Lobato, I. N., & Velasco, C. (2018). Efficiency improvements for minimum distance estimation of causal and invertible ARMA models. *Economics Letters*, 162, pp. 150-152.

DOI: 10.1016/j.econlet.2017.11.013

Efficiency Improvements for Minimum Distance Estimation of Causal and Invertible ARMA Models*

Ignacio N. Lobato and Carlos Velasco

Centro de Investigación Económica, ITAM and Universidad Carlos III de Madrid

ilobato@itam.mx carlos.velasco@uc3m.es

October 30, 2017

Abstract

In this note we analyze efficiency improvements over the Gaussian maximum likelihood (ML) estimator for frequency domain minimum distance (MD) estimation for causal and invertible autoregressive moving average (ARMA) models. The analysis complements Velasco and Lobato (2017) where optimal MD estimation, which employs information in higher order moments, is studied for the general possibly non causal or nonivertible case. We consider MD estimation that combines in two manners the information contained in second, third, and fourth moments. We show that for both MD estimators efficiency improvements over the Gaussian ML occur when the distribution of the innovations is platykurtic. In addition, we show that asymmetry alone is not associated with efficiency improvements.

Keywords: higher-order moments; efficiency; kurtosis.

JEL codes: C22, C13

Corressponding Author: Ignacio Lobato

*Lobato acknowledges financial support from Asociación Mexicana de Cultura and from the Mexican Consejo Nacional de Ciencia y Tecnología (CONACYT) under project grant 151624. Velasco acknowledges financial support from the Ministerio Economía y Competitividad (Spain), grants ECO2012-31748, ECO2014-57007p and MDM 2014-0431, and Comunidad de Madrid, MadEco-CM (S2015/HUM-3444).

1. INTRODUCTION

In a recent article Velasco and Lobato (2017, VL hereinafter) propose optimal minimum distance techniques for estimating general (possibly non invertible and/or non causal) linear time series models. Non-causal or non-invertible ARMA models have been recently emphasized in economics, see Alessi, Barigozzi, and Cano (2011) or Leeper, Walter, and Yang (2013). VL propose frequency domain techniques to combine the information about the parameters of interest contained in second, third and fourth moments. A main objective in VL is the identification of the parameters of interest, an so, the information contained in third and fourth moments is crucial in the general case that allows for non causal or non invertible solutions. For efficiency, second, third and fourth moments are in principle relevant even under assumptions on causality and invertibility.

Estimation in VL is based on minimizing the distance between the model reflected in the (higher-order) spectral densities and the data reflected in their sample analogues, the (higher order) periodograms. VL considered two estimators. First, one that minimizes a weighted average of three individual objective functions (OF) based on the second, third and fourth moments, respectively. Second, an estimator that automatically combines the individual scores instead of the corresponding OF. The advantage of the first estimator is that it allows to derive explicit expressions for the optimal weights for each individual OF in the particular case of invertible and causal ARMA models. These optimal weights illuminate the cases where efficiency improvements over the Gaussian ML can be expected. The second estimator is theoretically optimal and, in addition, VL showed that it presents a better finite sample performance than the Gaussian maximum likelihood (ML) estimator in some situations. In particular, in VL it was noticed that in finite samples efficiency improvements are typically associated with innovations with platykurtic distributions. However, in VL there is not a theoretical basis to support this result. The reason is that this second estimator automatically combines the information contained in second, third and fourth moments, preventing the assessment of each individual contribution.

The structure of this note is the following. Section 2 contains notation and some relevant results from VL, and Section 3 provides the analysis on efficiency improvements associated to optimal MD weights.

2. PRELIMINARIES

Consider a causal and invertible ARMA(p, q) models

$$\alpha(L)Y_t = \mu + \beta(L)\varepsilon_t,$$

where ε_t is an independent identically distributed (iid) sequence with zero mean, variance $\kappa_2 > 0$, and bounded eighth moment. The polynomials $\alpha(L) = 1 - \sum_{j=1}^{p} \alpha_j L$ and $\beta(L) = 1 + \sum_{j=1}^{q} \beta_j L$ are of order p and q, respectively, have all their roots outside the unit circle, and do not have any common roots. This model establishes a structure on the linear dependence of Y_t in terms of some parameter vector $\theta \in \mathbb{R}^{p+q}$, and the target is the estimation of θ. VL consider the general case where the polynomials $\alpha(L)$ and $\beta(L)$ may also have some roots inside the unit circle. For completeness in notation we call κ_k to the k-th order cumulant of ε_t, and we also introduce the standardized cumulant of order k,

$$\nu_k = \frac{\kappa_k}{\kappa_2^{k/2}},$$

so that, ν_3 and ν_4 are the skewness and kurtosis coefficients, respectively. In addition, call $\bar{\mu}_4 = \mu_4/\kappa_2^2 = \nu_4 + 3$, $\bar{\mu}_5 = \mu_5/\kappa_2^{5/2} = \nu_5 + 10\nu_3$ and $\bar{\mu}_6 = \mu_6/\kappa_2^3 = \nu_6 + 10\nu_3^2 + 15\nu_4 + 15$ to the standardized fourth, fifth and sixth moments, respectively.

Following Brillinger (1985) and Terdik (1999), among others, VL considered employing MD estimators in the frequency domain. In particular, in order to make inference on the parameter vector θ, VL initially considered MD estimators based on the normalized distance between the data, reflected in the higher-order periodogram, and the model, reflected in the parameterization of the spectral density of order k. Denote by $L_{kT}^*(\theta)$ the individual OF based just on the information associated to the k-th cumulant.

VL considered two manners of combining the information contained in the higher order cumulants so that identification in the general non causal non invertible case, and potential efficiency improvements could be achieved.

The first approach is to construct a general OF by weighting the individual OF, $L_{kT}^*(\theta)$. In particular, VL initially propose as estimator of θ the minimizer of

$$L_{wT}^*(\theta) = w_2 L_{2T}^*(\theta) + w_3 L_{3T}^*(\theta) + w_4 L_{4T}^*(\theta),$$

where $\underline{w} = (w_2, w_3, w_4)'$ are some non negative weights to be chosen. This estimator, called θ_{wT}, is consistent as long as $\sum_{j=3}^{4} w_j \nu_j^2 > 0$. Note that this condition allows for some zero cumulants and weights. It is also asymptotically normal, and, in order to express its asymptotic variance, denote the asymptotic variance of the Gaussian ML estimator by Φ_0^{-1}, for an expression of this variance see for instance Theorem 10.8.2 in Brockwell and Davis (1991). Then, the asymptotic variance of θ_{wT} is given by $\Sigma_0^{-1}\Omega_0\Sigma_0^{-1}$ where

$$\Sigma_0 = \Sigma_0(\underline{w}) = \left(w_2 + w_3\nu_3^2 + w_4\nu_4^2\right)\Phi_0,$$

and

$$\Omega_0 = \Omega_0(\underline{w}) = \sum_{j,k=2}^{4} w_j w_k \mathbb{V}_{j-1,k-1},$$

where $\mathbb{V}_{j,k}$ denote the elements of the matrix $\mathbb{V}$, which is given by

$$\mathbb{V} = \begin{pmatrix} 1 & \nu_3^2 & \bar{\mu}_4\nu_4 \\ \nu_3^2 & \nu_3^2\left(2+\nu_4\right) & \nu_3\nu_4\left(\bar{\mu}_5-\nu_3\right) \\ \bar{\mu}_4\nu_4 & \nu_3\nu_4\left(\bar{\mu}_5-\nu_3\right) & \nu_4^2\left(\bar{\mu}_6-\nu_3^2\right) \end{pmatrix} \otimes \Phi_0.$$

Note that when $\underline{w} = (1,0,0)$, the asymptotic variance of θ_{wT} is Φ_0^{-1}. In fact, VL showed that the estimator based on minimizing $L_{2T}^*(\theta)$ is asymptotically equivalent to the Gaussian ML estimator. In addition, in VL it was noted that θ_{wT} can be asymptotically more efficient than the Gaussian ML estimator in some circumstances for some weighting schemes. The disadvantage of θ_{wT} is that it is impossible to derive optimal weights for the possibly non-invertible non-causal case because in that general case the asymptotic variance is not proportional to Φ_0^{-1}.

The second approach is to construct a MD estimator based on the three scores of the individual OF, $L_{kT}^*(\theta)$. This is the manner eventually recommended in VL. This estimator, called $\hat{\theta}_T$, is consistent and efficient. In particular, its asymptotic variance is given by $(\mathbb{H}'\mathbb{V}^-\mathbb{H})^{-1}$, where $\mathbb{V}^-$ is a generalized inverse of the matrix $\mathbb{V}$, and, for the causal and invertible case, $\mathbb{H} = \begin{pmatrix} 1, & \nu_3^2, & \nu_4^2 \end{pmatrix}' \otimes \Phi_0$. Remarks 4 and 5 in VL show that $\hat{\theta}_T$ is asymptotically at least as efficient as θ_{wT} for any choice of the weights $\underline{w}$.

From a practical point of view, the clear advantage of $\hat{\theta}_T$ is that there is no need of choosing explicitly the optimal weights for the scores since they are automatically optimally weighted. However, since optimal weights are only implicit there is not a clear understanding about under which circumstances efficiency improvements can be expected with respect to the Gaussian ML estimator. The simulations results in VL indicate that for platykurtic distributions the $\hat{\theta}_T$ estimator typically presents finite sample efficiency improvements respect to an (unfeasible) Gaussian ML estimator that presumes identification, whereas for leptokurtic distributions that advantage disappears. It also appears that asymmetry does not seem to play any significant role for efficiency in simulations.

3. OPTIMAL SELECTION OF WEIGHTS FOR INVERTIBLE AND CAUSAL ARMA MODELS

In this section we target to analyze the optimal weights for the first estimator proposed in VL. Given Theorem 3 in VL, it is possible to derive the optimal weights (w_2, w_3, w_4) in the sense of minimizing the asymptotic variance of θ_{wT} for a given value of the higher order cumulants $(\nu_3, \nu_4, \nu_5, \nu_6)$ when the model is invertible and causal, since in this case the problem reduces to a univariate one. This analysis is not reported here since the expressions for the optimal weights are complicated and not intuitive. Here, instead, we will provide three lemmas with the particular cases that occur when we set one weight equal to zero. This analysis is interesting since relatively simple and intuitive solutions are available for these estimators that employ only the information contained in two moments.

Lemma 1 considers the case of the optimal selection of w_2 and w_3 when w_4 is restricted to be zero. This result is of interest since it shows that adding $L^*_{3T}(\theta)$ to $L^*_{2T}(\theta)$ can not make any improvement from an efficiency point of view. Hence, asymmetry alone does not improve efficiency.

Lemma 2 addresses the case where w_3 is restricted to 0, for instance, due to anticipation of symmetric innovations**.** This case is of interest because for the platykurtic case ($\nu_4 < 0$), employing the optimal w_4 delivers an estimator whose asymptotic variance is lower than the one for the Gaussian ML estimator. The gain in efficiency depends on the specific values for the cumulants. In the other situation, the leptokurtic case ($\nu_4 > 0$), Lemma 2 shows that there can not be efficiency improvements with respect to the Gaussian ML estimator.

For completeness, Lemma 3 considers the case where w_2 is restricted to be 0.

Proofs of the lemmas are in the Appendix.

Lemma 1: Restricting to the weights $(w_2, w_3, w_4) = (1, w_3, 0)$, the optimal value for w_3 is $w_3^* = 0$.

Lemma 2: Restricting to the weights $(w_2, w_3, w_4) = (1, 0, w_4)$, if $\nu_4 \neq 0$, the optimal value for w_4 is

$$w_4^* = \max\left\{\frac{3}{\nu_4^2\bar{\mu}_4 - \nu_4\left(\bar{\mu}_6 - \nu_3^2\right)}, 0\right\}.$$

It is immediate to see that w_4^* is positive when $\nu_4 < 0$ because $\bar{\mu}_4 > 0$ and $\bar{\mu}_6 - \nu_3^2 > 0$, which implies that the information contained in fourth order cumulants should be employed if the distribution tails of the innovations are lighter than the Gaussian ones. In this case the asymptotic variance of θ_{wT} is lower than that of the Gaussian ML estimator. The gain in efficiency varies according to the distribution of ε_t, as commented above.

When $\nu_4 > 0$, there are not possible gains from using fourth order cumulants since for w_4^* to be positive one would need $\nu_4\left(\bar{\mu}_6 - \nu_3^2\right) < \nu_4^2\bar{\mu}_4$ or equivalently

$$\frac{\bar{\mu}_6 - \nu_3^2}{\bar{\mu}_4^2} < \frac{\nu_4}{\bar{\mu}_4} = \frac{\bar{\mu}_4 - 3}{\bar{\mu}_4},$$

where the right hand side is smaller than one. However, the left hand side can be shown to be no lower than 1 (e.g. Corollary 1 in VL) and, therefore, w_4^* can not be positive when $\nu_4 > 0$.

Lemma 3: Restricting to the weights $(w_2, w_3, w_4) = (0, 1, w_4)$, if $\nu_3\nu_4 \neq 0$, the optimal value for w_4 is

$$w_4^\dagger = \max\left\{\frac{\nu_3\nu_4\left(\bar{\mu}_4 - 1\right) - \nu_3^2\left(\bar{\mu}_5 - \nu_3\right)}{\nu_3\nu_4\left(\bar{\mu}_6 - \nu_3^2\right) - \nu_4^2\left(\bar{\mu}_5 - \nu_3\right)}, 0\right\}.$$

In principle $w_4^\dagger$ can be positive either when $\nu_4 < 0$ or when $\nu_4 > 0$, depending on the values of ν_3, ν_5 and ν_6. The weights $w_4^\dagger$ are positive, for instance, for the case $\nu_4 < 0$ when ν_3 and ν_5 have the same sign, since in this case both numerator and denominator are negative. In particular, if $\nu_3 > 0$ and $\nu_5 > 0$, the numerator can be written as

$$\nu_3 \left\{ \nu_4 \left(\bar{\mu}_4 - 1 \right) - 9\nu_3^2 \right\} - \nu_3^2 \nu_5 < 0,$$

because $\bar{\mu}_5 = \nu_5 + 10\nu_3$, and $\bar{\mu}_4 - 1 = \nu_4 + 2 > 0$ because $\nu_4 > -2$, while the denominator is

$$\nu_3 \left\{ \nu_4 \left(\bar{\mu}_6 - \nu_3^2 \right) - 9\nu_4^2 \right\} - \nu_4^2 \nu_5 < 0$$

because $\bar{\mu}_6 - \nu_3^2 > 0$ by Cauchy-Schwarz inequality.

REMARK: The previous lemmas showed that adding the information contained in third moments does not improve on efficiency with respect to the Gaussian ML estimator, but adding the information contained in fourth moments leads to a θ_{wT} estimator that is more efficient than the Gaussian ML estimator in the platykurtic case. Evaluation of the asymptotic variance of the efficient $\hat{\theta}_T$ estimator, $(\mathbb{H}'\mathbb{V}^{-}\mathbb{H})^{-1}$, at the specific situations of symmetric innovations ($\nu_3 = 0$) or zero excess kurtosis ($\nu_4 = 0$) provides similar messages. In particular, imposing $\nu_4 = 0$, then $(\mathbb{H}'\mathbb{V}^{-}\mathbb{H})^{-1} = \Phi_0^{-1}$, confirming Lemma 1 in that the value of ν_3 does not play any role for efficiency on its own. In addition, imposing $\nu_3 = 0$ on the asymptotic variance, $(\mathbb{H}'\mathbb{V}^{-}\mathbb{H})^{-1}$, leads to a factor in front of Φ_0^{-1} that is $(\bar{\mu}_6 - \bar{\mu}_4^2) / \left(\bar{\mu}_6 - 2\nu_4 \bar{\mu}_4 + \nu_4^2 \right)$. This factor is less than 1 when ν_4 is negative, corroborating the results in Lemma 2.

Appendix

Proof of Lemma 1. For this case the asymptotic variance of the θ_{wT} estimator is given by

$$\frac{1 + 2w_3\nu_3^2 + w_3^2\nu_3^2 \left(\nu_4 + 2 \right)}{1 + 2w_3\nu_3^2 + w_3^2\nu_3^4} \Phi_0^{-1}.$$

The factor in front of Φ_0^{-1} is always larger than 1 if $w_3 > 0$ because $\nu_4 + 2 = \bar{\mu}_4 - 1 \geq \nu_3^2 > 0$, see Rao (1973, p.143). □

Proof of Lemma 2. For this case the asymptotic variance is given by

$$\frac{1 + 2w_4\nu_4\bar{\mu}_4 + w_4^2\nu_4^2 \left(\bar{\mu}_6 - \nu_3^2 \right)}{1 + 2w_4\nu_4^2 + w_4^2\nu_4^4} \Phi_0^{-1}.$$

Minimizing the factor in front of Φ_0^{-1} with respect to w_4, provides the FOC (as long as $\nu_4 \neq 0$),

$$0 = \nu_4 - \bar{\mu}_4 + \bar{\mu}_4\nu_4^2 w_4 + \nu_3^2\nu_4 w_4 - \nu_4\bar{\mu}_6 w_4,$$

so that

$$w_4^* = \frac{-3}{\nu_4 \left(\bar{\mu}_6 - \nu_3^2 - \bar{\mu}_4^2 + 3\bar{\mu}_4 \right)} = \frac{-3}{\nu_4 \left(\bar{\mu}_6 - \nu_3^2 - \bar{\mu}_4\nu_4 \right)},$$

which is positive when $\nu_4 < 0$ because $\bar{\mu}_6 - \nu_3^2 > 0$ and $\bar{\mu}_4 \nu_4 < 0$. $\square$

Proof of Lemma 3. In this case the asymptotic variance is given by

$$\frac{\nu_3^2 \left(\bar{\mu}_4 - 1\right) + w_4^2 \nu_4^2 \left(\bar{\mu}_6 - \nu_3^2\right) + 2 w_4 \nu_3 \nu_4 \left(\bar{\mu}_5 - \nu_3\right)}{\left(\nu_3^2 + w_4 \nu_4^2\right)^2} \Phi_0^{-1}.$$

Minimizing the factor in front of Φ_0^{-1} with respect to w_4, provides the FOC:

$$-2\nu_3 \frac{\nu_4}{\left(\nu_3^2 + w_4 \nu_4^2\right)^3} \left(\nu_3^3 - \nu_3 \nu_4 - \nu_3^2 \bar{\mu}_5 - \nu_3 \nu_4^2 w_4 + \bar{\mu}_5 \nu_4^2 w_4 + \nu_3^3 \nu_4 w_4 + \bar{\mu}_4 \nu_3 \nu_4 - \nu_3 \nu_4 \bar{\mu}_6 w_4\right) = 0,$$

so that, when $\nu_3 \nu_4 \neq 0$, gives the expression for the optimal weights

$$w_4^{\dagger} = \frac{\nu_3 \nu_4 \left(\bar{\mu}_4 - 1\right) - \nu_3^2 \left(\bar{\mu}_5 - \nu_3\right)}{\nu_3 \nu_4 \left(\bar{\mu}_6 - \nu_3^2\right) - \nu_4^2 \left(\bar{\mu}_5 - \nu_3\right)},$$

when the right hand side is positive, and the lemma follows. $\square$

REFERENCES

Alessi L., Barigozzi, M. and M. Capasso (2011), “Non-Fundamentalness in Structural Econometric Models: A Review”, *International Statistical Review*, 79, 16-47.

Brillinger, D.R. (1985), "Fourier inference: some methods for the analysis of array and nongaussian series data", *Water Resources Bulletin*, 21, 744-756.

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

Leeper, E.M., Walter, T.B. and S. Yang (2013), “Fiscal Foresight and Information Flows", *Econometrica*, 81, 1115-1145.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, Wiley, New York.

Terdik, G. (1999), *Bilinear Stochastic Models and Related Problems of Nonlinear Time Series Analysis*, Springer, New York.

Velasco, C. and I.N. Lobato (2017), "Frequency Domain Minimum Distance Inference for Possibly Noninvertible and Noncausal ARMA models," forthcoming, *Annals of Statistics*, available at http://imstat.org/aos/future_papers.html.