FORECASTING ACCURACY AND THE IMPORTANCE OF FINDING THE GLOBAL MAXIMA OF THE LIKELIHOOD FUNCTION

Mahbuba Yeasmin[1] & Maxwell L. King

Department of Econometrics & Business Statistics, Monash University, Clayton, Victoria 3800, Australia

Abstract

The method of maximum likelihood estimation is widely used in econometrics because it can be applied to a wide range of different parametric models and the resultant estimators have good asymptotic properties. Modern statistical computer packages allow one to carry out maximum likelihood estimation on reasonably complicated models with some degree of ease. An unfortunate drawback of the method, particularly when numerical methods are used to maximise the likelihood function, is that we can sometimes end up with a local maxima rather than the global maxima. This is a well-known problem in econometrics, although a survey of recent textbooks suggests that the consequences of accepting a local maxima instead of the global maxima are not well articulated. While the consequences for the estimation of parameters of interest might seem obvious, less obvious is what effect using parameter estimates from a local maximum could have on the small sample forecasting performance of a model.

This paper considers this problem in the context of the linear regression model with first-order moving average MA(1) errors; a model that can have local and global maxima with one at $\gamma = 1$ or $\gamma = -1$ where γ is the MA(1) parameter. We compare the accuracy of forecasts using five different estimation strategies. The first involves accepting the maximum that comes from maximising the likelihood from one fixed starting point. The second and third involve taking the best result from three fixed starting points either always or when the estimated value of γ is 1 or –1, and the fourth and fifth involve the same approach but using a total of 21 different starting values. We find that for this particular case the extra care taken to find the global maximum can, in some circumstances, have dramatic effects on forecasting accuracy.

Key words: General linear model; Local maxima; MA errors; Marginal likelihood; Concentrated likelihood; Forecasting performance

JEL classification: C5; C53

[1] Corresponding author: Tel +61-3-99055180; Fax: +61-3-9905-5474; Mahbuba.yeasmin@buseco.monash.edu.au

1. Introduction

Forecasting plays an important role in the field of econometrics, statistics and many other branches of science. One standard approach to the problem of producing forecasts is to construct a statistical or econometric model. This typically involves the estimation of unknown parameters and there is a common view that forecasts can only be as good as the quality of the parameter estimates. Researchers are often concerned about estimating unknown parameters of a model with forecasting in mind. The method of maximum likelihood has become very popular because it can be applied to a wide range of different parametric models and the resultant estimators have good asymptotic properties (for example, see Cramer (1986)).

One of the major problems of maximising the likelihood function is that the numerical optimization procedure sometimes leads to local maxima instead of the global maxima. For example, when there are constraints on a parameter or parameters being estimated, there is often a local maxima on the boundary of the constrained parameter set and the numerical procedure can sometimes end up at this local maxima. Although this is a well-known problem in econometrics, the consequences for subsequent hypothesis testing and forecasting are not well known. Yeasmin and King (2000) showed that accepting local maxima can have an important effect on the small sample properties of a hypothesis test. They demonstrated that finding the global maxima can dramatically improve the actual size of the test. They found circumstances in which without taking care over finding the global maxima, one might unknowingly be applying a test at the 0.20 level rather than at approximately the 0.05 level.

The aim of this paper is to consider the effect of the above-mentioned problem on forecasting performance. Can taking extra care to finding the global maxima when estimating parameters improve forecasting accuracy? This is investigated in the context of the linear regression model when the errors follow a MA(1) process with MA(1) parameter γ. We compare the accuracy of forecasts using five different estimation strategies. The first involves accepting the maximum that comes from maximizing the likelihood from one fixed starting point. The second and third involve taking the best result from an additional three fixed starting points either always or when initially the estimated value of γ is at a boundary point. The fourth and fifth involve the same approach but taking even greater care to find the global maxima by using 21 different starting values. We find that for this particular case, the extra care taken to find the global maximum can, in some circumstances, have dramatic effects on forecasting accuracy. This paper also compares the estimation of the moving average parameter from the concentrated likelihood and the marginal likelihood. There is a growing literature that suggests that the use of the marginal likelihood in place of the concentrated likelihood can help reduce estimation bias (see for example, Cooper and Thompson (1977), Rahman and King (1998) and Lasker and King (1998)). We therefore expect that the parameters estimated via the marginal likelihood to provide more accurate forecasts compared to forecasts from estimates based on the concentrated likelihood.

This paper is organised as follows. In section 2, we discuss the model and the consequences of accepting local maxima for the estimating parameter based on marginal likelihood and concentrated likelihood in the linear model. In section 3, we outline the design of a Monte Carlo experiment to estimate the average mean square forecasting error for different strategies for different sample sizes, different values of the parameter and different design matrices. Section 4, contains a brief discussion of the results of the Monte Carlo study. Section 5, presents some concluding remarks.

2. The Model

Consider the linear regression model with non-spherical disturbances

$$y = X\beta + u \tag{1}$$

where y is an $n \times 1$ vector, X is an $n \times k$ matrix of known values and of full column rank, β is a k-dimensional vector of unknown parameters. The elements of u are assumed to follow the MA(1) process.

$$u_t = \varepsilon_t + \gamma\varepsilon_{t-1},\ -1 \le \gamma \le 1, \text{ where } \varepsilon_t \sim IIN(0, \sigma^2) \tag{2}$$

which implies that $u \sim N(0, \sigma^2\Sigma(\gamma))$ where $\Sigma(\gamma)$ is an tridiagonal symmetric matrix of the form

$$\Sigma(\gamma) = \begin{pmatrix} 1+\gamma^2 & \gamma & 0 & \cdots & 0 \\ \gamma & 1+\gamma^2 & \gamma & \cdots & 0 \\ 0 & \gamma & 1+\gamma^2 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & \cdots & & 1+\gamma^2 \end{pmatrix}.$$

One way to tackle the estimation problem for this model is to transform using the Cholesky decomposition. Let $\Sigma(\gamma) = LL'$ where L is the upper triangular matrix of $\Sigma(\gamma)$, defined as

$$L = \begin{pmatrix} l_{11} & 0 & 0 & \cdots & 0 \\ l_{12} & l_{22} & 0 & \cdots & 0 \\ 0 & l_{23} & l_{33} & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & & & l_{nn} \end{pmatrix}.$$

The non zero elements of L can be recursively obtained by using

$l_{11} = (1+\gamma^2)^{1/2}$; $l_{i,i-1} = \dfrac{\gamma}{l_{i-1,i-1}}$ and $l_{ii} = (1+\gamma^2 - l_{i,i-1}^2)^{\frac{1}{2}}$ where $i = 2,\ 3,\ \cdots\ ,\ n.$

The inverse of the matrix L is used to transform model (1) to find the generalised least squares (GLS) estimator of β. This implies $Ly^* = y$ and $LX^* = X$ where y^* and X^* are the transformed variables. Using these expressions it is easy to obtained the formula for y^* which can be written as

$$y_1^* = y_1 / (1+\gamma^2)^{1/2}$$

$$y_2^* = (y_2 - l_{21} y_1^*) / l_{22}$$

$$y_3^* = (y_3 - l_{32} y_2^*) / l_{33}$$

and so on. This same transformation is also applied to each column of X in turn in order to obtain X^*. The first element of the intercept will be $1/l_{11}$ and the other values of the intercept will be $1/(l_{i,i-1} + l_{ii})$. The resultant transformed model can be written as

$$y^* = X^*\beta + u^*$$

which implies $\mathrm{E}(u^*) = 0$ and $\mathrm{E}(u^* u^{*\prime}) = \sigma^2 I$.

The GLS estimator can be found by applying the usual ordinary least squares estimator to the above transformed model. Unless the value of γ is known, the above technique cannot be applied. For a known value of γ, the GLS estimator of $\hat{\beta}$ is the best linear unbiased estimator of β. Unfortunately, the value of the parameters γ is typically unknown and therefore it needs to be estimated. For example, we can obtain the estimated values of γ by maximising the concentrated likelihood or the marginal likelihood of γ. The log concentrated likelihood is

$$l_c(y;\gamma) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\{s^2(\gamma)/n\} - \frac{1}{2}\log|\Sigma(\gamma)| - \frac{n}{2} \tag{3}$$

where σ^2 and β are replaced by their estimated values

$$s^2(\gamma) = \{y - X\hat{\beta}(\gamma)\}' \Sigma^{-1}(\gamma)\{y - X\hat{\beta}(\gamma)\}$$

and

$$\hat{\beta}(\gamma) = \{X'\Sigma^{-1}(\gamma)X\}^{-1} X'\Sigma^{-1}(\gamma)y \tag{4}$$

and the log marginal likelihood according to Tunnicliffe Wilson (1989) is defined by

$$l_m(y;\gamma) = -\frac{1}{2}\log|\Sigma(\gamma)| - \frac{1}{2}\log|X'\Sigma^{-1}(\gamma)X| - \frac{m}{2}\log|s^2(\gamma)| \tag{5}$$

where $m = n - k$.

To obtain a final estimate of β, we need to estimate the value of γ by maximizing either equation (3) or (5) and replace γ by the estimated value in equation (4) which can be written as

$$\tilde{\beta}^{(j)} = \{X'\Sigma^{-1}(\hat{\gamma}^{(j)})X\}^{-1} X'\Sigma^{-1}(\hat{\gamma}^{(j)})y \tag{8}$$

where $j = 1, 2$ in which $j = 1$ indicates the estimated value of γ comes from method of maximum likelihood (maximising (3)) and $j = 2$ indicates the estimated value comes from the marginal likelihood (maximising (5)).

Our interest is in measuring the forecasting accuracy of this model using our estimates and different strategies for finding the required maximum of the likelihood. Suppose we wish to find the one-step-ahead forecasts of y in model (1) which can be written as

$$y_{n+1} = x'_{n+1}\beta + u_{n+1} \tag{6}$$

where y_{n+1} is the next value of y, x_{n+1} is the $k \times 1$ vector of observations on the regressors at time $n+1$ and u_{n+1} is its associated disturbance term. The predicted value of y_{n+1} can be written as

$$\hat{y}_{n+1}^{(j)} = x'_{n+1}\tilde{\beta}^{(j)} + \hat{\gamma}^{(j)}\hat{\varepsilon}_n^{(j)} \tag{7}$$

Following King and McAleer (1987), the prediction error $\hat{\varepsilon}_n^{(j)}$ is obtained via the recursive procedure starting with

$$\hat{\varepsilon}_0^{(j)} = 0$$

and then calculating

$$\hat{y}_t^{(j)} = x'_t\tilde{\beta}^{(j)} + \hat{\gamma}^{(j)}\hat{\varepsilon}_{t-1}^{(j)}$$

and

$$\hat{\varepsilon}_t^{(j)} = y_t - \hat{y}_t^{(j)} \tag{9}$$

for $t = 1, 2, \cdots, n$ and $j = 1, 2$.

The form of log likelihoods (3) and (5) reveals that we are unable to maximize the likelihoods analytically in order to estimate γ. In this situation, the best way to solve the estimation problem is by optimising the log likelihood using a suitable numerical optimization program. The model given by (1) and (2) has a minor identification problem in that $\sigma^2\Sigma(\gamma)$ and $\sigma_*^2\Sigma(\gamma_*)$ take exactly the same value whenever $\sigma_*^2 = \sigma^2\gamma^2$ and $\gamma_* = \frac{1}{\gamma}$. As a consequence of this, we can easily show that

$$\begin{aligned}\hat{\beta}(1/\gamma) &= \left[X'\Sigma^{-1}(1/\gamma)X\right]^{-1}X'\Sigma^{-1}(1/\gamma)y\\ &= \left[X'\gamma^2\Sigma^{-1}(\gamma)X\right]^{-1}X'\gamma^2\Sigma^{-1}(\gamma)y\\ &= \left[X'\Sigma^{-1}(\gamma)X\right]^{-1}X'\Sigma^{-1}(\gamma)y\\ &= \hat{\beta}(\gamma)\end{aligned}$$

$$
\begin{aligned}
s^2(1/\gamma) &= [y - X\hat{\beta}(1/\gamma)]' \Sigma^{-1}(1/\gamma)[y - X\hat{\beta}(1/\gamma)] \\
&= [y - X\hat{\beta}(\gamma)]' \gamma^2 \Sigma^{-1}(\gamma)[y - X\hat{\beta}(\gamma)] \\
&= \gamma^2 [y - X\hat{\beta}(\gamma)]' \Sigma^{-1}(\gamma)[y - X\hat{\beta}(\gamma)] \\
&= \gamma^2 s^2(\gamma)
\end{aligned}
$$

$$
\begin{aligned}
l_c(y;1/\gamma) &= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\{s^2(1/\gamma)/n\} - \frac{1}{2}\log|\Sigma(1/\gamma)| - \frac{n}{2} \\
&= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\{\gamma^2 s^2(\gamma)/n\} - \frac{1}{2}\log|1/\gamma^2\Sigma(\gamma)| - \frac{n}{2} \\
&= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\gamma^2 - \frac{n}{2}\log\{s^2(\gamma)/n\} - \frac{1}{2}\log\{(1/\gamma^2)^n\} \\
&\quad - \frac{1}{2}\log|\Sigma(\gamma)| - \frac{n}{2} \\
&= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\{s^2(\gamma)/n\} - \frac{1}{2}\log|\Sigma(\gamma)| - \frac{n}{2} \\
&= l_c(y;\gamma)
\end{aligned}
$$

and

$$
\begin{aligned}
l_m(y;1/\gamma) &= -\frac{1}{2}\log|\Sigma(1/\gamma)| - \frac{1}{2}\log|X'\Sigma^{-1}(1/\gamma)X| - \frac{m}{2}\log\{s^2(1/\gamma)/m\} \\
&= -\frac{1}{2}\log|1/\gamma^2\Sigma(\gamma)| - \frac{1}{2}\log|\gamma^2 X'\Sigma^{-1}(\gamma)X| - \frac{m}{2}\log\{\gamma^2 s^2(\gamma)/m\} \\
&= -\frac{1}{2}\log|\Sigma(\gamma)| - \frac{n}{2}\log\{1/\gamma^2\} - \frac{1}{2}\log(\gamma^2)^k - \frac{m}{2}\log|X'\Sigma^{-1}(\gamma)X| \\
&\quad - \frac{m}{2}\log\{\gamma^2\} - \frac{m}{2}\log\{s^2(\gamma)/m\} \\
&= -\frac{1}{2}\log|\Sigma(\gamma)| - \frac{m}{2}\log|X'\Sigma^{-1}(\gamma)X| - \frac{m}{2}\log\{s^2(\gamma)/m\} \\
&= l_m(y;\gamma)
\end{aligned}
$$

As a consequence of this identification problem, both likelihoods have turning points (which may be a maximum or a minimum) at $|\gamma| = 1$. This can result in the global maximum being at $|\gamma| = 1$, a point that is well recognised in the literature (see Kang (1975), Dunsmuir (1981) Cryer and Ledolter (1981)). It can also result in a local maxima being at $|\gamma| = 1$. Consequently, when as can often happen, maximising the likelihood results in $\left|\hat{\gamma}\right| = 1$, one might be suspicious that one has a local maxima rather than a global maxima. This could be investigated by trying a number of different starting values for the optimization routine and seeing if another estimate of γ results in a large maximised likelihood.

3. The Design of Experiment

A Monte Carlo study was conducted to estimate and compare the forecasting accuracy of the linear model when the error term follows the MA(1) process. The comparison of forecasts is based on maximum likelihood and maximum marginal likelihood. Five strategies were used for obtaining the estimated value of the parameter. These strategies are (i) using one starting value and accepting the outcome as a global maxima. As there is always doubt that, at the first attempt, we have found the global maxima, the remaining four strategies involve searching for the real global maxima by taking some further steps.

Firstly we consider two further strategies which are (ii) using a further three optimizations with respective starting values of $\gamma = -0.5, 0, 0.5$ and (iii) using further twenty-one optimizations with respective starting values of $\gamma = \pm.95, \pm.90, \pm.80 \cdots \pm.10, 0$. In both cases, we take as the estimated value, the results from the optimization which has the largest maximized log likelihood function. As already mentioned, an estimate of γ on the boundary may indicate a local maxima. Therefore we considered further two strategies with this in mind.

These are:

(iv) If the resultant estimate from strategy (i) is on the boundary then we use $\hat{\gamma}$ as the final estimate of the parameter that comes from strategy (ii) otherwise we use the estimated value of γ from strategy (i).

(v) If the resultant estimate from strategy (i) is on the boundary then we consider $\hat{\gamma}$ as the final estimate of the parameter that comes from strategy (iii) otherwise take use the estimated value of γ from strategy (i).

In all five cases, the initial starting value was set to

$$\hat{\gamma}_0 = \begin{cases} -0.999, & \text{if } \hat{\rho} < -0.499, \\ 2\hat{\rho}, & \text{if } -0.499 \le \hat{\rho} \le 0.499, \\ 0.999, & \text{if } \hat{\rho} > 0.499, \end{cases}$$

where $\hat{\rho} = \sum_{t=2}^{n} z_t z_{t-1} / \sum_{t=1}^{n} z_t^{\,2}$ in which z_t, for $t = 1, \cdots, n$ are the OLS residuals from (1). In this study two step estimators were used, that is in first step we obtain the OLS estimate of γ, and in second step this estimate is used as an initial value to obtain maximum likelihood and maximum marginal likelihood estimates by minimising the equation (2) and (3). This maximisation was carried out by using the Gauss Constrained optimization algorithm (Aptech systems, 1996)

For this study the data matrices were generated by using the general form $x_{it} = \rho x_{it-1} + v_t$ where $v_t \sim IN(0,1)$, $i = 1, 2, \cdots, k$. This expression allows us to work with different types of design matrices for different values of ρ such as

X_1 A constant or intercept plus two white noise regressors generated as $x_{it} = v_t$, $i = 1, 2.$

X_2 A constant or intercept plus two autoregressive regressors generated as

$$x_{it} = 0.8x_{it-1} + v_t, \quad i = 1, 2. \qquad (10)$$

X_3 A constant or intercept plus five autoregressive regressors generate from (10).

X_4 A constant or intercept plus two random walk regressors generated as $x_{it} = x_{it-1} + v_t \quad i = 1, 2.$

X_5 A constant or intercept plus two explosive regressors generated as $x_{it} = 1.02 x_{it-1} + v_t \quad i = 1, 2.$

Throughout the experiment, we used two regressors for each of the design matrices except for X_3 for which we used five regressors, in order to check whether the number of regressors has an effect on forecasting performance. The sample sizes we used were $n = 30,\ 60$ and 120; the values of γ were $\gamma =$ ±0.90, ±0.75, ±0.60, ±0.45, ±0.30, ±0.15, 0, and 1000 iterations were used in the experiment.

Our aim is to measure the forecasting accuracy and compare the results, which were obtained using the several different methods for the above model. In order to measure the accuracy of forecasts we used MSFE (mean square forecast error) and MAE (mean absolute error), these being average over the 1000 iterations to produce the average MSFE (AMSFE) and average MAE (AMAE) respectively.

4. Results of the Simulation

The average mean square forecasts of the five different strategies are reported in Tables 1 to 6 for different sample sizes n, different design matrices and different values of γ. A feature of Tables 1 to 6 is that the AMSFE calculated using marginal likelihood provides better results for both large and small samples as compared to maximum likelihood. This is also true for different design matrices and the different values of γ.

From the results in Table 1 to 6, we see that AMSFE decreases as the sample size increases for each design matrix and the value of parameter γ. For example, when the sample sizes are n=30, 60 and 120, then the estimated AMSFE are 11.525, 6.387 and 2.844, respectively for marginal likelihood values of parameter γ and the design matrix X_1.

The improvements in AMSFE also suggest the importance of finding the global maxima when estimating γ. The best example of searching for the global maxima using extra effort can improve the AMSFE dramatically in the sense that it gives the minimum AMSFE, which can be seen in Table 3 for $\gamma = 0.60$.

Initially without taking any effort, we accept the outcome from strategy (i) as our global maxima. As a consequence it provides very unacceptably high AMSFE.
All tables show that when extra effort is taken to calculate the estimated value can reduce the estimation bias as well as provide forecasts with more acceptable AMSFE. Among the four such strategies (ii) to (v) the improvement of forecasting accuracy for strategy (v) is very high. We conclude that the results from strategy (v) based on the marginal likelihood are best overall and this highly recommended.

5. Conclusion

In this study, we examined the forecasting performance in the context of the linear regression model when the error follows a first order moving average process. We investigated the accuracy of its forecasting performance on the basis of estimation via the marginal likelihood and concentrated likelihood. But there is always the issue of whether the estimates have come from a global maxima or a local maxima. We investigated five strategies to deal with the problem of local maxima and found they improve the accuracy of forecasting performance. We recommended that the strategy (v) combined with the use of the marginal likelihood be used for forecasting purposes.

Acknowledgments

The research reported in this paper was funded by an ARC grant.

References

Aptech System Inc., 1996, Gauss constrained optimization (Aptech Systems Inc, Maple Valley, WA).

Cooper, D.M. and R. Thompson, 1977, A note on the estimation of the parameters of the autoregressive process, Biometrika, 64, 525-628.

Cramer, J.S., 1986, Econometric applications of maximum likelihood methods (Cambridge University Cambridge).

Cryer, J.D. and J. Ledolter, (1981), Small-sample properties of the maximum likelihood estimator in the first-order moving average model, Biometrika, 68, 691-694.

Dunsmuir, W.T.M, (1981), Estimation for stationary time series when data are irregularly spaced or missing, in D.F. Findley (ed.), Applied Time Series II, pp609-649. New York, Academic press.

Kang, K.M., (1975), A comparison of estimators for moving average processes, Unpublished manuscript, Australian Bureau of Statistics.

King, M.L. and M. McAleer, 1987, Further results on testing AR(1) against MA (1) disturbances in the linear regression model, Review of Economic Studies 54, 649-663.

Lasker, M.R. and M.L. King, 1998, Estimation and testing of regression disturbances based on modified likelihood functions, Journal of Statistical Planning and Inference, 71, 75-92.

Rahman, S. and M.L. King, 1998, Marginal likelihood score based tests of regression disturbances in the presence of nuisance parameters, Journal of Econometrics, 82,81-106.

Tunnicliffe Wilson, G., 1998, On the use of marginal likelihood in time series model estimation, Journal of the Royal Statistical Society, B, 51, 15-27.

Yeasmin, M. and M.L. King, 2000, Test sizes and the importance of finding global maxima of the likelihood function, unpublished paper.

Table1

The Average Mean Squared Forecasts Error and Mean Absolute Errors of different strategies for finding the global maxima of different design matrix for $n = 30$ and the design matrix are X_1 (white noise) X_2 (autoregressive process).

	X_1				X_2			
No of Starting Value	**AMSE** ML	AMSE MML	MAE ML	MAE MML	AMSE ML	AMSE MML	MAE ML	MAE MML
				$\gamma = -.90$				
1	7.148	5.108	2.008	1.615	51.427	36.598	5.331	4.109
3	7.207	4.603	2.018	1.516	51.402	33.727	5.294	3.853
21	7.140	4.603	2.007	1.516	51.358	33.500	5.287	3.831
3*	7.235	4.602	2.023	1.516	51.057	31.209	5.258	3.644
21*	7.140	4.603	2.007	1.516	50.864	30.963	5.240	3.619
				$\gamma = -.60$				
1	27.237	11.667	3.568	1.960	71.202	35.933	4.907	2.971
3	26.956	8.625	3.540	1.622	66.264	26.873	4.665	2.477
21	26.544	8.621	3.492	1.622	63.339	26.225	4.500	2.438
3*	28.039	8.530	3.662	1.615	65.350	25.687	4.617	2.394
21*	26.544	8.525	3.492	1.615	61.018	24.678	4.370	2.327
				$\gamma = -.30$				
1	30.544	9.205	2.847	1.391	65.636	30.578	3.655	2.031
3	30.070	5.623	2.801	1.161	61.593	18.163	3.547	1.603
21	26.729	5.579	2.603	1.156	54.515	18.073	3.226	1.594
3*	33.726	5.623	3.058	1.161	63.871	18.242	3.663	1.596
21*	26.728	5.579	2.603	1.156	52.281	17.919	3.127	1.574
				$\gamma = 0$				
1	10.787	2.610	1.400	0.958	47.177	9.694	2.749	1.294
3	11.864	2.283	1.454	0.935	49.744	11.672	3.028	1.342
21	10.697	2.274	1.389	0.933	41.405	9.073	2.643	1.262
3*	17.241	2.354	1.698	0.941	58.345	9.942	3.498	1.284
21*	10.635	2.210	1.386	0.929	37.368	8.269	2.561	1.228
				$\gamma = .30$				
1	4.563	2.646	1.159	1.037	50.614	14.445	3.862	1.717
3	6.740	1.906	1.179	0.953	48.843	10.051	3.916	1.493
21	3.832	1.840	1.070	0.942	44.097	10.048	3.646	1.493
3*	9.780	2.028	1.276	0.964	56.725	10.172	4.308	1.498
21*	3.708	1.812	1.057	0.938	43.343	9.919	3.614	1.483
				$\gamma = .60$				
1	4.842	3.946	1.342	1.269	34.607	16.533	4.254	2.402
3	6.158	2.866	1.205	1.087	36.865	13.233	4.357	2.066
21	3.602	2.818	1.137	1.077	33.835	13.231	4.166	2.065
3*	6.650	2.639	1.190	1.050	37.546	12.995	4.433	2.045
21*	3.339	2.590	1.089	1.0405	33.768	12.940	4.159	2.037
				$\gamma = .90$				
1	2.733	2.595	1.282	1.247	14.626	9.690	2.968	2.2395
3	2.762	2.424	1.252	1.204	14.751	8.821	2.9859	2.1195
21	2.751	2.418	1.249	1.202	14.594	8.815	2.9641	2.1183
3*	2.494	2.216	1.178	1.148	14.866	8.682	2.9992	2.1005
21*	2.476	2.206	1.173	1.145	14.594	8.676	2.9641	2.0993

* only used of initial estimate is on the boundary.

Table 2
The Average Mean Squared Forecasts Error and Mean Absolute Errors of different strategies for finding the global maxima of different design matrix for $n = 60$ and the design matrix are X_1 (white noise) X_2 (autoregressive process).

	X_1				X_2			
No of Starting Value	**AMSE** ML	AMSE MML	MAE ML	MAE MML	AMSE ML	AMSE MML	MAE ML	MAE MML
				$\gamma = -.90$				
1	4.612	3.570	1.654	1.415	2.618	2.263	1.275	1.172
3	4.537	3.003	1.637	1.295	2.535	2.047	1.254	1.112
21	4.534	3.003	1.636	1.295	2.535	2.046	1.254	1.111
3*	4.421	2.787	1.615	1.246	2.455	1.870	1.230	1.060
21*	4.414	2.785	1.613	1.245	2.450	1.867	1.228	1.059
				$\gamma = -.60$				
1	9.741	6.378	1.892	1.477	4.850	3.256	1.429	1.168
3	6.855	2.390	1.543	1.000	3.675	1.844	1.253	0.952
21	6.444	2.392	1.496	1.000	3.584	1.806	1.236	0.947
3*	7.171	2.151	1.581	0.971	3.594	1.681	1.236	0.934
21*	6.418	2.154	1.491	0.972	3.425	1.613	1.207	0.922
				$\gamma = -.30$				
1	2.047	1.806	0.937	0.903	1.485	1.227	0.883	0.844
3	1.498	1.211	0.883	0.845	1.314	1.073	0.863	0.826
21	1.484	1.211	0.881	0.845	1.311	1.073	0.862	0.826
3*	1.789	1.111	0.911	0.836	1.336	1.099	0.867	0.829
21*	1.484	1.111	0.881	0.836	1.276	1.073	0.858	0.826
				$\gamma = 0$				
1	1.084	1.077	0.830	0.828	1.366	1.067	0.857	0.824
3	1.084	1.077	0.830	0.828	1.428	1.067	0.864	0.824
21	1.084	1.077	0.830	0.828	1.352	1.067	0.854	0.824
3*	1.177	1.077	0.837	0.828	1.713	1.067	0.877	0.824
21*	1.084	1.077	0.830	0.828	1.328	1.067	0.852	0.824
				$\gamma = .30$				
1	1.342	1.324	0.863	0.859	2.875	1.451	1.009	0.872
3	1.093	1.075	0.831	0.827	2.662	1.152	0.985	0.837
21	1.093	1.075	0.831	0.827	2.502	1.103	0.968	0.832
3*	1.103	1.075	0.832	0.827	3.561	1.152	1.064	0.837
21*	1.079	1.075	0.829	0.827	2.502	1.102	0.968	0.832
				$\gamma = .60$				
1	3.109	2.924	1.164	1.136	10.235	4.300	2.041	1.275
3	1.424	1.300	0.887	0.868	9.728	2.182	1.966	0.978
21	1.415	1.300	0.885	0.868	8.932	2.134	1.868	0.973
3*	1.243	1.178	0.865	0.851	9.808	2.046	1.988	0.961
21*	1.212	1.178	0.860	0.851	8.729	1.999	1.846	0.956
				$\gamma = .90$				
1	2.206	2.122	1.153	1.127	6.124	4.384	1.928	1.550
3	1.984	1.923	1.091	1.072	6.172	3.903	1.936	1.438
21	1.981	1.918	1.090	1.070	6.099	3.903	1.922	1.438
3*	1.691	1.610	1.011	0.986	6.181	3.614	1.938	1.380
21*	1.689	1.605	1.010	0.984	6.096	3.614	1.921	1.380

* only used of initial estimate is on the boundary.

Table 3

The Average Mean Squared Forecasts Error and Mean Absolute Errors of different strategies for finding the global maxima of different design matrix for $n = 120$ and the design matrix are X_1 (white noise) X_2 (autoregressive process).

	X_1				X_2			
No of Starting Value	**AMSE** ML	AMSE MML	MAE ML	MAE MML	AMSE ML	AMSE MML	MAE ML	MAE MML
				$\gamma = -.90$				
1	2.511	2.021	1.248	1.102	3.578	2.682	1.434	1.229
3	2.486	1.754	1.240	1.024	3.495	2.400	1.412	1.161
21	2.468	1.754	1.235	1.024	3.463	2.390	1.406	1.158
3*	2.423	1.562	1.222	0.967	3.115	1.872	1.325	1.038
21*	2.405	1.562	1.216	0.967	3.083	1.862	1.319	1.035
				$\gamma = -0.60$				
1	3.050	2.814	1.140	1.098	3.664	3.415	1.119	1.077
3	1.395	1.075	0.878	0.824	1.398	1.122	0.883	0.839
21	1.318	1.075	0.867	0.824	1.333	1.123	0.876	0.839
3*	1.390	1.075	0.878	0.824	1.406	1.100	0.885	0.836
21*	1.283	1.075	0.861	0.824	1.333	1.100	0.876	0.836
				$\gamma = -0.30$				
1	1.164	1.158	0.825	0.823	1.079	1.075	0.823	0.821
3	1.048	1.042	0.817	0.815	1.079	1.060	0.823	0.819
21	1.048	1.042	0.817	0.815	1.079	1.060	0.823	0.819
3*	1.048	1.042	0.817	0.815	1.125	1.060	0.829	0.819
21*	1.048	1.042	0.817	0.815	1.079	1.060	0.823	0.819
				$\gamma = 0$				
1	1.035	1.034	0.812	0.811	1.039	1.037	0.813	0.813
3	1.035	1.034	0.812	0.811	1.039	1.037	0.813	0.813
21	1.035	1.034	0.812	0.811	1.039	1.037	0.813	0.813
3*	1.035	1.034	0.812	0.811	1.039	1.037	0.813	0.813
21*	1.035	1.034	0.812	0.811	1.039	1.037	0.813	0.813
				$\gamma = 0.30$				
1	1.032	1.032	0.811	0.811	1.082	1.028	0.817	0.810
3	1.032	1.032	0.811	0.811	1.032	1.028	0.811	0.810
21	1.032	1.032	0.811	0.811	1.032	1.028	0.811	0.810
3*	1.032	1.032	0.811	0.811	1.032	1.028	0.811	0.810
21*	1.032	1.032	0.811	0.811	1.032	1.028	0.811	0.810
				$\gamma = 0.60$				
1	3.023	3.016	1.088	1.086	2.177	1.937	1.022	0.973
3	1.130	1.088	0.832	0.826	1.378	1.049	0.881	0.816
21	1.128	1.088	0.832	0.826	1.357	1.049	0.876	0.816
3*	1.102	1.081	0.830	0.824	1.379	1.049	0.881	0.816
21*	1.093	1.081	0.827	0.824	1.347	1.049	0.874	0.816
				$\gamma = 0.90$				
1	2.447	2.184	1.205	1.133	2.523	2.040	1.255	1.109
3	2.135	1.738	1.122	1.018	2.437	1.831	1.230	1.049
21	2.122	1.738	1.119	1.018	2.429	1.824	1.228	1.047
3*	1.803	1.397	1.032	0.922	2.270	1.581	1.181	0.977
21*	1.796	1.397	1.031	0.922	2.263	1.573	1.179	0.975

* only used of initial estimate is on the boundary.

Table 4

The Average Mean Squared Forecasts Error and Mean Absolute Errors of different strategies for finding the global maxima of different design matrix for $n = 30$ and the design matrix are X_3 (random walk) X_4 (explosive process).

	X_3				X_4			
No of Starting Value	**AMSE** ML	AMSE MML	MAE ML	MAE MML	AMSE ML	AMSE MML	MAE ML	MAE MML
				$\gamma = -.90$				
1	93.945	62.947	7.281	5.213	6985.212	5021.612	50.864	38.031
3	93.947	53.038	7.282	4.541	6988.732	4061.889	50.901	32.052
21	93.945	52.889	7.281	4.531	6985.212	4034.622	50.864	31.898
3*	94.045	53.056	7.294	4.543	7015.989	4059.004	51.045	32.053
21*	93.945	52.796	7.281	4.524	6985.212	4029.738	50.864	31.856
				$\gamma = -0.60$				
1	184.639	84.369	9.852	5.140	10627.040	5509.106	60.056	33.065
3	183.435	67.208	9.821	4.196	10623.840	3078.306	59.755	21.186
21	182.670	67.190	9.767	4.196	10481.310	3069.394	58.919	21.117
3*	188.659	67.208	10.069	4.196	10886.160	3068.493	61.137	21.115
21*	182.670	67.190	9.767	4.196	10481.310	3059.579	58.919	21.046
				$\gamma = -0.30$				
1	279.162	64.627	9.852	3.000	10263.040	3150.451	44.644	16.989
3	278.530	43.932	9.825	2.283	10125.750	1129.838	43.076	9.628
21	265.831	43.931	9.417	2.283	9269.741	1129.475	40.628	9.616
3*	319.838	43.932	11.053	2.283	11604.220	1129.840	47.988	9.628
21*	265.831	43.931	9.417	2.283	9269.741	1129.476	40.628	9.616
				$\gamma = 0$				
1	186.481	27.639	5.231	1.498	4855.581	511.788	18.880	5.001
3	216.147	14.525	5.763	1.362	5163.961	260.278	19.854	4.280
21	185.993	13.929	5.209	1.340	4801.152	260.278	18.573	4.280
3*	295.988	14.405	7.549	1.352	7092.873	263.213	25.310	4.313
21*	185.990	13.809	5.209	1.330	4801.152	260.231	18.573	4.276
				$\gamma = 0.30$				
1	52.961	2.078	2.164	1.035	2126.341	355.317	8.628	3.798
3	87.698	1.829	2.854	1.003	2642.560	466.288	10.378	3.966
21	52.720	1.791	2.134	0.998	1978.029	349.458	8.308	3.720
3*	142.624	1.796	3.863	0.999	3345.891	466.198	12.613	3.959
21*	52.668	1.750	2.127	0.992	1978.016	349.375	8.307	3.712
				$\gamma = 0.60$				
1	22.351	2.900	1.431	1.195	1419.626	17.367	5.181	2.860
3	31.160	2.266	1.535	1.087	1976.276	16.844	6.857	2.805
21	21.794	2.204	1.332	1.077	1470.975	16.818	5.300	2.802
3*	41.635	2.076	1.686	1.057	2102.241	16.645	7.176	2.787
21*	21.501	2.007	1.289	1.045	1418.917	16.585	5.110	2.779
				$\gamma = 0.90$				
1	12.824	2.201	1.333	1.167	94.130	13.441	2.917	2.557
3	12.756	2.096	1.314	1.138	184.189	13.439	3.185	2.555
21	12.740	2.086	1.310	1.135	184.313	13.533	3.196	2.564
3*	18.302	1.933	1.335	1.092	184.092	13.311	3.172	2.536
21*	12.576	1.922	1.263	1.089	184.207	13.425	3.183	2.548

* only used of initial estimate is on the boundary.

Table 5

The Average Mean Squared Forecasts Error and Mean Absolute Errors of different strategies for finding the global maxima of different design matrix for $n = 60$ and the design matrix are X_3 (random walk) X_4 (explosive process).

	X_3				X_4			
No of Starting Value	**AMSE** ML	AMSE MML	MAE ML	MAE MML	AMSE ML	AMSE MML	MAE ML	MAE MML
				$\gamma = -0.90$				
1	13.601	10.445	2.419	2.044	1149.441	716.947	20.104	13.854
3	13.664	8.016	2.426	1.752	1145.368	642.529	20.016	12.711
21	13.601	8.016	2.419	1.752	1139.864	642.375	19.942	12.710
3*	13.673	7.683	2.428	1.716	1145.178	622.470	20.006	12.472
21*	13.601	7.643	2.419	1.712	1139.674	620.299	19.932	12.442
				$\gamma = -0.60$				
1	40.350	21.208	3.563	2.253	2615.827	1055.210	23.932	11.554
3	33.324	5.482	3.094	1.256	2074.367	295.772	19.823	5.619
21	32.522	5.482	3.042	1.256	1998.397	295.771	19.239	5.619
3*	34.551	5.128	3.178	1.237	2152.960	278.926	20.545	5.524
21*	32.487	5.127	3.036	1.237	1986.023	278.908	19.104	5.524
				$\gamma = -0.30$				
1	11.466	4.436	1.461	1.061	890.383	240.279	8.245	4.381
3	9.298	1.473	1.310	0.922	611.386	48.226	6.881	3.239
21	7.752	1.473	1.267	0.922	576.390	48.226	6.708	3.239
3*	12.553	1.473	1.470	0.922	860.206	48.226	8.159	3.239
21*	7.752	1.473	1.267	0.922	576.390	48.226	6.708	3.239
				$\gamma = 0$				
1	2.194	1.751	0.942	0.906	69.020	16.347	3.126	2.756
3	2.451	1.263	0.953	0.893	69.020	16.347	3.126	2.756
21	2.194	1.263	0.942	0.893	69.020	16.347	3.126	2.756
3*	3.360	1.263	0.989	0.893	86.188	16.347	3.230	2.756
21*	2.194	1.263	0.942	0.893	69.020	16.347	3.126	2.756
				$\gamma = 0.30$				
1	1.405	1.387	0.903	0.897	15.936	14.971	2.682	2.605
3	1.254	1.235	0.887	0.881	15.044	14.077	2.651	2.573
21	1.254	1.235	0.887	0.881	15.044	14.076	2.651	2.573
3*	1.239	1.221	0.885	0.879	15.044	14.077	2.651	2.573
21*	1.239	1.221	0.885	0.879	15.044	14.076	2.651	2.573
				$\gamma = 0.60$				
1	2.387	2.381	1.073	1.071	19.710	19.343	2.908	2.881
3	1.328	1.414	0.894	0.902	14.220	13.954	2.584	2.566
21	1.329	1.414	0.894	0.902	14.219	13.954	2.584	2.566
3*	1.266	1.261	0.885	0.884	14.132	13.742	2.575	2.545
21*	1.267	1.261	0.886	0.884	14.132	13.742	2.575	2.545
				$\gamma = 0.90$				
1	1.935	1.939	1.089	1.091	11.933	11.854	2.433	2.426
3	1.794	1.790	1.047	1.045	11.483	11.367	2.379	2.372
21	1.792	1.789	1.046	1.045	11.471	11.361	2.378	2.372
3*	1.545	1.542	0.975	0.974	11.027	10.870	2.325	2.309
21*	1.543	1.541	0.975	0.974	11.016	10.892	2.324	2.312

* only used of initial estimate is on the boundary.

Table 6

The Average Mean Squared Forecasts Error and Mean Absolute Errors of different strategies for finding the global maxima of different design matrix for $n = 120$ and the design matrix are X_3 (random walk) X_4 (explosive process).

	X_3				X_4			
No of Starting Value	**AMSE** ML	AMSE MML	MAE ML	MAE MML	AMSE ML	AMSE MML	MAE ML	MAE MML
				$\gamma = -0.90$				
1	22.866	15.454	3.359	2.540	38.432	25.778	3.987	3.035
3	21.164	11.028	3.211	2.104	37.561	18.947	3.927	2.492
21	21.066	11.027	3.201	2.104	37.407	18.946	3.910	2.492
3*	18.619	7.823	2.987	1.761	37.160	14.603	3.901	2.177
21*	18.521	7.823	2.978	1.761	37.007	14.601	3.884	2.177
				$\gamma = -0.60$				
1	34.485	33.785	2.402	2.308	118.984	0 95.229	4.265	3.475
3	2.775	1.328	1.047	0.891	27.536	4.152	2.098	1.300
21	2.704	1.328	1.038	0.891	25.783	4.152	2.035	1.300
3*	2.483	1.328	1.033	0.891	30.031	3.291	2.152	1.273
21*	2.387	1.328	1.019	0.891	25.154	3.291	2.006	1.274
				$\gamma = -0.30$				
1	2.160	1.167	0.894	0.861	3.250	2.605	1.267	1.214
3	2.153	1.160	0.893	0.860	2.713	2.511	1.246	1.207
21	2.153	1.160	0.893	0.860	2.713	2.511	1.246	1.207
3*	2.153	1.160	0.893	0.860	2.713	2.511	1.246	1.207
21*	2.153	1.160	0.893	0.860	2.713	2.511	1.246	1.207
				$\gamma = 0$				
1	1.145	1.136	0.853	0.850	2.550	2.460	1.212	1.193
3	1.145	1.136	0.853	0.850	2.550	2.460	1.212	1.193
21	1.145	1.136	0.853	0.850	2.550	2.460	1.212	1.193
3*	1.145	1.136	0.853	0.850	2.550	2.460	1.212	1.193
21*	1.145	1.136	0.853	0.850	2.550	2.460	1.212	1.193
				$\gamma = .30$				
1	1.126	1.122	0.847	0.846	2.440	2.395	1.183	1.173
3	1.126	1.122	0.847	0.846	2.440	2.395	1.183	1.173
21	1.126	1.122	0.847	0.846	2.440	2.395	1.183	1.173
3*	1.126	1.122	0.847	0.846	2.440	2.395	1.183	1.173
21*	1.126	1.122	0.847	0.846	2.440	2.395	1.183	1.173
				$\gamma = .60$				
1	2.836	2.823	1.071	1.069	3.460	3.451	1.328	1.326
3	1.153	1.164	0.854	0.853	2.452	2.403	1.190	1.183
21	1.135	1.164	0.850	0.853	2.448	2.403	1.189	1.183
3*	1.153	1.134	0.854	0.850	2.425	2.402	1.187	1.182
21*	1.135	1.134	0.850	0.850	2.421	2.402	1.186	1.182
				$\gamma = 0.90$				
1	2.205	2.140	1.137	1.120	3.189	3.162	1.372	1.366
3	1.818	1.834	1.040	1.044	2.972	2.980	1.326	1.327
21	1.815	1.834	1.039	1.044	2.972	2.980	1.326	1.327
3*	1.450	1.343	0.941	0.911	2.732	2.644	1.271	1.252
21*	1.446	1.343	0.940	0.911	2.732	2.644	1.272	1.252

* only used of initial estimate is on the boundary.