Non redundancy of high order moment conditions for efficient GMM estimation of weak AR processes

Laurence Broze[a], Christian Francq[b], Jean-Michel Zakoïan[c,*]

[a] *GREMARS, Université de Lille 3 et CORE, Université catholique de Louvain*

broze@core.ucl.ac.be

[b] *LMPA J. Liouville, Université du Littoral–Côte d'Opale*

francq@lma.univ-littoral.fr

[c] *GREMARS, Université de Lille 3 et CREST*

zakoian@ensae.fr

Abstract

This paper considers GMM estimation of autoregressive processes. It is shown that, contrary to the case where the noise is independent (see Kim, Qian and Schmidt (1999)), using high-order moments can provide substantial efficiency gains for estimating the AR(p) model when the noise is only uncorrelated.

Keywords: Autoregressive process; Efficiency gains; GMM; Empirical autocorrelations; Yule-Walker estimator.

JEL classification: C13; C22

---

*We thank Luc Bauwens for useful comments.

1 Introduction

Since the seminal paper by Mann and Wald (1943), the estimation of autoregressive (AR) models has attracted the attention of a large number of statisticians and econometricians. In a recent paper, Kim, Qian and Schmidt (1999) considered minimum distance (MD) and generalized method of moments (GMM) estimation (see Hansen (1982)) of the univariate AR(p) process:

$$y_t = \phi_1^0 y_{t-1} + \phi_2^0 y_{t-2} + \cdots + \phi_p^0 y_{t-p} + \epsilon_t \tag{1}$$

where (ϵ_t) is a sequence of uncorrelated and centered variables, with common variance σ_ϵ^2. We assume that the parameter $\phi_0 = (\phi_1^0, \ldots, \phi_p^0)'$ satisfies the usual 'roots outside the unit circle' requirement to ensure stationarity. Kim et al. (1999) consider the MD estimator that minimizes the optimal quadratic form in the difference $\hat{\rho} - \rho(\phi)$, where $\rho(\phi_0)$ is the vector of the first m ($m \geq p$) population autocorrelations and $\hat{\rho}$ contains the corresponding sample autocorrelations. The two main results of Kim et al. paper are that: *i)* the MD estimator is asymptotically equivalent to the GMM estimator that asserts uncorrelatedness of y_{t-j} and ϵ_t, for $j = 1, \ldots, m$; *ii)* for $m > p$, the last $(m-p)$ moment conditions are redundant, in the sense that they do not increase asymptotic efficiency compared to the estimator using the first p moments (or autocorrelations). Moreover Kim et al. argue that these results do not depend on normality assumptions on the noise process (ϵ_t). An implicit assumption in their paper, however, is that the noise is a strong one, *i.e.* it is a sequence of independent variables. Our main objective in writing this note is to draw the attention to the following point: when the independence assumption does not hold (*i.e.* in the general case where the ϵ_t's are dependent though uncorrelated), the redundancy result can be in failure. In other words, *the additional $m - p$ moments can provide substantial asymptotic efficiency gains over the GMM estimator based on the first p moments.* This is an important point for practical applications since the 'strong linearity' of most economic series has been seriously questioned during the last twenty years. Despite the success of nonlinear time series models, the autoregressive processes can still provide good approximations of economic dynamics but it is necessary to relax the independence assumption on the noise. One reason for this, is that transformations commonly used in applied research, such as aggregation, marginalization of vector time series are well known to preserve linearity, but they fail to preserve *strong* linearity (see Francq and Zakoïan, 1998).

Instead of deriving a general theory of GMM estimation under weak assumptions, which is far beyond the scope of this paper, we shall take a simple example to illustrate our point. The advantage of this simplicity is that explicit calculations can be done.

2 GMM estimation of an AR(1) model

Consider the AR(1) model

$$y_t = \phi_0 y_{t-1} + \epsilon_t, \quad |\phi_0| < 1 \tag{2}$$

where $\epsilon_t = \eta_t \eta_{t-1}$ and (η_t) is a sequence of *independent* and identically distributed IID (0,1) variables with finite fourth-order moment $\kappa = E(\eta_t^4)$. We assume that $V(\eta_t^2) > 0$, i.e. $\kappa \neq 1$. The variables ϵ_t are clearly centered, uncorrelated with unit variance. Therefore (ϵ_t) is a white noise. But, for instance, the sequence of the squares is not uncorrelated: $\text{cov}(\epsilon_t^2, \epsilon_{t-1}^2) = E(\eta_t^2 \eta_{t-1}^4 \eta_{t-2}^2) - 1 = \kappa - 1 \neq 0$. Hence (ϵ_t) is not an *independent* white noise.

The parameter to be estimated is ϕ, with true value ϕ_0. Following the notations of Kim et al. (1999), let for $m \geq 0$, $y_t^{(m)} = (y_t, \ldots, y_{t-m})'$ and consider the moment conditions $E[h(y_t^{(m)}, \phi_0)] = 0$, for $m \geq 1$, where

$$h(y_t^{(m)}, \phi) = y_{t-1}^{(m-1)} \cdot (y_t - \phi y_{t-1}) = \begin{bmatrix} y_{t-1}(y_t - \phi y_{t-1}) \\ \vdots \\ y_{t-m}(y_t - \phi y_{t-1}) \end{bmatrix}.$$

For a set of T observations, $y_1, \ldots, y_T$, let $\overline{h}_T(\phi) = T^{-1} \sum_{t=m+1}^{T} h(y_t^{(m)}, \phi)$ and let Ω_T a non singular symmetric $m \times m$ matrix, converging in probability to a non-random positive definite matrix Ω. Then $\hat{\phi}^{(m)}$ is a GMM estimator of ϕ based on the moment conditions $E[h(y_t^{(m)}, \phi_0)] = 0$, if it is solution of

$$\hat{\phi}^{(m)} = \arg\min_{\phi \in]-1,1[} \overline{h}_T(\phi)' \Omega_T^{-1} \overline{h}_T(\phi).$$

For simplicity, consider the cases $m = 1$ and $m = 2$ for which closed form GMM estimators can be derived. Obviously, the GMM estimator based a single moment condition is the usual Yule-Walker estimator (and it is also the least-squares estimator) given by

$$\hat{\phi}^{(1)} = \frac{\sum_{t=2}^{T} y_t y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2}.$$

Moreover, straightforward calculations yield

$$\begin{aligned}
\hat{\phi}^{(2)} &= \Bigg[\omega_{T,11}\left(\sum_{t=3}^{T} y_t y_{t-1}\right)\left(\sum_{t=3}^{T} y_{t-1}^2\right) \\
&\quad +\omega_{T,12}\left\{\left(\sum_{t=3}^{T} y_t y_{t-1}\right)^2 + \left(\sum_{t=3}^{T} y_t y_{t-2}\right)\left(\sum_{t=3}^{T} y_{t-1}^2\right)\right\} \\
&\quad + \omega_{T,22}\left(\sum_{t=3}^{T} y_t y_{t-2}\right)\left(\sum_{t=3}^{T} y_t y_{t-1}\right)\Bigg] \\
\times &\quad \left[\omega_{T,11}\left(\sum_{t=3}^{T} y_{t-1}^2\right)^2 + 2\omega_{T,12}\left(\sum_{t=3}^{T} y_t y_{t-1}\right)\left(\sum_{t=3}^{T} y_{t-1}^2\right) + \omega_{T,22}\left(\sum_{t=3}^{T} y_t y_{t-1}\right)^2\right]^{-1} \\
&\quad +o_P(1)
\end{aligned}$$

where $(\omega_{T,ij})$ is the element on the ith row and jth column of Ω_T^{-1}. Hence it can be shown using standard arguments that

$$\sqrt{T}(\hat{\phi}^{(1)} - \phi_0) = \frac{\frac{1}{\sqrt{T}}\sum_{t=2}^{T} \epsilon_t y_{t-1}}{\frac{1}{T}\sum_{t=2}^{T} y_{t-1}^2}$$

has an asymptotic centered normal distribution, with asymptotic variance given by $V_{\mathrm{as}}\{\sqrt{T}(\hat{\phi}^{(1)} - \phi_0)\} = I_{11}J^{-2}$ where for all positive integers i and j, $I_{i,j} = \sum_{k=-\infty}^{+\infty} \mathrm{cov}(\epsilon_t y_{t-i}, \epsilon_{t-k} y_{t-j-k})$ and $J = E(y_t^2) = 1/(1-\phi_0{}^2)$.

To compute the sums $I_{i,j}$ we define the cumulants $C(i,j,k) = E(\epsilon_t \epsilon_{t-i} \epsilon_{t-j} \epsilon_{t-k})$. For $0 \le i \le j \le k$, it is clear that $C(i,j,k) = 0$ unless $i = 0$ and $j = k$. Moreover, we have

$$C(0,0,0) = \kappa^2, \quad C(0,1,1) = \kappa \quad \text{and} \quad C(0,k,k) = 1, \quad \text{for} \quad k > 1.$$

Therefore, using the expansion $y_t = \sum_{\ell=0}^{+\infty} \phi_0^{\ell} \epsilon_{t-\ell}$, we have

$$I_{i,j} = \sum_{k=-\infty}^{+\infty} \sum_{\ell,n=0}^{+\infty} \phi_0^{\ell+n} E(\epsilon_t \epsilon_{t-i-\ell} \epsilon_{t-k} \epsilon_{t-j-k-n}) = \sum_{\ell,n=0}^{+\infty} \phi_0^{\ell+n} C(0, i+\ell, j+n).$$

Hence

$$I_{1,1} = C(0,1,1) + \sum_{\ell=1}^{+\infty} \phi_0^{2\ell} = \kappa + \frac{\phi_0^2}{1-\phi_0^2} \ ,$$

and for $i = \max(i,j) > 1$

$$I_{i,j} = \sum_{\ell=0}^{+\infty} \phi_0^{2\ell+i-j} = \frac{\phi_0^{i-j}}{1-\phi_0^2} \ .$$

Thus

$$V_{\mathrm{as}}\{\sqrt{T}(\hat{\phi}^{(1)} - \phi_0)\} = (1-{\phi_0}^2)\{1+(1-\kappa)({\phi_0}^2-1)\}. \tag{3}$$

Note in particular that this asymptotic variance is always larger than $1-{\phi_0}^2$, which corresponds to the case where the ϵ_t's are i.i.d. We next derive the asymptotic distribution of $\sqrt{T}\hat{\phi}^{(2)}$. Remark that

$$\sqrt{T}(\hat{\phi}^{(2)} - \phi_0) = \lambda_T \frac{1}{\sqrt{T}} \sum_{t=3}^{T} \epsilon_t y_{t-1} + \mu_T \frac{1}{\sqrt{T}} \sum_{t=3}^{T} \epsilon_t y_{t-2} + o_P(1),$$

where $\lambda_T :=$

$$\frac{\omega_{T,11}\left(\frac{1}{T}\sum_{t=3}^{T} y_{t-1}^2\right) + \omega_{T,12}\left(\frac{1}{T}\sum_{t=3}^{T} y_t y_{t-1}\right)}{\omega_{T,11}\left(\frac{1}{T}\sum_{t=3}^{T} y_{t-1}^2\right)^2 + 2\omega_{T,12}\left(\frac{1}{T}\sum_{t=3}^{T} y_t y_{t-1}\right)\left(\frac{1}{T}\sum_{t=3}^{T} y_{t-1}^2\right) + \omega_{T,22}\left(\frac{1}{T}\sum_{t=3}^{T} y_t y_{t-1}\right)^2}$$

and μ_T is obtained by replacing $\omega_{T,11}$ by $\omega_{T,12}$ and $\omega_{T,12}$ by $\omega_{T,22}$ in the numerator of λ_T. Hence, if we set $(\omega_{ij}) = \Omega^{-1}$, (λ_T, μ_T) converges in probability to

$$(\lambda, \mu) := \left(\frac{(\omega_{11} + \phi_0\omega_{12})(1-\phi_0^2)}{\omega_{11} + 2\omega_{12}\phi_0 + \omega_{22}\phi_0^2}, \quad \frac{(\omega_{12} + \phi_0\omega_{22})(1-\phi_0^2)}{\omega_{11} + 2\omega_{12}\phi_0 + \omega_{22}\phi_0^2} \right).$$

Furthermore we have the following convergence in distribution

$$\begin{pmatrix} \frac{1}{\sqrt{T}} \sum_{t=3}^{T} \epsilon_t y_{t-1} \\ \frac{1}{\sqrt{T}} \sum_{t=3}^{T} \epsilon_t y_{t-2} \end{pmatrix} \Longrightarrow \mathcal{N}\left(0, \begin{bmatrix} I_{11} & I_{12} \\ I_{12} & I_{22} \end{bmatrix}\right) := \mathcal{N}(0, \Sigma).$$

Finally, $\sqrt{T}(\hat{\phi}^{(2)} - \phi_0)$ has an asymptotic centered normal distribution, with $V\text{as}\{\sqrt{T}(\hat{\phi}^{(2)} - \phi_0)\} = (\lambda, \mu)\Sigma(\lambda, \mu)'$. Now we consider the optimal choice of the matrix Ω, which is given by

$$\Omega^* = V\text{as}\{\sqrt{T}\overline{h}_T(\phi_0)\} = \Sigma.$$

Simple computations show that

$$\Omega^{*-1} = \Sigma^{-1} = \frac{1}{\kappa}\begin{pmatrix} 1 & -\phi_0 \\ -\phi_0 & \kappa(1-\phi_0^2) + \phi_0^2 \end{pmatrix}.$$

Hence, (λ, μ) takes the value

$$(\lambda^*, \mu^*) = \left(\frac{1 - {\phi_0}^2}{1 - {\phi_0}^2 + \kappa\phi_0^2}, \frac{(1 - {\phi_0}^2)\phi_0(\kappa - 1)}{1 - {\phi_0}^2 + \kappa\phi_0^2}\right).$$

Finally, the asymptotic variance of the optimal GMM estimator based on the moment condition $E[h(y_t^{(2)}, \phi_0)] = 0$ is given by

$$V\text{as}\{\sqrt{T}(\hat{\phi}^{(2)} - \phi_0)\} = (\lambda^*, \mu^*)\Sigma(\lambda^*, \mu^*)' = \frac{\kappa(1 - {\phi_0}^2)}{1 + {\phi_0}^2(\kappa - 1)}. \quad (4)$$

In view of (3) and (4) we have

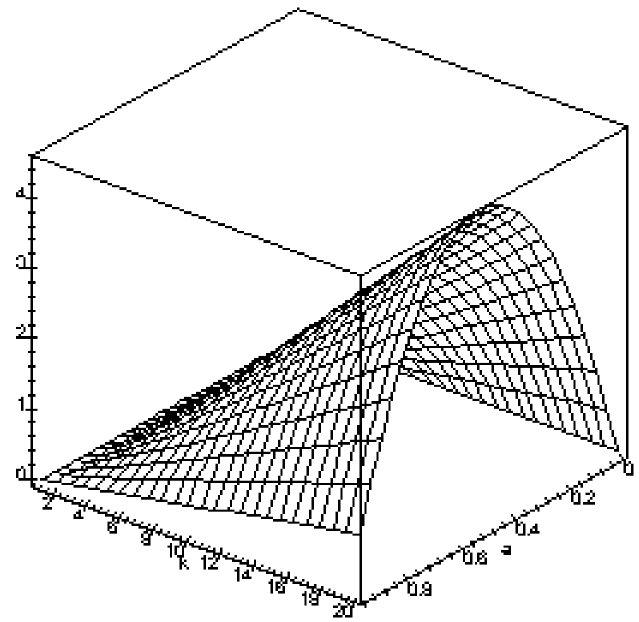
$$V\text{as}\{\sqrt{T}(\hat{\phi}^{(1)} - \phi_0)\} - V\text{as}\{\sqrt{T}(\hat{\phi}^{(2)} - \phi_0)\} = \frac{(1-\kappa)^2(1 - {\phi_0}^2)^2{\phi_0}^2}{1 + {\phi_0}^2(\kappa - 1)}$$

which confirms that the GMM estimator based on two moments strictly dominates, asymptotically, the GMM estimator based on the first moment condition unless when $\phi_0 = 0$. We have reported in Figure 1 the quantity $\frac{V\text{as}\{\sqrt{T}(\hat{\phi}^{(1)} - \phi_0)\}}{V\text{as}\{\sqrt{T}(\hat{\phi}^{(2)} - \phi_0)\}} - 1$ as a function of κ and ϕ_0^2. The figure clearly indicates that when the distribution of (η_t) has fat tails ($\kappa \to \infty$), the efficiency gains obtained from using two moments can be very high.

3 Concluding remarks

The above analysis suggests that, even if p moment conditions are sufficient to consistently estimate an AR(p) model, using additional moments can provide substantial gains in terms of asymptotic accuracy. Similar results are well-known to hold for MA or ARMA processes, even when the noise is independent. In the autoregressive case, the redundancy property crucially relies on independence. This should have empirically relevant implications in situations where only the uncorrelatedness of the noise can be assumed.

Figure 1: $\dfrac{V\mathrm{as}\{\sqrt{T}(\hat{\phi}^{(1)}-\phi_0)\}}{V\mathrm{as}\{\sqrt{T}(\hat{\phi}^{(2)}-\phi_0)\}}-1$ as a function of κ and ϕ_0^2.



References

Francq, C., Zakoïan, J-M., 1998. Estimating Linear Representations of Nonlinear Processes. Journal of Statistical Planning and Inference 68, 145–165.

Hansen, L.P., 1982. Large Sample Properties of Generalized Method of Moments Estimators. Econometrica 50, 1029–1054.

Kim, Y., Quian, H., Schmidt, P., 1999. Efficient GMM and MD estimation of autoregressive models. Economics Letters 62, 265–270.

Mann, H.B., Wald, A., 1943. On the statistical treatment of linear stochastic differential equations. Econometrica 11, 173–200.