Quaderni di Statistica
Vol. 2, 2000

The Variance of CLS estimators for a simple Bilinear model

Francesco Giordano
Dipartimento di Scienze Economiche, Università degli Studi di Salerno
E-mail: giordano@unisa.it

Summary: The aim of the paper is to get the asymptotic variance of the CLS estimators in a particular bilinear model, improving the approximation given in Giordano and Vitale (2000) and evaluating the performance of this variance with bands for inference about the unknown parameters.

Keywords: CLS estimators, ergodicity, asymptotic variance.

1. Introduction

This paper follows another one submitted to SIS 2000 by Giordano and Vitale. The problem, to be investigated, is to improve the approximation of the variance for Conditional Least Squares (CLS) estimators and evaluate the degree of convergence. The model is a particular bilinear one, that is:

$$X_t = b\boldsymbol{e}_{t-1} X_{t-2} + \boldsymbol{e}_t \qquad (1)$$

where

$\boldsymbol{e}_t \sim N(0, \boldsymbol{s}^2)$ $\qquad \forall t \qquad$ Cov($\boldsymbol{e}_t, \boldsymbol{e}_s$)=0 $\qquad t \neq s$

b and s^2 are the unknown parameters to estimate.

This model belongs to the "type I" of Grahn (1995). The new method (CLS) to estimate the parameters, introduced by Grahn, is appealing for the fast computation respect to the classical Maximum Likelihood by Subba Rao (1981). The CLS estimators, under mild conditions, have good properties for the strong consistency and limiting distribution. The main problem is to make inference for the unknown parameters, given that it is particularly difficult to get an explicit formulation for the variance of these estimators.

In this paper the objective is to analyse the model (1) for the variance of the CLS estimator for the parameter b.

The paragraph 2 gives the theoretical background to build the asymptotic variance; the third paragraph shows, by a simulation, the performance of this variance using asymptotic bands for the unknown parameter b; finally the conclusions and remarks are in paragraph 4.

2. Main results

Following Grahn (1995) the CLS estimator for b is:

$$\hat{b} = \frac{\sum_{t=3}^{N} X_t X_{t-1} X_{t-2}}{\sum_{t=3}^{N} X_{t-2}^2} / \hat{s}^2$$

where the estimator $\hat{s}^2$ is the intercept in the solution of this linear regression problem:

$$\sum_{t=3}^{N} \left\{ X_t^2 - (s^2 + b_{12} X_{t-2}^2) \right\}^2$$

with N the number of observations and $\boldsymbol{b}_{12}$ another parameter to estimate.

Grahn (1995) shows that $\sqrt{N}(\hat{b}-b)$ converges to a Normal distribution if $E(X_t^8)<\infty$.

If $E(X_t^4)<\infty$ then $\hat{b}$ converges, with probability one, to b with the law of Iterated Logarithm.
Given the model (1) it can be shown:

$b^2 \boldsymbol{s}^2 < 1$	$\Rightarrow$	X_t is Stationary, Ergodic and Causal;
$b^2(1+\boldsymbol{s}^2)<1$	$\Rightarrow$	X_t is invertible;
$b^4 \boldsymbol{s}^4 < 1/3$	$\Rightarrow$	$E(X_t^4)<\infty$ (Strong Consistency);
$b^8 \boldsymbol{s}^8 < 1/105$	$\Rightarrow$	$E(X_t^8)<\infty$ (Normality).

The moments of X_t have the following values:

- $E(X_t) = 0$ $\quad \boldsymbol{m}_2^X \equiv \mathrm{Var}(X_t) = \dfrac{\boldsymbol{s}^2}{1-b^2\boldsymbol{s}^2}$ $\quad \forall t$;
- $\mathrm{Cov}(X_t, X_s)=0$ $\quad \forall t \neq s$;
- $E(X_t, X_{t-1}, X_{t-2}) = b\boldsymbol{s}^2 \boldsymbol{m}_2^X$ $\quad \forall t$.

Let $\tilde{b}$ be another estimator for b, given by

$$\tilde{b} = \frac{\left(\sum_{t=3}^{N} X_t X_{t-1} X_{t-2}\right)/N}{\boldsymbol{s}^2 \boldsymbol{m}_2^X}.$$

The estimator $\tilde{b}$ is equal to $\hat{b}$ if $\boldsymbol{s}^2$ and $\boldsymbol{m}_2^X$ are known.

It is easy to show that $E(\tilde{b}) = \dfrac{N-2}{N} b$, that is asymptotically correct.

Lemma 1.

If X_t is as in model (1) and $b^4\sigma^4 < 1/3$ then

$\tilde{b} \xrightarrow{wp1} b$ *and* $|\tilde{b} - b| = O(L_N)$

where $L_N = \left(\dfrac{N}{\log\log N}\right)^{-1/2}$

with N>exp(1) (law of Iterated Logarithm).

Proof.

The condition above $b^4\sigma^4 < 1/3$ implies $b^2\sigma^2 < 1$ then X_t is stationary, ergodic and causal. Now the empirical mixed moment

$$\hat{M}_3 = \frac{\sum_{t=3}^{N} X_t X_{t-1} X_{t-2}}{N}$$

is such that

$$\lim_{N\to\infty} \hat{M}_3 = M_3 \equiv \mathrm{E}(X_t X_{t-1} X_{t-2}) = b\sigma^2 \mu_2^X \qquad \text{by ergodicity.}$$

Then $\tilde{b} \xrightarrow{wp1} b$.

To prove the convergence for the order of Iterated Logarithm, it is sufficient to use Theorem 3.1 in Grahn (1995). ♦

Another auxiliary result is necessary. This is stated in:

Lemma 2.

If X_t is as in model (1) and $b^4\sigma^4 < 1/3$ then

$|\tilde{b} - \hat{b}| \xrightarrow{wp1} 0$ *and* $|\tilde{b} - \hat{b}| = O(L_N)$ *at most*.

Proof.

From Grahn, 1995 follows that $|\hat{b} - b| \xrightarrow{wp1} 0$ with $O(L_N)$ and from lemma 1 above $|\tilde{b} - b| \xrightarrow{wp1} 0$ with $O(L_N)$.

Then $|\tilde{b} - \hat{b}| \le |\tilde{b} - b| + |\hat{b} - b|$.

But, in this case, the *wp1* convergence is equivalent to L^2 convergence as in Liu and Brockwell, 1988 and Grahn, 1995 looking at the random variables (mixed moments) in C[0,1] space that is the space of continuous functions defined on the interval [0,1].

Hence it is sufficient to prove that $\int \left(\left|\hat{b} - b\right| + \left|\tilde{b} - b\right| \right)^2 dP(X) \to 0$ when $n \to \infty$.

Here $P(X)$ is the distribution function of X.

As $\int \left(\hat{b} - b\right)^2 dP(X) \to 0$ and $\int \left(\tilde{b} - b\right)^2 dP(X) \to 0$ when $n \to \infty$ and applying the Schwarz inequality for $\int \left|\hat{b} - b\right| \left|\tilde{b} - b\right| dP(X)$ then the result holds.

For the order of the convergence, the result is immediate from the arguments above with the theorem 3.1 of Grahn, 1995. ♦

Now it is possible to state the following theorem:

Theorem.

If X_t is as in model (1) and $b^8 \boldsymbol{s}^8 < 1/105$ then

$$\sqrt{N}\left(\hat{b} - b\right) \xrightarrow{D} N(0, V)$$

where

$$V = \frac{\boldsymbol{s}^2}{m_2^X}\left(1 + 22b^2 \boldsymbol{s}^2 + 9b^2 \boldsymbol{s}^4 - 6b^2 m_2^X\right) + o(b^2 \boldsymbol{s}^2)$$

Proof.

From lemma 1 and lemma 2 it follows that $\sqrt{N}(\hat{b} - b)$ has the same asymptotic distribution as $\sqrt{N}(\tilde{b} - b)$. Then, under the above hypothesis,

$$\sqrt{N}(\hat{b} - b) \xrightarrow{D} N(0, V_1) \text{ and } \quad \sqrt{N}(\tilde{b} - b) \xrightarrow{D} N(0, V_2).$$

It implies that $V_1=V_2\equiv V$.
Let $Y_t= X_tX_{t-1}X_{t-2}$.

To get the formula above for V, it is necessary to make a direct computation of

$$\frac{1}{N}\mathrm{V}ar\left[\left(\sum_{\mathrm{t}} Y_t\right)\right].$$

After some algebraic manipulations, the variance above is:

$$\frac{1}{N}\mathrm{V}ar\left[\left(\sum_{\mathrm{t}} Y_t\right)\right] = \mathrm{V}ar(Y_t)+2\mathrm{Cov}(Y_{t+1},Y_t)+2\mathrm{Cov}(Y_{t+2}, Y_t)+$$
$$+2\mathrm{Cov}(Y_{t+3}, Y_t)+2\mathrm{Cov}(Y_{t+4}, Y_t)+o(b^2 s^2)$$

given that $\mathrm{Cov}(Y_{t+s}, Y_t)= o(b^2 s^2)$ for $s>4$

Finally, expanding the above operators, the result holds. ♦

The result for the asymptotic variance of $\hat{b}$ is exact unless the terms with a power greater than $b^2 s^2$ at least for b. In this way there is an improvement with respect to the approximation for the variance of $\hat{b}$ in Giordano and Vitale, 2000, where a sixth mixed moment is approximated with a univariate moment of order six.

3. Numerical results

To give an idea of the performance of this variance of $\hat{b}$, an experiment of simulation is to be done. The aim is to analyse the actual coverage of the bands against the nominal one. Besides, the average length of the bands is returned. All that is made in order to evaluate the different performances of the variance V in the theorem above, and of the approximation given in the paper of Giordano and Vitale, 2000.

The structure of the simulation is organised with s^2 fixed to 1 and the number of iterations equals to 1000. For each iteration there is a generation of a series with length (N) of 200 observations with b fixed. Then the CLS method is used to obtain the estimators $\hat{b}$ and $\hat{s}^2$. Finally the bands are built as

$$\hat{b} + 1.96 * \sqrt{Var(\hat{b})} \qquad \text{and} \qquad \hat{b} - 1.96 * \sqrt{Var(\hat{b})}\,.$$

The nominal coverage is fixed to 95%. Let $V_{(2)}$ be the variance of $\hat{b}$ above and $V_{(1)}$ be the variance of $\hat{b}$ in Giordano and Vitale, 2000. Recall that

$$V_{(1)} = \frac{1}{N}\frac{1}{s^2(1-15b^6s^6)}\left[\frac{1-b^2s^2}{1-3b^4s^4}(183b^6s^6 + 42b^4s^4\right.$$

$$\left. + 14b^2s^2 + 1)\right]$$

The table below shows the results.

b	Actual number	Actual coverage $V_{(1)}$ %	Average Length $V_{(1)}$	Actual coverage $V_{(2)}$ %	Average Length $V_{(2)}$
0.1	1000	98.10	0.3088	97.70	0.3221
-0.1	1000	97.70	0.3067	97.30	0.3189
0.2	997	96.29	0.3661	96.09	0.3858
-0.2	998	95.79	0.3608	95.49	0.3820
0.4	910	93.52	0.5337	95.15	0.4737
-0.4	920	91.74	0.5416	94.78	0.4806
0.6	624	81.89	0.4676	84.30	0.3641
-0.6	612	81.37	0.4648	83.01	0.3568

The first column contains the different values of the parameter *b* in model (1). The second shows the actual number of iterations which respects the condition of normality. The other columns are referred to the actual coverage and the average length of the bands for the variance $V_{(1)}$ and $V_{(2)}$, respectively.

The most important behaviour is to be discovered in the "Actual coverage" for $V_{(1)}$ respect to the same for $V_{(2)}$. In fact it is clear that small values of *b* bear to an overestimate of the variance. However, the overall performance is better for $V_{(2)}$. rather than for $V_{(1)}$. Instead for great values of *b* there is an underestimate of the variance. As in the previous case, $V_{(2)}$ still performs better than $V_{(1)}$.

Another limitation for $V_{(1)}$ is the missing in computation of the effect due to covariances. This explains the behaviour observed in the table above from the "Actual coverage" and "Average length".

4. Conclusions and remarks

The simple model in (1) is appealing because it looks like a White Noise if we consider only the first and second moments as seen above. So this model can be fitted to the residuals of other linear or non linear models for time series in order to capture, for example, the skewness or kurtosis.

This paper can be considered as a step to understand the techniques underlying the bilinear model. Projects for future research include the overcoming of the problems related to the estimate of the variance for the CLS estimators in a more general model.

Acknowledgements: The paper is supported by MURST98 "Modelli statistici per l'analisi delle serie temporali".

References

Bera A. K., Higgins M. L. (1997) Arch and Bilinearity as Competing Models for Nonlinear Dependence, *Journal of Business & Economic Statistics*, 15, 43-50

Brillinger D. R. (1981) *Time Series: Data Analysis and theory*, Holt, Rinehart & Winston, New York

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

Giordano, F. and Vitale C. (2000), Intervalli di confidenza asintotici per modelli bilineari in serie storiche "*Atti della XL Riunione Scientifica SIS*", 623-626

Grahn T. (1995) A Conditional Least Squares approach to Bilinear Time Series Estimation, *Journal of Time Series Analysis*, 16, 509-529

Granger C. W., Andersen A. P. (1978) *An introduction to bilinear time series models*, Vandehoek and Ruprecht, Gottigen

Guegan D. (1987) Different Representations for Bilinear models, *Journal of Time Series Analysis*, 8, 389-408

Liu J. and Brockwell, P. (1988) On the general time series model, *J. Appl. Prob.* 25, 553-564

Liu S. I. (1985) Theory of Bilinear Time Series models, *Comm. Statist. Theor. Meth.*, 14, 2549-2561

Pham D. T. (1985) Bilinear Markovian Representation and Bilinear models, *Stochastic Processes and their Applications*, 20, 295-306

Subba Rao T. (1981) On the Theory of Bilinear Time Series models, *J. R. Statist. Soc. B*, 43, 244-255

Subba Rao T., da Silva M. E. A. (1992) Identification of Bilinear Time Series models BL(p,0,p,1), *Statistica Sinica*, 2, 465-478

Tong H. (1990) *Non – linear Time Series*, Oxford University Press, Oxford