Proceedings of the World Congress on Engineering 2007 Vol II
WCE 2007, July 2 - 4, 2007, London, U.K.

Procedures of Parameters'estimation of AR(1) models into lineal state-space models

Rouhia NOOMENE
rouhia.noomene@upc.edu
Polytechnical University of Catalonia
Department of Statistics and Operational Research *

Abstract—**The objective of this paper is to study how algorithms of optimization affect the parameters-estimation of Autoregressive AR(1)Models. In our research we have represented the AR(1) models in linear state space form and applied the Kalman Filters to estimate the different unknown parameters of the model. Many methods have been proposed by researchers for the estimation of the parameters in the case of the linear state space models. In our work we have emphasized on the estimation through the Maximum Likelihood (ML). Statisticians have used many algorithms to optimise the likelihood function and they have proposed many filters; publishing their results in many papers. In spite of the fact that this field is so extended, we have emphasized our study in the financial field. Two quasi-Newton algorithms: Berndt, Hall, Hall, and Hausman (BHHH) and Broyden-Fletcher-Goldfarb-Shanno (BFGS), and the Expectation-Maximization (EM) algorithm have been chosen for this study. A practical study of these algorithms applied to the maximization of likelihood by means of the Kalman Filter have been done. The results are focused on efficiency in time of computation and precision of the unknown parameters estimation. A simulation study has been carried out, using as true values the parameters of this model published in the literature, in order to test the efficiency and precision of our implemented algorithms.**

Keywords: State space model, Kalman filer, maximum likelihood, BHHH, BFGS and EM

1 Introduction

The class of state-space models provides a flexible framework for modelling and describing a wide range of time series in a variety of disciplines. In the lineal and Gaussian case, many studies tried to derive the Kalman filter algorithm, such as derivations based on normality assumption for error terms, derivations related to the mixed estimation approach, so-called Goldberger-Theil estimator, where we do not have to impose normality assumption for error terms because it is based on the Generalized Least Square estimation, and the third one is interpreted as minimum square linear estimator (Tanizaki,1996). For the last decade, some researchers have studied the procedure of maximum likelihood, which is used to maximize an innovation form of the likelihood function, thus they have worked on many algorithms to optimize the likelihood function and they have proposed many filters and smoothers and given their results in many papers of work and many publications. In spite of the fact that this field is so extended, we are going to emphasize our study in the econometric field. We will propose three algorithms: two quasi-Newton; Berndt, Hall and Hausman (BHHH, 1974) and Broyden-Fletcher-Goldfarb-Shanno (BFGS, 1960), and the other optimizer is Expectation-Maximization (EM) algorithm(Shumway and Stoffer, 1982). The general objective is to find the algorithm (and its computations implementation) more efficient in question of precision and time of computation to maximize the maximum likelihood for the estimation of the parameters of the AR(1)models.

2 The state space form

In a lineal gaussian state space, the state space model is formulated as follows: For all t=1,.....,n

$$x_t = \Phi x_{t-1} + w_t,\ w_t \sim N(0, Q) \tag{1}$$

x_t is a $(p \times 1)$ vector, Φ is a $(p \times p)$ matrix and w_t is a $(p \times 1)$ vector.

(2.1) is called the Transition equation or State equation. and

$$y_t = Ax_t + v_t,\ v_t \sim N(0, R) \tag{2}$$

y_t is a $(q \times 1)$ vector, A_t is a $(q \times p)$ matrix and v_t is a $(q \times 1)$ vector. This equation is called Measurement equation or Observation equation. where
y: represents the observed serie
x: is the series which represents the non observed state
Φ: Parameter of the autoregressive process
R$(q \times q)$: covariance matrix of the measurement equation

*My acknowledgment to P.Munyoz for her support and help to achieve this work

(v_t)
Q($p\times p$): covariance matrix of the transition equation(w_t)
n: length of the series

2.1 The Kalman Filter

Filter purpose: our problem is to estimate the state x_t using information of the y_t up to time s:

$$Y_s = \{y_1, ...y_s\} \tag{3}$$

we calculate the mean and the covariance of x_{t-1} given the information at time $t-1$ by Y_{t-1}:

$$E[x_{t-1} \mid Y_{t-1}] \equiv x_{t-1}^{t-1} \tag{4}$$

$$Cov(x_{t-1} \mid Y_{t-1}) \equiv P_{t-1}^{t-1} \tag{5}$$

Prediction :
Thus using the initial conditions $x_0^0 = \mu_0$ and $P_0^0 = \Sigma_0$ and assuming known the estimations given at time $t-1$, we can obtain the predicted estimation x_t^{t-1} and P_t^{t-1} at time t, for $t = 1, ..., n$using the State equation.

$$x_t^{t-1} = E[x_t \mid Y_{t-1}] = E(\Phi x_{t-1} + w_t \mid Y_{t-1}) = \Phi x_{t-1}^{t-1} \tag{6}$$

and being w_t independent of observations and states, with

$$E[w_t] = 0 \tag{7}$$

$$P_t^{t-1} = \Phi P_{t-1}^{t-1} \Phi' + Q$$

Then we estimate the mean and the variance of prediction data of y_t that will be observed at time t using the observation equation, which are respectively:

$$y_t^{t-1} = E[y_t \mid Y_{t-1}] = A_t x_t^{t-1} \tag{8}$$

When the new data arrives y_t the innovation (or prediction error or one-step ahead forecast error) called ϵ_t is obtained as below:

$$\epsilon_t = A_t(x_t - x_t^{t-1}) + v_t$$

The innovations $\epsilon_1, ..., \epsilon_n$ are independent Gaussian random vectors with zero means and covariance matrices:

$$\begin{aligned} F_t^{t-1} &= Var(\epsilon_t \mid Y_{t-1}) = var(y_t - y_t^{t-1} \mid Y_{t-1}) \\ &= A_t P_t^{t-1} A_t' + R \end{aligned}$$

The mean $E(\epsilon_t \mid Y_{t-1})$ and the covariance $Cov(\epsilon_t, y_j)$ for $j < t$ of the innovations are zero as:

$$E(\epsilon_t \mid Y_{t-1}) = 0$$

and

$$\begin{aligned} Cov(\epsilon_t, y_j) &= E[y_j E[\epsilon_t \mid_{t-1}]] \\ &= 0 \qquad for \quad j < t \end{aligned} \tag{9}$$

Thus, the 1-step ahead forecast error is uncorrelated (hence, independent in the Gaussian case) with y_t for $j < t$.

3 Maximum Likelihood Estimation

For the linear Gaussian model,using the Kalman filter equation and ignoring a constant the log-likelihood can be written as:

$$-2 \ln L_Y(\Theta) = \sum_{t=1}^{n} \log |F_t^{t-1}(\Theta)| + \sum_{t=1}^{n} \epsilon_t(\Theta)' F_t^{t-1}(\Theta)^{-1} \epsilon_t(\Theta) \tag{10}$$

The loglikelihood can be maximized by means of iterative numerical procedures. Using a numerical method, like BHHH algorithm, we can obtain the minimum of this function. Shumway and Stoffer (1982) presented a conceptually simpler estimation procedure based on EM (expectation-maximization)algorithm, and Broyden-Fletcher-Goldfarb-Shanno used the BFGS algorithm.

3.1 Numerical maximization algorithms

A wide range of numerical search algorithms are available for maximizing the loglikelihood. Many of these are based on Newton's method which solves the equation:

$$\partial_1(\Theta) = \frac{\partial \log L_Y(\Theta)}{\partial \Theta} = 0 \tag{11}$$

using a first-order Taylor series, we obtain:

$$\partial_1(\Theta) \simeq \widetilde{\partial}_1(\Theta) + \widetilde{\partial}_2(\Theta)(\Theta - \widetilde{\Theta}) \tag{12}$$

for a random value of $\widetilde{\Theta}$, where:

$$\widetilde{\partial}_1(\Theta) = \partial_1(\Theta)|_{\Theta=\widetilde{\Theta}} \quad ; \tag{13}$$

$$\widetilde{\partial}_2(\Theta) = \partial_2(\Theta)|_{\Theta=\widetilde{\Theta}} \tag{14}$$

and

$$\partial_2(\Theta) = \frac{\partial^2 \log L_Y(\Theta)}{\partial \Theta \partial \Theta'} \tag{15}$$

$$\overline{\Theta} = \widetilde{\Theta} - \widetilde{\partial}_2(\Theta)^{-1} \widetilde{\partial}_1(\Theta) \tag{16}$$

This process is repeated until it converges. If the Hessian matrix $\partial_2(\Theta)$ is negative definite for all Θ, the loglikelihood is said to be concave and a unique maximum exists for the likelihood. The *gradient* $\partial_1(\Theta)$ determines the direction of the step taken to the optimum and the Hessian modifies the size of the step. In practice it is often computationally demanding or impossible to compute $\partial_1(\Theta)$ and $\partial_2(\Theta)$ analytically. Numerical evaluation of $\partial_1(\Theta)$ is usually feasible. A variety of computational devices are available to approximate $\partial_2(\Theta)$ in order to avoid computing it analytically.

3.1.1 BFGS algorithm

The BFGS method ensures that the approximate Hessian matrix remains negative definite. The details and derivations of the Newton's method of optimization and the BFGS method in particular can be found (Fletcher, 1987). The BFGS method is one of the most famous quasi-Newton algorithms for unconstrained optimization. It is possible to overstep the maximum in the direction determined by the vector $-\widetilde{\partial}_2(\Theta)^{-1}\widetilde{\partial}_1(\Theta)$ and therefore a line search along the gradient vector within the optimization process. We obtain the algorithm:

$$\overline{\Theta} = \widetilde{\Theta} + s\widetilde{\partial}_2(\Theta)^{-1}\widetilde{\partial}_1(\Theta) \tag{17}$$

where various methods are available to find the optimum value for s which is usually found to be between 0 and 1 The approximation of the Hessian matrix uses a device in which at each new value for Θ a new approximate inverse Hessian matrix is obtained via the recursion:

$$\widetilde{\partial}_2(\theta)^{-1} = \widetilde{\partial}_2(\theta)^{-1} + (s + \frac{g'g^*}{(\widetilde{\partial}_2(\Theta)^{-1}\widetilde{\partial}_1(\Theta))'g}) \frac{\widetilde{\partial}_2(\Theta)^{-1}\widetilde{\partial}_1(\Theta)(\widetilde{\partial}_2(\Theta)^{-1}\widetilde{\partial}_1(\Theta))'}{(\widetilde{\partial}_2(\Theta)^{-1}\widetilde{\partial}_1(\Theta))'g} - \frac{\widetilde{\partial}_2(\Theta)^{-1}\widetilde{\partial}_1(\Theta)g^{*'} + g^*(\widetilde{\partial}_2(\Theta)^{-1}\widetilde{\partial}_1(\Theta))'}{(\widetilde{\partial}_2(\Theta)^{-1}\widetilde{\partial}_1(\Theta))'g}$$

where g is defined as the difference between the gradient $\widetilde{\partial}_1(\Theta)$ and the gradient for a trial value of Θ prior to $\widetilde{\Theta}$, and g^* is defined as:

$$g^* = \widetilde{\partial}_2(\Theta)^{-1}g \tag{18}$$

3.1.2 The EM algorithm

The EM (expectation-maximization)algorithm is a well-known tool for iterative maximum likelihood estimation. The earlier EM methods for the state space model were developed by Shumway and Stoffer (1982) and Watson and Engle (1983). In addition to Newton-Raphson, Shumway and Stoffer (1982)presented conceptually simpler estimation procedure based on the EM algorithm (Dempster et al.1977). The basic idea is that if we can observe the states, $X_n = x_0, x_1, ..., x_n$ and $Y_n = y_0, y_1, ..., y_n$, then we would consider X_n, Y_n as the complete data, with the joint density:

$$f_\Theta(X_n, Y_n) = f_{\mu_0,\Sigma_0}(x_0)\prod_{t=1}^{n} f_{\Phi,Q}(x_t/x_{t-1})\prod_{t=1}^{n} f_R(y_t/x_t) \tag{19}$$

Under the Gaussian assumption and ignoring constants, the complete data likelihood can be written as :

$$-2lnL_{X,Y}(\Theta) = ln|\Sigma_0| + (x_0-\mu_0)'\Sigma_0^{-1}(x_0-\mu_0) + ln|Q| + \sum_{t=1}^{n}(x_t-\Phi x_{t-1})'Q^{-1}(x_t-\Phi x_{t-1}) \tag{20}$$
$$+ln|R| + \sum_{t=1}^{n}(y_t - A_t x_t)'R^{-1}(y_t - A_t x_t)$$

Thus, if we did have the complete data, we could then use the results from multivariate normal theory to easily obtain the MLEs of Θ. If we do not have the complete data, the EM algorithm gives us an iterative method for finding the MLEs of Θ based on the incomplete data, Y_n, by successively maximizing the conditional expectation of the complete data likelihood. To implement the EM algorithm, we write, at iteration j, (j =1,2,....),

$$Q(\Theta/\Theta^{(j-1)}) = E\{-2lnL_{X,Y}(\Theta)/Y_n, \Theta^{(j-1)}\} \tag{21}$$

Expectation − step : When we calculate the equation above we have the expectation step. Of course, given the current value of the parameters, $\Theta^{(j-1)}$, we can obtain the desired conditional expectations as smoothers.

$$Q(\Theta/\Theta^{(j-1)}) = ln|\Sigma_0| + tr\Sigma_0^{-1}[P_0^n + (x_0^n-\mu_0)(x_0^n-\mu_0)'] + ln|Q| + trQ^{-1}[S_{11} - S_{10}\Phi' - \Phi S_{10}' + \Phi S_{00}\Phi'] \tag{22}$$
$$+ln|R| + tr\{R^{-1}\sum_{t=1}^{n}[(y_t - A_t x_t^n)(y_t - A_t x_t^n)' + A_t P_t^n A_t']\}$$

where

$$S_{11} = \sum_{t=1}^{n}(x_t^n x_t^{n'} + P_t^n) \tag{23}$$

$$S_{10} = \sum_{t=1}^{n}(x_t^n x_{t-1}^{n'} + P_{t,t-1}^n) \tag{24}$$

$$S_{00} = \sum_{t=1}^{n}(x_{t-1}^n x_{t-1}^{n'} + P_{t-1}^n) \tag{25}$$

Maximization − step : Maximizing this equation with respect to the parameters, at iteration j, constitutes the maximization step, and is analogous to the usual multivariate regression approach, which yields the updated estimates

$$\Phi^{(j)} = S_{10}S_{00}^{-1} \tag{26}$$

$$Q^{(j)} = n^{-1}(S_{11} - S_{10}S_{00}^{-1}S_{10}') \tag{27}$$

$$R^{(j)} = n^{-1}\sum_{t=1}^{n}[(y_t - A_t x_t^n)(y_t - A_t x_t^n)' + A_t P_t^n A_t'] \tag{28}$$

The initial mean and covariance cannot be estimated simultaneously, so it is conventional to fix both or to fix the covariance matrix and use the estimator:

$$\mu_0^{(j)} = x_0^n \tag{29}$$

obtained from minimizing the equation above under that assumption.

3.1.3 The BHHH algorithm

Going from the density function $f_\Theta(y_n, ..., y_1)$, as we defined, the gradient in (2.36) and the Hessian Matrix in(2.40), the notation $\partial_1(\theta_0)$ and $\partial_2(\theta_0)$ indicates the derivative evaluated at θ_0, where θ_0 is known as Information Matrix.

$$Var[\partial_1(\theta_0)] = -E[\partial_1(\theta_0)] \tag{30}$$

The asymptotic covariance matrix of the maximum likelihood estimator is a matrix of parameters that must be estimated (that is, it is a function of the θ_0 that is being estimated). If the form of the expected values of the second derivatives of the log-likelihood is known, then:

$$[I(\theta_0)]^{-1} = \{-E[\frac{\partial^2 lnL_Y(\theta_0)}{\partial\theta_0\partial\theta_0'}]\}^{-1} \tag{31}$$

can be evaluated at $\hat{\theta}$ to estimate the covariance matrix for the MLE. This estimator will rarely be available. The second derivatives of the log-likelihood will almost always be complicated nonlinear functions of the data whose exact expected values will be unknown. However, the expected second derivatives matrix is the covariance matrix of the first derivatives vector is:

$$[\hat{\hat{I}}(\hat{\theta_0})]^{-1} = [\sum_{i=1}^{n} \partial_{1i}(\hat{\Theta})\partial_{1i}(\hat{\Theta})']^{-1} = [\hat{G}'\hat{G}]^{-1} \tag{32}$$

Where

$$\partial_{1i}(\hat{\Theta}) = \frac{\partial ln f_{\hat{\theta}}(y_i)}{\partial\hat{\theta}} \tag{33}$$

$$\hat{G} = [\partial_{11}, \partial_{12}, ..., \partial_{1n}]' \tag{34}$$

$\hat{G}$ is an $n \times k$ matrix with ith row equal to the transpose of the ith vector of derivatives in the terms of the log-likelihood function with respect to the k elements of Θ. For a single parameter, this estimator is just the reciprocal of the sum of squares of the first derivatives. $[\hat{\hat{I}}(\hat{\theta_0})]^{-1}$ is called the BHHH estimator.

4 Results

In this part, we are going to display the different results of the estimation of the parameters for an AR (1) model. As we have said that the estimation of the parameters is not easy, nevertheless, in our work, we are going to apply the three algorithms cited above in the AR(1) process to see which one is the best in estimating the unknowing parameters. We have chosen the case of $\phi = 0.5$, the different results are shown in the table 1 with the corresponding Quadratic Medium Error (QEM) for each parameter ϕ, Q and R. We Started with a simulation process of the data with a length of $n = 2500$, and then applied the Kalman Filter. We noted from programming and the estimation process, that using the BHHH algorithm, the time of computation was extremely slow, and the results are not so precise regarding to the error and the estimation value of the parameters. This issue can be related to the fact that the Package used in R programm is still not so exact in the process of estimation, for this reason, one of the step I am going to achieve is to try to programm with the S_+ package to give more precise conclusions regarding the optimization with the three different algorithms.

Table.1. Quadratic medium error for AR(1) model

	EM	L-BFGS-B	BHHH
$\phi = 0.5$	0.190	0.353	0.497
$Q = 1$	0.197	0.771	0.923
$R = 1$	0.179	0.658	0.978
$\phi = 0.5$	0.187	0.402	0.501
$Q = 10$	0.186	0.883	0.657
$R = 1$	0.168	0.635	0.886
$\phi = 0.5$	0.181	0.421	0.833
$Q = 1$	0.097	0.564	0.433
$R = 10$	0.135	0.398	0.432

5 Conclusions and future work

Comparing the two algorithms L-BFGS-B and EM, we have concluded that the EM algorithm is the most efficient in time of computation and the most precise with the lowest rate of error obtained in the estimation process.In the case of AR(1) model, from the results of the estimation of the parameters of the models and the corresponding errors, we can confirm that the L-BFGS-B algorithm gives a good optimization of the value estimated of the parameters.

References

[1] Barndorff-Nielsen, O.E., Nicolato, E. and Shephard, N.G.: Some Recent Developments in Stochastic Volatility Modelling. Quantitative Finance 2(2002) 11-23.

[2] Broto, C. and Ruiz, E.,: Unobserved Component Models with Asymmmetric Conditional Variances, Computational Statistics and Data Analysis, 50, 9 (2005a)2146-2166.

[3] Broto, C. and Ruiz, E.,: Using Auxiliary Residuals to detect Conditional heteroscedasticity in inflation,(2005b). http://docubib.uc3m.es/WORKINGPAPERS/WS/ws060402.pdf

[4] Broto, C. and Ruiz, E.: Estimation Methods for Stochastic Volatility Models: a Survey. Journal of Econometrics 18 N5(2004) 614-649.

[5] Carnero, A., Pea, D.,and Ruiz, E.: Persistence and Curtosis in Autoregressive Stochastic Volatility Models (manuscript, Universidad Carlos III de Madrid) (2003).

[6] Danielsson, J.: Stochastic Volatility in Asset Prices: Estimation with Simulated Maximum Likelihood. Journal of Econometrics 64(1994) 375-400.

[7] Davis, R.A. and Yam, G.B.: Estimation for State-Space Models Based on a Likelihood Approximation. Statistica Sinica 15 (2005) 381-406.

[8] Greene, P.: Maximum Likelihood Estimation-Chap.17(2002).

[9] Jacquier, E., Polsen, N.G. and Rossi, P. E.: Bayesian Analysis of Stochastic Volatility Models with Fat-tails and Correlated Errors. Journal of Econometrics 122 (2004) 185-212.

[10] Kalman, R.E.: A new approach to linear filtering and prediction problems. Journal of basic engineering, transactions ASME, Ser.D 82(1960) 35-45.

[11] Shephard, N.: Statistical Aspects of ARCH and Stochastic Volatility. In O.E. Burndorff-Nielsen, D. R. Cox and D.V. Hinkley (eds). Statistical Models in Econometrics(1996).

[12] Shumway,R.H. and Stoffer, D.S.: An appraoch to time series smoothing and forecasting using the EM algorithm. Journal of Times series analysis 3 (1982) 253-264.

[13] Shumway,R.H. and Stoffer, D.S.: Time Series and Its Applications (2002) 301-321.

[14] Tanizaki, H.: Nonlinear Filters: Estimation and Applications (Springer-Verlag)(1996).

[15] Tsay, R.S.: Analysis of Financial Time Series.2nd Ed. John Wiley and Sons Inc. (2005).