*	-0.013	0.022	0.015	0.005	0.116*	0.006*	-0.007
	(ST.ERROR)	(0.012)	(0.009)	(0.009)	(0.012)	(0.011)	(0.011)	(0.015)	(0.012)	(0.010)
	ASYM.BIAS	0.127	0.083	-0.015	0.031	0.007	-0.015	0.193	0.085	-0.015
400	EST.BIAS	0.031*	0.001*	-0.004	0.015	0.011	0.006	0.052*	0.005*	-0.007
	(ST.ERROR)	(0.008)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.008)	(0.009)	(0.007)
	ASYM.BIAS	0.063	0.041	-0.008	0.015	0.003	-0.008	0.096	0.042	-0.008

Table 4. Estimated bias (with standard error) and asymptotic bias for moment, least squares and maximum likelihood estimators. MA(1) model.

T		α = −0.4			α = 0.2			α = 0.8		
		MM	LS	ML	MM	LS	ML	MM	LS	ML
50	EST.BIAS	-0.248*	-0.002	-0.022	-0.229*	0.034	0.014*	-0.174*	0.012*	-0.004
	(ST.ERROR)	(0.015)	(0.019)	(0.019)	(0.015)	(0.018)	(0.018)	(0.022)	(0.022)	(0.022)
	ASYM.BIAS	0.004	-0.020	-0.040	-0.013	0.005	-0.040	0.254	0.513	-0.040
100	EST.BIAS	-0.132*	0.008	-0.002	-0.133*	0.018	0.007	-0.090*	0.023*	0.013
	(ST.ERROR)	(0.012)	(0.013)	(0.013)	(0.013)	(0.014)	(0.013)	(0.015)	(0.016)	(0.016)
	ASYM.BIAS	0.002	-0.010	-0.020	0.007	0.002	-0.020	0.127	0.256	-0.020
200	EST.BIAS	-0.078*	-0.007	-0.012	-0.078*	-0.002	-0.007	-0.047*	0.014*	0.007
	(ST.ERROR)	(0.009)	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)
	ASYM.BIAS	0.001	-0.005	-0.010	0.003	0.001	-0.010	0.063	0.128	-0.010
400	EST.BIAS	-0.032	-0.002	-0.004	-0.046*	-0.007	-0.009	-0.022*	0.009*	0.007
	(ST.ERROR)	(0.012)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)
	ASYM.BIAS	0.000	-0.002	-0.005	0.002	0.0006	-0.005	0.031	0.064	-0.005

Table 5. Estimated bias (with standard error) and asymptotic bias for moment, least squares and maximum likelihood estimators. MA(2) model.

T		$\alpha_1 = -1.1 \quad \alpha_2 = 0.2$			$\alpha_1 = 1.1 \quad \alpha_2 = 0.2$			$\alpha_1 = 0.3 \quad \alpha_2 = -0.6$		
		MM	LS	ML	MM	LS	ML	MM	LS	ML
50	EST.BIAS	-0.146*	-0.013	-0.062	-0.136*	0.031	-0.032	-0.172*	0.016	-0.021
	(ST.ERROR)	(0.019)	(0.018)	(0.017)	(0.021)	(0.022)	(0.021)	(0.016)	(0.017)	(0.016)
	ASYM.BIAS	0.173		-0.060	-0.362		-0.060	0.198		-0.060
100	EST.BIAS	-0.069*	-0.023	-0.043*	-0.053*	0.026	0.004	-0.093*	0.001	-0.018
	(ST.ERROR)	(0.016)	(0.014)	(0.014)	(0.016)	(0.016)	(0.016)	(0.014)	(0.014)	(0.013)
	ASYM.BIAS	0.086		-0.030	-0.181		-0.030	0.099		-0.030
200	EST.BIAS	-0.044**	-0.009	-0.019	-0.020	0.023	0.012	-0.051*	-0.007	-0.016
	(ST.ERROR)	(0.010)	(0.010)	(0.010)	(0.010)	(0.013)	(0.013)	(0.011)	(0.011)	(0.011)
	ASYM.BIAS	0.043		-0.015	-0.090		-0.015	0.049		-0.015
400	EST.BIAS	-0.018*	-0.005	-0.010	-0.007*	0.011	0.006	-0.025*	-0.009	-0.013
	(ST.ERROR)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
	ASYM.BIAS	0.022		-0.008	-0.045		-0.008	0.025		-0.008

Table 6. Estimated bias (with standard error) and asymptotic bias for moment, least squares and maximum likelihood estimators. ARMA(1,1) model.

T		$\alpha = 0.9$ $\beta = 0.3$			$\alpha = -0.8$ $\beta = -0.6$			$\alpha = 0.4$ $\beta = -0.7$		
		MM	LS	ML	MM	LS	ML	MM	LS	ML
50	EST.BIAS	-0.013*	0.027	-0.030	0.014	0.018	-0.026	0.049*	0.039	-0.001*
	(ST.ERROR)	(0.019)	(0.019)	(0.018)	(0.023)	(0.022)	(0.021)	(0.018)	(0.018)	(0.016)
	ASYM.BIAS	-0.254		-0.060	-0.011		-0.060	0.218		-0.060
100	EST.BIAS	0.001*	0.020	-0.011	0.021	0.028	0.006	0.011*	0.016	-0.004
	(ST.ERROR)	(0.014)	(0.014)	(0.013)	(0.017)	(0.017)	(0.016)	(0.013)	(0.013)	(0.013)
	ASYM.BIAS	-0.127		-0.030	-0.006		-0.030	0.109		-0.030
200	EST.BIAS	-0.009*	0.004	-0.009	0.005	0.014	0.003	0.000*	0.006	-0.004
	(ST.ERROR)	(0.011)	(0.009)	(0.009)	(0.015)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
	ASYM.BIAS	-0.063		-0.015	-0.003		-0.015	0.054		-0.015
400	EST.BIAS	0.010*	0.002	-0.005	0.028	0.009	0.004	-0.006*	-0.003	-0.008
	(ST.ERROR)	(0.009)	(0.006)	(0.006)	(0.019)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
	ASYM.BIAS	-0.031		-0.008	-0.001		-0.008	0.027		-0.008

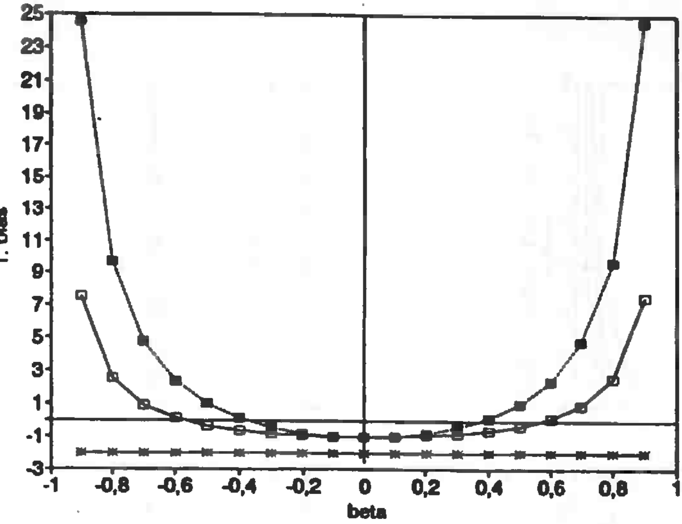

Figure 1. (a) Graph of T.E $(\widehat{\sigma_a^2} - \sigma_a^2)$, when $\sigma_a^2 = 1$, for AR(1); MM (■----■), LS (□—□) and ML (∗···∗)

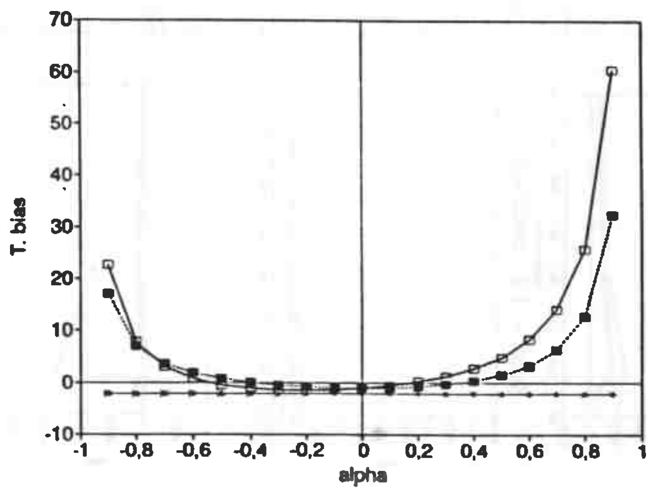

Figure 1. (b) Graph of T.E $(\widehat{\sigma_a^2} - \sigma_a^2)$, when $\sigma_a^2 = 1$, for MA(1); MM (■----■), LS (□—□) and ML (∗···∗)

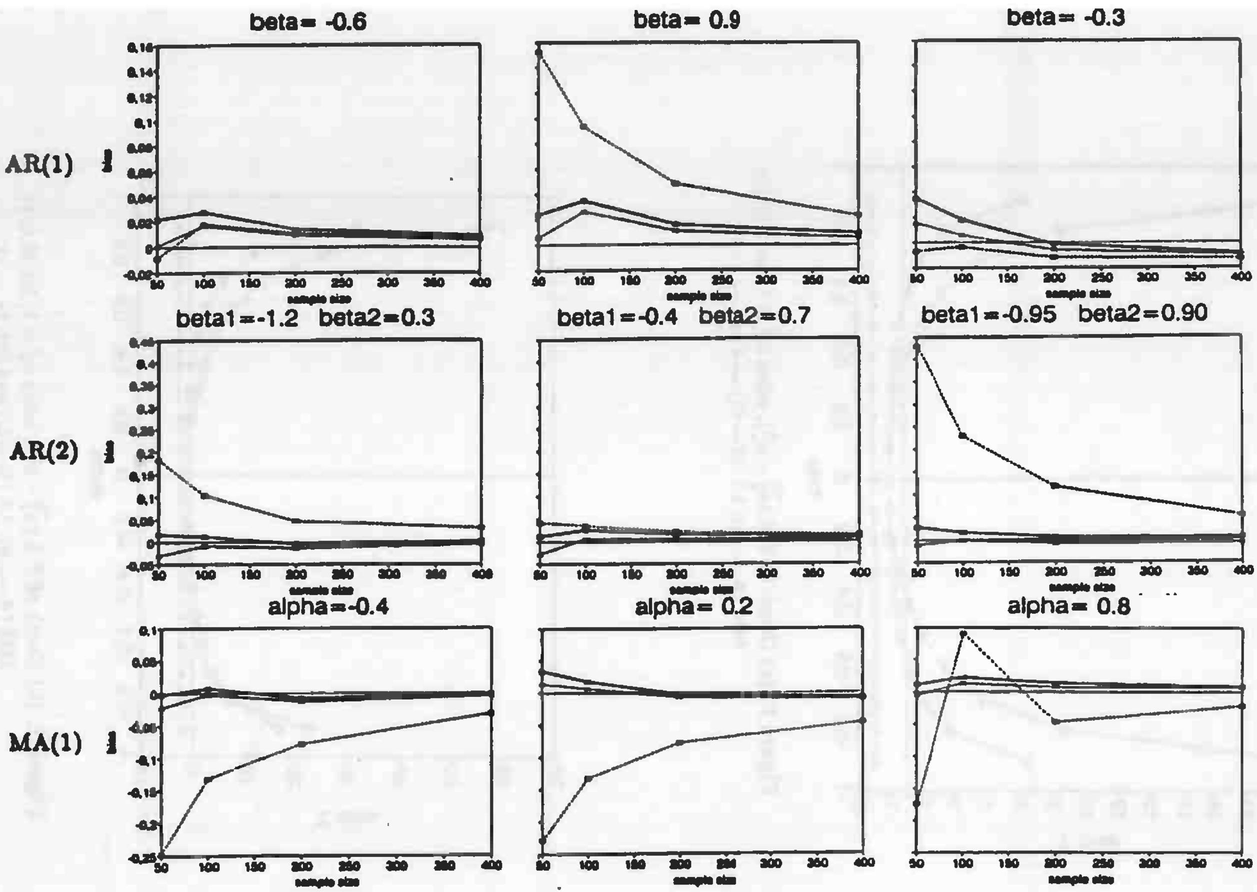

Figure 2. Estimated bias for method of moment (■---■), least squares (□—□) and maximum likelihood (∗···∗)

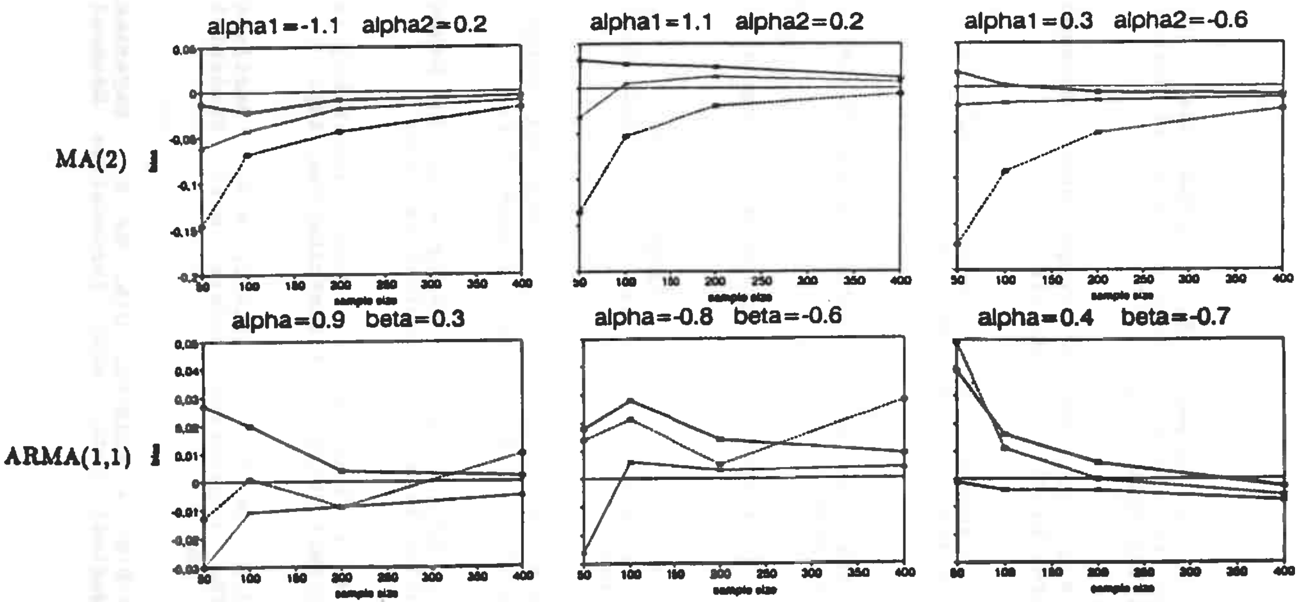

Figure 2. Continuation

ULTIMOS RELATORIOS TECNICOS PUBLICADOS

1993

9301 - BUENO, V.C. & ARIZONO, H. Comparisons for Maintenance Policies Involving Complete and Minimal Repair Through Compensator Transform, 12p.

9302 - BUENO, V.C. Maintenance Comparisons Through Compensator Transform: Block Policies, 13p.

9303 - FERRARI, P.A. & FONTES, L.R.G. Shock Fluctuations in the Asymmetric Simple Exclusion Process, 17p.

9304 - CORDEIRO, G. M., BOTTER, D.A. & FERRARI, S.L.P. Nonnull Asymptotic Distributions of Three Classic Criteria in Generalised Linear Models, 21p.

9305 - TANAKA, N.I. & TAVARES, H.R. Introdução à Teoria da Percolação, 38p.

9306 - BOLFARINE, H., ZACKS, S. & SANDOVAL, M.C. On Predicting the Population Total Under Regression Models with Measurement Erros, 10p.

9307 - FERRARI, P.A. & FONTES, L.R.G. The Net Output Process of a System with Infinitely Many Queues, 20p.

9308 - HO, L.L. & BUSSAB, W.O. Decomposição de Dados Categóricos com Auxílio de Árvores Dicotômicas, 23p.

9309 - BOLFARINE, H. & VALLE, R.B.A. Robust Modelling in Measurement Error Models Using the t Distribution, 14p.

9310 - FERRARI, S.L. de P. & ARELLANO, R.B. Bartlett Corrected Test for Regression Models with Student-t Independent Errors, 30p.

9311 - PAULINO, C.D.M. & PEREIRA, C.A. de B. Bayesian Methods for Categorical Data Under Informative General Censoring, 13p.

9312 - INOUE, L.Y.T., PASCHOA, O.E.D. & WECHSLER, S. Bayesian Assessment of Bioequivalence, 9p.

9313 - MADRUGA, M.R., PEREIRA, C.A. de B. & RABELLO-GAY, M.N. Bayesian Dosimetry: Radiation Dose Versus Frequencies of Cells with Aberrations, 15p.

The complete list of Relatórios do Departamento de Estatística, IME-USP, will be sent upon request.

- Departamento de Estatística
  IME-USP
  Caixa Postal 20.570
  01498-970 - São Paulo, Brasil